Discussion Paper Series – CRC TR 224

Discussion Paper No. 655
Project A 01

Sticky Models

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

February 2025

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

February 17, 2025

Abstract: People often form mental models based on incomplete information, revising them as new relevant data becomes available. In this paper, we experimentally investigate how individuals update their models when data on predictive variables are gradually revealed. We find that people's models tend to be 'sticky,' as their final models remain strongly influenced by earlier models formed using a subset of variables. Guided by a simple framework highlighting the role of attention in shaping model revisions, we document that only participants who exert lower cognitive effort during the revising stage, relative to the initial model formation stage – as proxied by time spent – exhibit significant model stickiness. Additionally, subjects' final models are strongly predicted by their reasoning type – their self-described approach to extracting models from multidimensional data. While model stickiness varies across reasoning types, effort allocation across stages remains a strong predictor of stickiness even when accounting for reasoning.

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. **Acknowledgements**: We thank Botond Kőszegi, Chris Roth, and Florian Zimmermann for their outstanding supervision. We thank Chiara Aina, Peter Andre, Benjamin Enke, Nicola Gennaioli, Duarte Gonçalves, Thomas Graeber, Luca Henkel, Ulrike Malmendier, Robin Musolff, Ryan Oprea, Joshua Schwartzstein, Ran Spiegler, Mark Toth, Emanuel Vespa, and participants at various conferences and seminars for helpful comments and discussions. **Funding**: Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project A01) and under 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. **Preregistration**: Our data collection was preregistered at `https://aspredicted.org/xqfv-9s2d.pdf`.

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 with limited data availability where not all relevant variables are observable at once, economic agents make decisions based on partial mental models that capture the relationship between a subset of variables. As new variables emerge, rational agents should revise their models in order to reflect the *joint* importance of both new and existing explanatory 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 seeking to predict expected asset returns forms a preliminary model based on a set of risk factors. When additional factors are proposed, she might fail to adopt the more complex model and instead shift paradigms 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. 2023; 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 additional variables become available. While extensive literature provides evidence on how people update their beliefs 'within' models as they gather more observations, less is known about how people learn 'between' models when faced with new *dimensions* of information. This raises two important questions: How do people revise their mental models in dynamic environments when confronted with new economic variables? Do individuals correctly revise their models, or are they prone to interpret new evidence through the lens of their pre-existing models?

We address these questions experimentally by investigating how people form mental models that capture the statistical relationships in data and how they revise these models once data on additional variables is revealed. We study path dependence in model formation, whereby pre-existing partial models shape final models. Specifically, we hypothesize that mental models are *sticky*, meaning that people insufficiently revise their initial mental models in such dynamic environments.

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 update 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.

To address our research questions, we designed an experimental paradigm that captures the essence of dynamic model learning. In each stage of our experiment, 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 dataset is such 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 a strong omitted variable bias. In each stage, subjects see 40 data points that are initially revealed one by one so that subjects 'experience' the data. Participants then encounter pairs of new projects with unknown outcomes that differ in the value of exactly one predictor variable 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 variable values. This set of beliefs constitutes a complete statistical model of the relationship between the predictor variables and the outcome which closely maps to a parametrized linear regression. To corroborate our measure for people's statistical models, we also ask subjects to state whether they expect a ceteris paribus variation in X or Z to affect the outcome at all.

Several experimental design features enable us to obtain clean proof-of-concept evidence for stickiness in model formation and to shed light on underlying cognitive mechanisms. First, the dataset in the second stage is identical for both the *X-first* and *Z-first* groups, ensuring that any treatment differences can be attributed to the variable observed in the first stage. Second, subjects first form partial models, 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. Third, given our objective to better understand how people integrate new variables into their models when learning from data, we explicitly ask subjects to describe their strategy for forming their conditional beliefs in the second stage. This allows us to gather rich qualitative process data that provides additional insights into the cognitive foundations of model stickiness.

We test our central hypothesis by measuring how the observed variable in the first stage affects subjects' models in the second stage. 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 intensive margin, subjects in the *X-first* group estimate the marginal effect of X on success to be, on average, 2.6 percentage points (pp)—or about 50%—higher than their counterparts in the *Z-first* group. On the extensive margin, *X-first* subjects are 10 percentage points more likely to agree that variable X conditionally affects the probability of a successful outcome. In contrast, both groups have similar beliefs regarding the marginal impact of Z on either margin. Evidence from choice data between projects with different predictor value combinations is qualitatively consistent with results on conditional beliefs, although the data is noisier overall, such that we partly do not find statistically significant differences between treatment groups.

We thus find evidence in support of our hypothesis. The *X-first* group incorporates (the conditionally predictive) variable Z to the same extent as the *Z-first* group but does not fully let go of what they initially learned about (the conditionally unpredictive) variable X.

To derive our hypotheses and highlight underlying cognitive mechanisms, we introduce a simple conceptual framework based on a two-step model of cognition (Ba et al., 2023). In the framework, dynamic model learning is driven by how agents allocate cognitive effort across stages and what summary statistics of the data set they focus on when revising their model in the second stage. Stickiness then arises from subjects exerting too little effort in revising their model and thus defaulting to what they already learned in the first stage.

We investigate how subjects attend to different statistics in the complete data set by analyzing the alignment of subjects' beliefs with different empirical benchmarks. Specifically, we regress subjects' beliefs on the empirical frequencies of success $P(Y|X = x)$ (Benchmark X) and $P(Y|Z = z)$ (Benchmark Z). This structural regression reveals that the treatment effects can be characterized by subjects' second-stage models more strongly incorporating the unconditional empirical benchmarks from the variable they already observed in their respective first stage.

But what drives model stickiness in the data? In line with the predictions from our theoretical framework, we show that the allocation of cognitive effort across stages strongly predicts model stickiness: We find that subjects who allocate below-median time to the second stage relative to the first stage exhibit significant stickiness, while

those who invest relatively more effort in stage two show no stickiness.

In addition to *how much* effort subjects allocate to revising their model in the second stage, how they reason about the problem – and therefore *what* summary statistics they attend to – may also empirically interact with model stickiness. To test the robustness of the proposed cognitive effort mechanism and assess the role of heterogeneous mental representations of the problem, we elicit how subjects go about revising their model in an open-ended question and use the coded answers to identify three distinct reasoning styles: *Frequentists* solve the model revision problem rationally by estimating the relative success likelihood for all variable value combinations, while *Separate* reasoners fail to account for the correlation between X and Z. *Absolute success* reasoners attend to the raw number of successes, thereby committing base-rate neglect. Our analyses show that subjects' final models are strongly predicted by their reasoning type. Moreover, while model stickiness varies across reasoning types, effort allocation remains a robust predictor even when accounting for reasoning type.

Related Literature: To the best of our knowledge, this is the first paper that empirically investigates how people revise their models when there is an exogenous change to the environment. Thereby, 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. (2024) use a closely related setting with two binary inputs and one output to study how people learn stochastic relationships from datasets. They focus on how the noise in the statistical relationship between variables affects people's prediction models and identify two frequently-made types of errors in static model learning from data: (1) failures to properly condition on the correlations of relevant variables and (2) failures to use the correlations in the data optimally. Similarly, Kendall and Oprea (2024) study how people learn models based on data but instead focus on deterministic rules of varying complexity. While the general setting of those papers is similar to ours to the extent that people extract models from datasets with an ex-ante unknown data-generating process, we significantly depart from past work by studying dynamics in data-driven model learning.

We also relate to another strand of the literature that focuses on models as causal narratives. Charles and Kendall (2024) experimentally illustrate how externally provided models can influence participants' data interpretation, supporting the core predictions of Eliaz and Spiegler (2020) in a setting closely related to ours. In a similar vein, Ambuehl and Thysen (2024) examine how individuals select among various causal models and find that people tend to prefer more complex models. While our design is not intended to disentangle all competing causal representations, it provides clean evidence on how initially perceived causal links can persist even when alternative explanations

better explain patterns in the data.

Regarding the empirical study of the dynamics of mental models, the most closely related paper is Esponda et al. (2024). The authors experimentally show that misspecified models can persist despite regular feedback in a belief formation environment where people are prone to base-rate neglect. We depart from their work in several important ways: First, instead of focusing purely on a notion of models capturing the mental representation of the task at hand and how to solve it, we study mental models capturing the relationships between multiple variables at both intensive and extensive margins. Our conceptualization of models is much less restrictive and applies to a vast set of problems where agents learn about the relationship between multiple inputs and one output. Second, instead of knowing all relevant parameters to solve the problem optimally, subjects in our experiment need to extract models from a dataset. Third and most importantly, we force a significant model revision for a randomized subsample of participants, which allows us to cleanly measure model revisions independently of prior mistakes or misrepresentations of the environment. Finally, we provide novel evidence on the cognitive mechanisms underlying failures in model revision.

Several findings from the field are consistent with broad interpretations of insufficient model revision, although it leaves open the question of whether it arises more generally. For instance, Hanna et al. (2014) show that seaweed farmers neglect pod size - a relevant production input - and just observing data from the experiment is not sufficient for them to attend to it. Providing summary statistics, however, helps farmers optimize along the neglected dimension. Their setting differs from ours in several aspects. Farmers in their setting could, in principle, attend to and optimize all relevant inputs from the beginning on. In contrast, we study model revision when relevant relationships in the data can only be learned later on and where strong priors are unlikely to constrain attention to variables ex-ante.

More closely mapping to the dynamics of our experiment is the work by Liu and Zhang (2024), who show that initial narratives on genetically modified mosquitoes shape how people interpret later complementary information despite having the opportunity to read opposing-side narratives. Similarly, Macchi (2023) shows that in a field setting, first impressions from personal meetings has a lasting effect on creditors' loan decisions despite detailed financial information being revealed subsequently. Building on this evidence, our experiment allows us to test whether such model stickiness "survives" in an abstract environment where we provide people with all available data to learn the relevant relationships between the variables throughout the experiment and prompt them to think about the relevant contingencies.

We also add to an extensive literature on path dependency in belief updating, such as confirmation bias (Rabin and Schrag, 1999) and prior-biased updating (see Benjamin

2019 for a review). Unlike previous theoretical and empirical work that holds the model space fixed, our study examines a dynamic setting in which the set of explanatory variables expands across two experimental stages. In the first stage, subjects form models based on a coarser distribution of $Y|X$ (or $Y|Z$), and in the second stage, they are required to revise their models and reason about the refined distribution $Y|X, Z$. Our findings not only strengthen existing evidence for path dependence — by showing that initial models continue to influence subsequent belief formation — but also shed light on the cognitive origins of these effects.

We also connect to the literature documenting belief biases in complex environments, such as subjects stating attenuated beliefs Enke and Graeber (2023), neglecting the signal-generating process underlying the information (Enke, 2020; Enke and Zimmermann, 2019), or failing to reason through all relevant contingencies (Niederle and Vespa, 2023).

Finally, this project relates to a recent and growing theoretical literature on misspecified models and their implications. In this literature, it is generally assumed that agents' (misspecified) models shape the way new observations are interpreted. Several underlying cognitive mechanisms have been proposed to give rise to model misspecification, among which most prominently are limited attention (e.g., Bordalo et al. 2024; 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: Gagnon-Bartsch et al. (2023) study the conditions under which agents become aware of model misspecification, while Lanzani (2024) studies how the concern of being misspecified affects dynamic model choice. Ba (2023) examines which misspecified models persist in the long run and finds that simple models can be more robust than correctly specified complex models. We provide a simple theoretical model to derive our main predictions and empirically inform the theoretical literature by providing evidence on how the environmental dynamics of data availability shape models.

2 Experimental Design

This experiment aims to study *path dependence* in model formation and revision. While the introductory examples showcased the broad empirical relevance of dynamic model formation, the co-occurrence of other, possibly confounding phenomena makes it challenging to investigate our research question using observational data.

An experiment with minimal framing is necessary to understand the mechanisms underlying model revision since it allows the disentangling of cognitive primitives from

alternative explanations such as context-driven motivated reasoning or heterogeneity in priors. Specifically, studying inertia in how individuals revise their models as they receive new dimensions of information without deception calls for an experimental paradigm that addresses the following requirements:

1. A sufficiently rich learning environment featuring a *multi-dimensional dataset*, where multiple input variables may help predict an outcome variable,
2. a dynamic setting encompassing *multiple stages*, including an initial model formation and a subsequent, deception-free model revision stage,
3. *exogenous variation in the dynamics of information provision* while holding constant the total amount of information received by the end of the experiment and
4. a data-generating process (DGP) where the degree of model revision can be mapped to a clear *statistical benchmark*.

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 past projects. 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

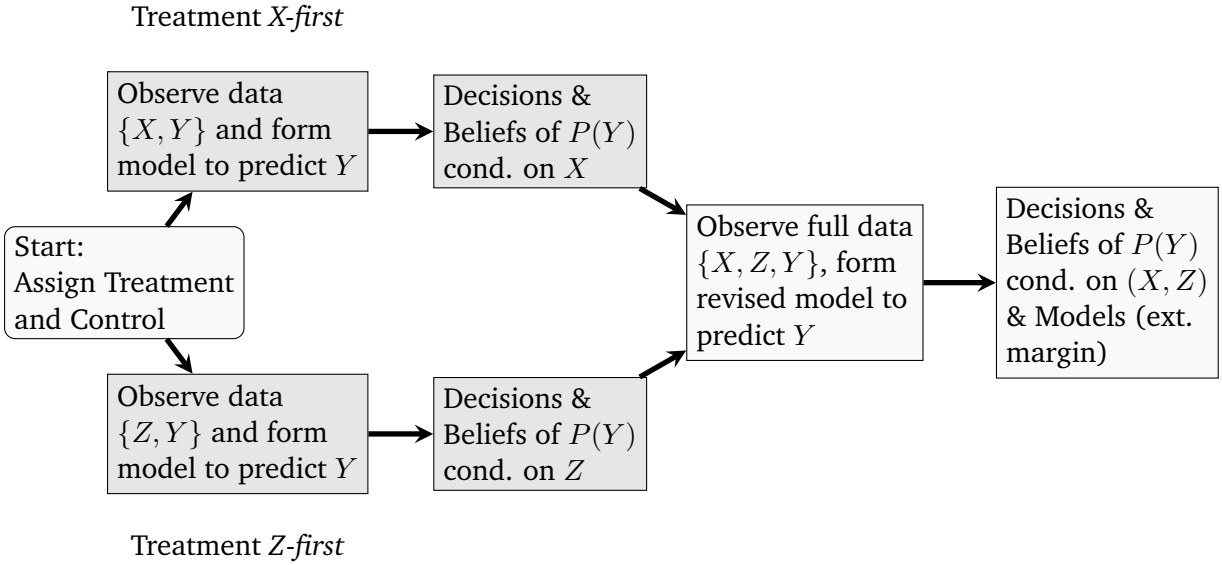


Figure 1 shows the timeline of our experiment. At the beginning of the experiment, subjects first pass a basic attention check before acquainting themselves with the general setting. Only subjects that pass comprehension questions can proceed to the main part. The main part consists of two stages. For the main part, subjects are randomly assigned to a treatment group that exogenously varies the predictor variable 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 features but that they are only able to observe one randomly selected variable[1].

In the second stage, the formerly missing second variable is revealed 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 variable. Further, in each stage, we ensure that subjects pay at least some attention to the data by forcing them to stay on the page while 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 predictor variables and the outcome using reported beliefs, binary project choices, and willingness to pay between projects. At the end of the second stage, we additionally ask subjects to describe their approach to the task in an open-ended question[2] and elicit an additional measure about the perceived marginal impact of each variable.

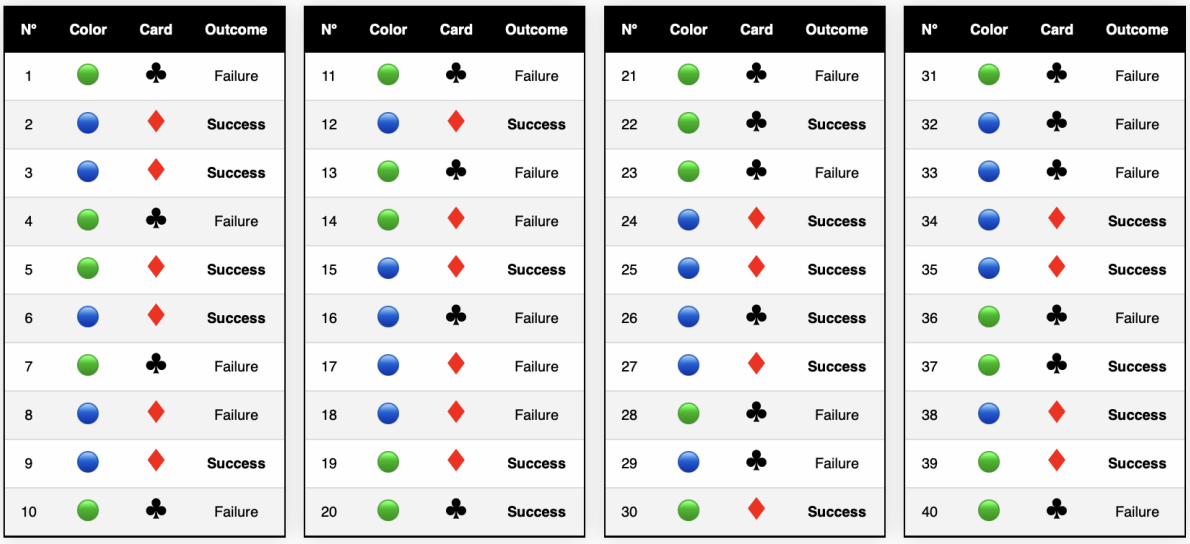
2.2 Data Generating Process

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

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

[1] Announcing the general information structure at the beginning avoids heterogeneity in beliefs about the experimental structure by treatment without making it explicit that more information will be available after the first stage

[2] We elicit subjects' reasoning last in order not to influence their other choices or beliefs.



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

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

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

N°	Color	Card	Outcome
31	●	♣	Failure
32	●	♣	Failure
33	●	♣	Failure
34	●	♦	**Success**
35	●	♦	**Success**
36	●	♣	Failure
37	●	♣	**Success**
38	●	♦	**Success**
39	●	♦	**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 participants' 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 this DGP, Z (Card) is highly predictive of Y ($P(Y|Z=1) - P(Y|Z=0) = 0.6$), 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 ($P(Y|X=1) - P(Y|X=0) = 0.3$)).

In the first stage, where subjects only encounter X or Z, the rational benchmark is 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., $P(Y|X,Z)$. 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[3].

2.3 Measurement of Models and Incentives

We measure subjects' statistical mental models using different measurements: 1) conditional success likelihoods for projects with different combinations of predictor variable

[3]Note that this data structure in principle may give rise to stickiness along two dimensions: subjects in *X-first* (a) attributing comparatively more (causal) weight to X and (b) putting less weight on Z than the *Z-first* group.

values, 2) binary choices between projects that differ in one independent variable, and 3) willingness to pay for a *ceteris paribus* variation in one independent variable. In the second stage, we additionally measure 4) agreement (and confidence in) the statements that each predictor variable has a marginal effect on the likelihood of success. For each future project, subjects observe the value of that project's independent variable, but the outcome is unknown.

We incentivize choices and beliefs by randomly selecting one of the choices throughout the experiment to count for their bonus payment for every tenth participant. For the choice that counts, the outcome was drawn after the experiment according to the conditional success probability described in 2.2.

Decisions (binary choice and willingness-to-pay): Using a two-step procedure, we elicit binary choice and willingness-to-pay between two future projects that differ in exactly one predictor variable's value and, accordingly, feature different success probabilities. In the first step, we ask subjects to make a binary choice and select their preferred project. Subjects are incentivized to choose the project they believe to have a higher likelihood of success, as they receive a bonus payment of $10 if the realization of the future project is a success and $0 if it is a failure.

Having decided, subjects are asked to indicate their willingness to pay to stay with their preferred project instead of switching to their less preferred project. For this purpose, subjects are presented with a multiple price list (MPL) consisting of 21 rows, where each row reflects a choice between their preferred project and the alternative, along with an additional fixed payment for choosing the alternative. The fixed payment increased from $0 to $10 in increments of $0.5. One of these rows is randomly selected for a potential bonus payment.

Mental models (intensive margin) - conditional success probability: After making their project decisions, 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. 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?

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 the true success likelihoods of the respective projects.

Model formation process: After eliciting conditional success likelihoods in the second stage, subjects describe in an open-text format how they formed their beliefs, i.e., how they determined the likelihood of success of the projects with different combinations of variable values. For this purpose, we use 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.

Mental models (extensive margin): At the end of the experiment, subjects are asked whether they believe each of the two statements below 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.

In addition, subjects provide their confidence in each of their answers on a slider from 0% (not at all confident) to 100% (extremely confident).

3 Framework and Research Hypotheses

To guide our analyses and derive hypotheses, we develop a stylized framework of dynamic model formation based on a two-step model of cognition (Bordalo et al. 2023, Ba et al. 2023). Subjects first form a mental representation of the statistical problem depending on their reasoning type before exerting cognitive effort attending to the summary statistics of the data set that are comprised in their representation. Both steps jointly inform the dynamics of model formation.

Framework. Subjects in our experiment are faced with constructing a (statistical) *mental model* of project success, that is, a complete set of beliefs about conditional success probabilities[4]. From this model, agents can directly derive the role of each variable in

[4]As agents only observe binary dependent and independent variables, this captures agents' full pre-

predicting project success. Agents form their mental models by attending to statistics about the observations of past projects that constitute a *data set*. While there are many statistics agents can extract from the data set, only some help to form the correct mental model. Every agent first forms a representation of the statistical problem. Each representation corresponds to an estimator $\hat{P}$ for the conditional success probabilities and is a function of various summary statistics extracted from the data. It therefore determines what summary statistics the agent attends to. In the second step, she (imperfectly) learns about these statistics by counting observations and combines them according to her representation to form a model about the success probability of projects with different features. The accuracy of extracting and combining statistics depends on the agent's cognitive effort.

Agents in our experiment face two statistical problems in the two stages of the experiment that are related to each other. In the first stage, agents face a simpler statistical problem of learning how one independent variable $S_1 \in \{X, Z\}$ affects a project's success probability Y. In the second stage, agents face a more complex statistical problem in learning how two independent variables X and Z affect the success probability separately or jointly. In both stages, subjects see a table with 40 rows, each representing the outcome of 1,000 project realizations, constituting a dataset of 40,000 draws from the joint distribution of (S_1, Y) and (X, Z, Y) respectively.[5] Based on these observations, agents learn about the success probability conditional on the available predictor variables.

In our setting, where agents observe a large number of signals, the importance of the prior is vanishing, which means that the rational Bayesian approach corresponds approximately to a Frequentist approach that equates conditional success probabilities with the corresponding empirical frequencies. The rational benchmark is therefore given by the data-generating process described in section 2.2.

Dynamics of model formation: In the first stage, subjects exert cognitive effort t_1 to form beliefs $\mu_{s_1}(t_1)$ about $P(Y = 1|S_1)$, the probability of success conditional on their observed first-stage variable S_1. Let $\hat{P}_{S_1}$ denote their representation of the statistical problem in the first stage. The accuracy of their beliefs depends on the time t_1 they spend in the first stage as well as their representation:

$$\mu_{s_1}(t_1) = \delta(t_1) \cdot \hat{P}_{S_1=s_1} + (1 - \delta(t_1)) \cdot 0.5 \tag{1}$$

diction model, i.e., the distribution of the outcome conditional on independent variables.

[5]We ensure in comprehension questions that subjects understand the large number of signals they observe.

In the second stage, the space expands for all subjects. For each of the 40 rows of the data set, subjects can now observe the second variable S_2 they did not observe in the first stage. As subjects move from the first to the second stage of the experiment, they need to expand their mental model and form beliefs μ_{s_1,s_2} about $P(Y = 1|S_1, S_2)$, the success probabilities based on both independent variables. As the second stage expands the sample space, subjects can attend to many more statistics. In particular, they can attend to the relative frequency of success conditional on any subset of the independent variables $\{X, Z\}$.[6] Let $\hat{P}_{(S_1,S_2)}$ denote the subject's representation of the statistical problem in the second stage, capturing the summary statistics that the subject relies on to estimate the probability.

Agents who don't attend to the additional data in the second stage fully rely on their corresponding first-stage beliefs. As agents attend to the novel information according to their representation $\hat{P}$, they move away from their reported first-stage belief μ_{s_1} towards a second-stage estimator $\hat{P}_{(S_1,S_2)}$. Their reported second-stage beliefs μ_{s_1,s_2} are therefore a convex combination of first-stage beliefs μ_{s_1} and the 'perfect' belief according to their mental representation $\hat{P}_{(S_1=s_1,S_2=s_2)}$, that is

$$\mu_{s_1,s_2}(t_2) = \gamma(t_2) \cdot \hat{P}_{(S_1=s_1,S_2=s_2)} + (1 - \gamma(t_2)) \cdot \mu_{s_1}(t_1) \tag{2}$$

$$= \gamma(t_2) \cdot \hat{P}_{(S_1=s_1,S_2=s_2)} + (1 - \gamma(t_2))\delta(t_1) \cdot \hat{P}_{S_1=s_1} + (1 - \gamma(t_2))(1 - \delta(t_1)) \cdot 0.5 \tag{3}$$

where t_2 reflects the time they spend in the second stage and $\gamma(t_2) \in [0, 1]$ is weakly increasing in $t_2 \in [0, T_{max})$.

Agents that update less in the second stage exhibit beliefs closer to their respective first-stage beliefs. As long as agents do not perfectly update in the second stage, this will lead to a stronger dependence of agents' second-stage models based on the features they already observed in the first stage.

As described in Hypothesis 1, we therefore expect agents to exhibit *sticky models*. In the first stage, both groups learn that their respective first-stage variable S_1 has a marginal impact on the probability of success. Therefore, we expect that agents assign (weakly) more weight to their first stage variable in explaining project success. This might be partly reflected in their actions.

Hypothesis 1 ***(Sticky models)*****:** *Agents exhibit path dependency in model formation, as the variable agents observe first predicts their final second-stage model. Agents' models are 'sticky,' as they tend to incorporate weakly more statistics already observable in their respective first stage compared to agents who observe these statistics only in the second*

[6] That is, the set of possible features a subject can attend to is the power set of the independent variables, i.e. $2^{\{X,Z\}} = \{\emptyset, \{X\}, \{Z\}, \{X, Z\}\}$

stage.

Second-stage cognitive effort. In the second stage, agents spend time attending to features unavailable in the first stage. The less time a subject spends in the second stage, the less they learn about the joint importance of the features of the second stage. As a result, their second-stage beliefs should be closer to the corresponding first-stage belief, increasing the difference between treatment groups. In the limit case, an agent who does not exert any cognitive effort in the second stage only incorporates first-stage statistics into their model independent of their representation of the problem. The accuracy of their first-stage beliefs and how strongly they internalize the importance of their first-stage feature depends on the time spent in the first stage. Consequently, the propensity for *sticky models* is more pronounced among agents who spend less time in the second stage, particularly in relation to their first-stage effort[7]. Agents investing more time in the second stage learn more precisely about the second-stage statistics they attend to. In the limit case, agents who exert sufficient cognitive effort in the second stage will learn all the statistics they attend to with high accuracy.

Hypothesis 2: ***(Sticky models are driven by cognitive effort)*****:** *The less time agents spend in the second stage, the more their final models incorporate the statistics of their first stage. The difference in final models is larger for agents exerting less cognitive effort in the second stage, particularly relative to their effort in the first stage.*

Reasoning types: In addition to cognitive effort governing how agents learn summary statistics that are attended to, *which* statistics an agent attends to in the first place crucially shapes the model $\hat{P}_{(S_1=s_1,S_2=s_2)}$ they would acquire given sufficient effort. We consider heterogeneity in subjects' mental representation of the problem by assigning each participant a reasoning type. Reasoning types differ from the cognitive effort discussed above, as two subjects that spend the same time (exerting the same cognitive effort) in the second stage can still attend to very different statistics and ultimately form very different models.

The *rational* reasoning type represents the conditional success probability as the unconditional success probability relative to the base rate. Their estimator $\hat{P}^R$ is equal to the number of successful projects with $(X = x, Z = z)$ relative to the overall number of projects with $(X = x, Z = z)$.

[7]There are two possible pathways through which this might operate. First, agents allocating attention bottom-up to all relevant features might still recall their first-stage features, leading to higher precision about these statistics. Second, for agents rationally allocating attention in the second stage, the first-stage attention allocation is akin to a cognitive 'sunk cost,' leading to an over-allocation of cognitive bandwidth to the first-stage features from the agent's perspective in the second stage. We don't take a stance here on which mechanism prevails.

Agents who don't attend to the base rate or incorporate other summary statistics would not learn the true data-generating process even if they exerted sufficient cognitive effort. In addition to the *rational* reasoning type, we consider two additional types, which we refer to as *Separate* ($\hat{P}^S$) and *Absolute Success* ($\hat{P}^A$). We describe these reasoning types in greater detail in Section 4.4. We expect heterogeneity in beliefs within both treatment groups based on their representation $\hat{P}_{(S_1,S_2)}$ of the statistical problem.

Hypothesis 3 ***(Reasoning shapes model formation)*****:** *Subjects' approach to extracting information from data or 'reasoning type,' as observed from their open text qualitative reasoning, predicts their second-stage model.*

Reasoning and cognitive effort. Subjects' reasoning types, as we define them, shape what a subject would learn if she exerted sufficient cognitive effort. As shown in equation 2, holding constant the reasoning type, higher cognitive effort in the second reasoning step should 'sharpen' a subject's final model, moving her beliefs from her first-stage model μ_{S_1} towards the model resulting from her mental representation. This need not move agents closer to the rational model. Conversely, lower cognitive effort in the second stage leads to greater model stickiness, across reasoning types.

Hypothesis 4 ***(Final models are determined by reasoning and cognitive effort)*****:**

a. *Lower cognitive effort in the second stage, measured through more time spent on the second stage in total and relative terms, increases stickiness across reasoning types.*

b. *The effect of increased cognitive effort in the second stage depends on agents' reasoning type.*

4 Results

In this section, we present the experimental results. First, we examine how subjects' second-stage models depend on the variables observed in the first stage using both reduced-form specifications and a structural approach. Second, we investigate the role of cognitive effort in path dependence using different timing-based measures. Finally, we examine the impact of reasoning on model formation and assess how the allocation of effort across stages affects path dependence while accounting for reasoning.

As pre-registered, all analyses exclude subjects who are among the 1% fastest and the 1% slowest as measured by the total time spent on our experiment to limit the results

being driven by outlier observations. This screening device excludes 16 subjects, which brings our total sample size to $n = 784$ subjects.

4.1 Summary Statistics

We ran the 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). 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[8]. 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, our subjects are broadly representative of the US population. However, our subjects tend to be younger and better educated, which is a common phenomenon in online samples. Across the *X-first* and *Z-first* treatment groups, demographics are well-balanced with no statistically significant differences. Therefore, we do not include demographic controls in our main regression specifications.

4.2 Path Dependence in Model Formation

We first present a set of reduced-form findings on how subjects' second-stage models and decisions vary by treatment. We then report structural findings on the extent to which subjects' second-stage beliefs depend on the information received in the first stage.

4.2.1 Reduced-form Evidence

To test our main hypothesis of path dependence in model formation, we first investigate second-stage models by treatment. The core feature of subjects' prediction model is the role they attribute to each variable in affecting project outcomes. As pre-registered, we thus examine how the variable subjects see in the first stage affects the final belief in the marginal impact of each predictor variable on project success. While path dependence in model formation refers to any significant difference in the final model between groups, stickiness implies a stronger reliance on the initial model. This leads to the hypothesis that subjects believe (at least weakly) in the greater marginal impact of the variable they encounter first compared to those who see the variable later. Moreover, this difference is strictly significant for at least one variable. Table 1 reports subjects' beliefs about the marginal impact of each predictor variable on the project's

[8] 72% of our subjects pass the comprehension check in the first attempt, 14% in the second attempt. 14% fail comprehension questions twice and cannot participate in our main experiment.

Table 1 Treatment effect on belief about impact of independent variables on success

	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 the treatment effect estimates from an exogenous manipulation of the first-stage independent variable on subjects' mental models of project success in the second stage. Columns (1) and (2) report the share of subjects agreeing to the statement that the predictor variables X and Z, respectively, have a *ceteris paribus* impact on the outcome, which measures subjects' models at the extensive margin. Columns (3) - (8) consider the intensive margin of model differences. Columns (3) and (4) report the differences in marginal beliefs for changes in the weak variable ΔX and the strong variable ΔZ, pooling both marginal beliefs for each subject. Columns (5) - (8) report the individual marginal belief differences of X and Z, holding constant the other variable at 0 and 1, respectively. 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.

success likelihood, using two different measures corresponding to the *extensive margin* (Columns 1 and 2) and the *intensive margin* (Columns 3 to 8), respectively. The *extensive margin* refers to whether subjects believe that a *ceteris paribus* change in a variable has any effect on the outcome. For the *intensive margin* we directly compare subjects' quantitative beliefs about the marginal effects of the variables — namely, how much a *ceteris paribus* change in a variable shifts the perceived probability of success. Columns (3) and (4) pool the beliefs in the marginal effect across the two possible values of the other variable, such that we obtain an indicator for the average belief in the marginal impact, which allows for a direct test of our hypothesis on model stickiness.[9] We also show the results from the disaggregated marginal effects that we used for our main pre-registered specification in columns (5) - (8).

The main finding is that subjects' beliefs exhibit path dependency at both the extensive and intensive margin. 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 compared to those who saw Z in the first stage. However, there are no significant differences in beliefs for a *ceteris paribus* change in Z.

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

[9]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.

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 no significant differences between groups ($p = 0.622$). The results thus show that initial exposure to a predictor variable whose importance fades when controlling for the other predictor has a lasting impact on its perceived relevance.

Turning to the intensive margin, we can observe that both groups predict a significantly higher success likelihood for projects that exhibit a high value of X (see column 3). The X-first treatment group estimates X's impact on success to be 2.6 pp higher ($p = 0.025$) 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 3). When disaggregating the pooled beliefs in the marginal effects (columns 5 to 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 about 4 pp 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.[10]

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.9, average confidence in the accuracy of subjects' prediction models decreases from about 68% in the first stage to 60% in the second stage. Crucially, there is no evidence that differentially perceived complexity drives the observed treatment effects: despite adopting different final models, the average model confidence is very similar in the second stage ($p = 0.439$).

Overall, these results align with our pre-registered hypothesis that subjects' beliefs exhibit path dependency at both margins. The identified path dependency indicates that initial models about the impact of a variable persist despite its role needing to be reconsidered after controlling for an additional predictor.

Result 1 (Sticky Models) Subjects' models are 'sticky': their final models depend on the variable observed in the first stage. Specifically, by the end of the second stage, subjects who initially observed a predictive first-stage variable (X) that later becomes conditionally unpredictive, compared to subjects who initially observed a predictive

[10] 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.

first-stage variable (Z) that remains conditionally predictive,

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

b.) believe that X's marginal effect on the outcome is higher.

4.2.2 Spillovers on Decision-making

We now examine whether the path dependency observed in subjects' beliefs extends to their decision-making, which we pre-registered as a secondary analysis. Table A.5 presents treatment differences in subjects' binary choices, while Table A.6 reports the differences in willingness-to-pay (WTP) for projects.

In the second stage, subjects face four binary decisions between pairs of projects that differ in one variable value. For each decision, we analyze the probability of choosing the project with X =1 vs. X =0 or Z =1 vs. Z =0. Under indifference, we expect random choice (50% probability), whereas a strict preference should lead to selecting the preferred project with 100% probability.

Table A.5 displays the differences in choice probabilities across project pairs. Subjects' choices are generally sensitive to the project variables, with the likelihood of choosing the project with $X = 1$ or $Z = 1$ ranging from about 50% to over 90%. Notably, subjects are, on average, 24pp more likely to select a project with $Z = 1$ over a project with $Z = 0$ (column 2), compared to selecting $X = 1$ over $X = 0$ (column 1), reflecting their higher belief in Z's conditional predictive power.

The pooled analysis of choice probabilities does not reveal a significant treatment effect (columns 1 and 2), although the signs of the *X-first* coefficient align with our belief analysis. The disaggregated estimates show that subjects who observed X first are about 7pp more likely to choose the project with the higher X value if Z = 1 compared to those who observed Z first ($p = 0.032$). This finding is consistent with the analogous effect observed in belief elicitation.

We analyze subjects' WTP for their preferred projects in Table A.6 to obtain a more detailed picture of model stickiness in decisions.

As before, columns (1) and (2) of Table A.6 pool decisions where projects differ in their X or Z variables, respectively. Columns (3)-(6) report the WTP for individual choices. As with beliefs and binary choices, subjects' WTP varies significantly across project comparisons, with variations in Z being, on average, valued at about \$3.6 more than variations in X.

Consistent with our pre-registration, results on path dependence for WTP exhibit a similar qualitative pattern to those for beliefs and binary choices but are generally

noisier. We thus find no significant treatment differences in WTP for either the pooled or disaggregated comparisons. One likely explanation is that the transmission of beliefs to decisions is attenuated by uncertainty over the optimal policy (Yang 2024, Enke et al. 2023). In support of such behavioral attenuation, we find that belief differences and WTP are only moderately correlated ($\rho = 0.48$), and scaled standard errors are larger for WTPs than for beliefs.

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. Finally, the structural approach enables a comprehensive examination of how effort and reasoning types affect model revision. We present the corresponding analyses for effort and reasoning types in Sections 4.3 and 4.4.

In this subsection, we focus on two key aspects: (1) alignment with the rational benchmark and (2) the weight subjects assign to the empirical benchmarks from their respective first stages. Notably, the empirical benchmarks from *Z-first* and *X-first* are linearly independent. We thus estimate the following regression model to identify differential reactions to the benchmarks:

$$\begin{aligned} \mu_{(x,z),i} &= \beta_0 + \beta_1 \text{BMX}_{(x,z),i} + \beta_2 \text{BMZ}_{(x,z),i} \\ &\quad + \beta_3(\text{BMX}_{(x,z),i} \times \textit{X-first}_i) + \beta_4(\text{BMZ}_{(x,z),i} \times \textit{X-first}_i) + \varepsilon_{(x,z),i}, \end{aligned} \tag{4}$$

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. The Benchmark X, BMX$_{(x,z),i}$, and the Benchmark Z, BMZ$_{(x,z),i}$, for the second stage beliefs are defined as if subjects were only focusing on that specific variable. In other words, BMX$_{(x,z),i}$ corresponds to the benchmarks of $P(Y|X = x)$ and BMZ$_{(x,z),i}$ corresponds to the benchmarks of $P(Y|Z = z)$. These benchmarks therefore serve as the rational benchmarks of the respective first stages of the two treatment groups. Since $P(Y|Z = z) = P(Y|(X, Z) = (x, z))$, the Benchmark Z also serves as the rational benchmark for the second stage.

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

Table 2 Path dependence - structural approach

	Subjective success probability (pooled), alternative specifications	
	Linear probability	Log odds
	(1)	(2)
Benchmark X	0.156***	0.156***
	(0.030)	(0.034)
Benchmark X	0.087**	0.103**
× X-first	(0.039)	(0.046)
Benchmark Z	0.500***	0.538***
	(0.021)	(0.023)
Benchmark Z	0.035	0.031
× X-first	(0.030)	(0.032)
Constant	47.348***	−0.164***
	(0.447)	(0.025)
Observations	3,136	3,093
R^2	0.363	0.342

Notes: This table analyzes second-stage model formation using structural regressions. Column 1 implements equation 4, regressing subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by the overall success frequency in the data $P_{emp}(Y)$. Column 2 presents the results for log-odds transformed variables, as specified in equation 5. 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.

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 differentially load on the benchmarks. Stickiness refers to 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.

Column 1 of Table 2 tests equation 4. Two main findings emerge. First, both treatment groups, on average, deviate from the rational benchmarks, overreacting to Benchmark X and underreacting to Benchmark Z. Second, the differences between treatment groups stem primarily from the *X-first* group loading more heavily on Benchmark X, and therefore, the information already provided in their first stage. In contrast, both groups similarly load on the 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 to Benchmark X at roughly $\beta_2 = 16\%$. Strikingly, the *X-first* group responds an additional $\beta_4 = 9$pp more, i.e., *X-first* loads over 50% more on Benchmark X than *Z-first* ($p = 0.025$). In contrast, the *Z-first* group follows Benchmark Z at about $\beta_3 = 50\%$, while the *X-first* group does so insignificantly more (54%, $p = 0.325$); hence, the weight on Z

does not differ significantly by treatment.

To further confirm that the observed differences are driven by reliance on first-stage information, we assess whether participants meaningfully update their beliefs in the first stage. The analysis of the first-stage beliefs is presented in Table A.2 in the Appendix. The regression results demonstrate that the treatment is effective, as participants' first-stage beliefs closely align with the empirical benchmarks associated with their respective first stage.

As a robustness check, Column 2 repeats the analysis using a logit transformation, i.e., regressing the log odds of both the reported beliefs and benchmarks, estimating

$$\begin{aligned} logit(\mu_{(x,z),i}) &= \gamma_0 + \gamma_1 \cdot logit(\text{BMX}_{(x,z),i}) + \gamma_2 \cdot logit(\text{BMZ}_{(x,z)}) \\ &+ \gamma_3 \cdot logit(\text{BMX}_{(x,z)}) \times \text{X-first}_i + \gamma_4 \cdot logit(\text{BMZ}_{(x,z)}) \times \text{X-first}_i + \eta_{(x,z),i}, \end{aligned} \tag{5}$$

where $logit(\theta_{x,z,i}) = ln(\frac{\theta(Y=1|X=x,Z=z)}{(1-\theta(Y=1|X=x,Z=z)})$ for $\theta \in \{P_i, \text{BMX}, \text{BMZ}\}$. We find very similar results, confirming that our specification choice does not drive the results. We use the linear probability model for the remainder of the paper for ease of interpretation.

4.3 The Role of Cognitive Effort

To assess the role of cognitive effort, we use response time data collected on every page of the experiment.[11] In the literature, response time data has frequently been used as a proxy for cognitive effort or mental deliberation (e.g., Caplin et al. 2020; Rubinstein 2016; Wilcox 1993).

As described in Section 3, we expect cognitive effort in the second stage to play a key role in driving stickiness. The less effort participants exert in the second stage, the more they rely on their first-stage beliefs, thereby amplifying group differences. Conversely, lower effort in the first stage decreases the likelihood of properly internalizing the first-stage variable's predictive importance, resulting in imprecise beliefs. Thus, how effort is allocated should strongly predict the extent of model stickiness.

Summary statistics are reported in Table A.8. On average, subjects spend 5:44 minutes in the first stage and 7:22 minutes in the second stage. Subjects in both treatment groups spend virtually the same amount of time in the first stage. The mean (median) *X-first* subject takes about 15 (36) seconds longer in the second stage than the mean (median) *Z-first* subject ($p = 0.09$).

[11] We pre-registered to measure cognitive effort based on whether participants clicked on a button to view the data table again after having seen it once at the beginning of each stage instead of using response times. However, due to a technical issue, click data was not collected on the pages where subjects report their willingness to pay for alternative projects. We therefore use response time data, which has been collected for every page of the experiment. An analysis using click data on the belief page can be found in the Appendix as Figure A.10.

Table 3 Path dependence by cognitive effort

	All	Relative Time S2 high	Relative Time S2 low	Total Time S2 high	Total Time S2 low
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.156***	0.163***	0.150***	0.151***	0.161***
	(0.030)	(0.043)	(0.042)	(0.038)	(0.046)
Benchmark X	0.087**	0.019	0.163***	0.045	0.135**
× X first	(0.039)	(0.052)	(0.059)	(0.048)	(0.062)
Benchmark Z	0.500***	0.626***	0.392***	0.607***	0.405***
	(0.021)	(0.030)	(0.027)	(0.029)	(0.029)
Benchmark Z	0.035	0.024	0.010	0.056	−0.011
× X first	(0.030)	(0.041)	(0.039)	(0.040)	(0.040)
Constant	47.348***	46.668***	48.028***	46.406***	48.290***
	(0.447)	(0.640)	(0.623)	(0.670)	(0.588)
Observations	3,136	1,568	1,568	1,568	1,568
R^2	0.363	0.476	0.260	0.460	0.272

Notes: This table analyzes the model formation in the second stage of the experiment. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2 (3) presents the analysis only for subjects that spent a larger (smaller) relative share of their time on the second stage compared to the first stage than the median participant. Column 4 (5) presents the analysis only for subjects that allocated above (below) median total time on the second stage, i.e. irrespective of the time spent in the first stage. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3 analyzes second-stage model formation by regressing subjects' reported success probabilities on the empirical benchmarks for X and Z. Column (1) uses the full sample, while Columns (2)-(3) divide the sample based on a median split of relative time spent in stage 2 compared to stage 1, reporting results separately for observations above and below the median. Similarly, Columns (4) and (5) split the sample based on total time spent in the second stage, reporting results for those above and below the median.

Four key findings emerge from the analysis. First, while both total and relative time splits yield similar results, the relative time split consistently results in slightly sharper differences between the groups. As relative time is also the more important metric in our theoretical framework, we use the relative time split from Columns (2) and (3) as a proxy of effort allocation throughout the remainder of the paper.

Second, subjects who spend relatively more time in stage two report models that are closer to the rational benchmark. These participants follow (the conditionally predictive) Benchmark Z to about 63% to 65%, which is roughly 23pp more than those who invest less effort ($p < 0.001$). Additionally, they also rely equally or less on (the conditionally un-predictive) Benchmark X. In the *X-first* group, high-effort subjects follow X

to 18% compared to 31% observed for low-effort subjects ($p = 0.010$). In contrast, the *Z-first* group shows negligible differences, with high- and low-effort subjects following X to 16% and 15%, respectively ($p = 0.824$).

Third, model stickiness is driven primarily by those exerting less cognitive effort in stage two. Among *X-first* subjects with lower relative effort (Column 3), adherence to Benchmark X is about 16pp higher than for their *Z-first* counterparts ($p = 0.006$). In contrast, among high-effort subjects (Column 2), the difference is only 2pp and not statistically significant ($p = 0.716$). We observe no significant treatment differences in adherence to Benchmark Z for either effort group.

Lastly, even subjects who invest more cognitive effort fall short of the rational benchmark, following Benchmark X to approximately $16 - 18\%$ (instead of completely disregarding X) and Benchmark Z to about $63 - 65\%$ (rather than fully incorporating Z). This may reflect other mistakes in model updating that persist across effort levels.

Result 2 (Cognitive effort and stickiness): Subjects' cognitive effort in stage two, both overall and relative to stage one predicts their model formation and stickiness. In particular,

a.) subjects that spend less time on stage two (in total or relative terms) exhibit larger stickiness and

b.) subjects that spend more time in stage two are closer to the rational model.

4.4 Reasoning and Model Formation

After showing that cognitive effort (as measured by response times) interacts strongly with our measures of model stickiness, we now examine how different reasoning approaches shape path dependence. Although lower cognitive effort is linked with increased stickiness, individuals vary not only in the effort they invest but also in the sophistication of their strategies. Differences in reasoning may affect both the required level of effort to operationalize it and the extent to which early models are maintained due to improper conditioning when revising the model. By classifying subjects' reasoning types, we can assess whether the effect of cognitive effort on stickiness persists after accounting for differences in strategy.

We begin by presenting a taxonomy of the most prominent reasoning types, then examine how these categories relate to second-stage belief formation, and finally explore the interplay between reasoning types and cognitive effort.

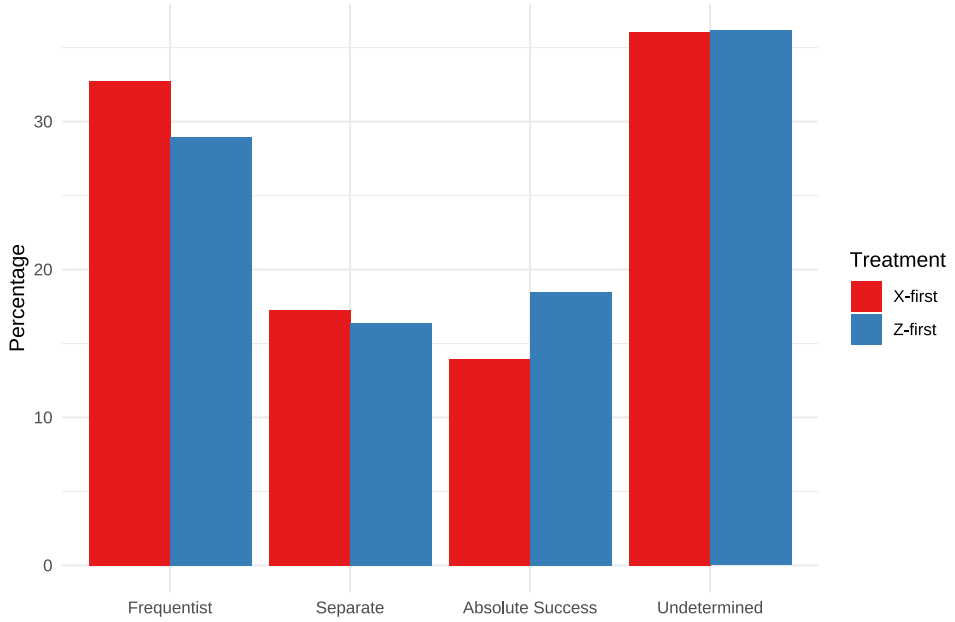


Figure 3 Reasoning Types by Treatment

4.4.1 Reasoning Types

To classify subjects' reasoning, we analyze responses to an open-text question in which subjects explained their approach to forming quantitative beliefs in the second stage.

Based on theoretical considerations (see Section 3) and pilot data, the authors first identified the most prominent types of reasoning and subsequently independently coded responses using a coding manual with mutually exclusive labeling. A response was assigned a reasoning type if at least two coders agreed. Otherwise, the response was labeled *Undetermined*. Table C.1 summarizes the coding scheme, and Figure 3 displays the distribution of reasoning types by treatment. To check the robustness of our coding approach, we employed OpenAI's ChatGPT-4o API in a zero-shot reasoning approach (Kojima et al. 2023), 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 C.

Frequentist: Subjects in this category estimate the project's success likelihood by counting the successes and failures of past projects with the same values for both independent variables and dividing the number of successes by the total count.[12] For example, one *Frequentist* explained:

"I counted each success in each combination presented. I compared that to the

[12]Note that in our instructions, we stressed that each data row represents a large number of identical observations, thereby minimizing the role of priors such that this reasoning category nests both Frequentist and Bayesian reasoning.

total number of that combination to come up with a percentage chance of success. I used those percentages to compare one project with another."

Separate: These subjects estimate success likelihoods by independently evaluating the success rates for X and Z. They then aggregate these individual effects, effectively considering $P(Y|X = x)$ and $P(Y|Z = z)$ separately instead of $P(Y|X = x, Z = z)$. One *separate* reasoner described:

"Looking at the Successes, the Blue Color was significantly more successful than the Green Color. I chose Blue options over Green options. When Clubs and Diamonds were introduced, I saw that the Diamonds were more successful than the Clubs. Therefore, I chose Diamonds over Clubs. [...]"

Absolute success: Subjects here focus on the number of successes associated with each combination of independent variable values. They condition either solely on the outcome or jointly on their first-stage variable and the outcome. This approach leads to *base-rate neglect*, as subjects fail to account for the frequency with which certain combinations appear in the data set. Their strategy is best captured by attending to $P(X, Z|Y = 1)$. For instance, one *absolute success* response stated:

"I revisited the past projects and looked at which combinations had the most success outcomes. The ones with more successful outcomes gave me more confidence in [choosing] that project"

Figure 3 shows that approximately 31% of responses were categorized as Frequentist, 17% as Separate, and 16% as Absolute Success, with the remaining 36% labeled Undetermined. Around two-thirds of our respondents can thus be assigned to one of the three most prominent reasoning types, indicating our collected data's high quality. The similar distribution of reasoning types across treatment groups ($\chi^2(3) = 3.44$, $p = 0.329$) suggests that the order of revealing variables does not significantly affect how subjects conceptualize the statistical problem.

4.4.2 Second Stage Models by Reasoning Type

We next examine how second-stage beliefs vary by reasoning type, focusing on how each group incorporates key empirical benchmarks: (i) success frequencies conditional on X only (Benchmark X), (ii) success frequencies conditional on Z only (Benchmark Z), and (iii) the demeaned total number of successes for each (X, Z) combination ("Number of successes,"), which captures attention to absolute successes. Table C.3 reports separate structural regressions for each reasoning type.

Frequentists. Frequentists rely most strongly on Benchmark Z (76%) and largely disregard Benchmark X (5%), with only a modest load on the number of successes (14%). These findings confirm that *Frequentist* reasoners most closely approximate the rational benchmark, adhering significantly more to Benchmark Z than subjects using other strategies ($p < 0.001$ for all pairwise comparisons).

Separate reasoners. This group integrates unconditional beliefs about both X and Z. They assign a weight of about 28% to Benchmark X (significantly more than any other group, $p < 0.015$ for any pairwise comparison) and 36% to Benchmark Z. Their pattern of partially weighting both benchmarks aligns with their self-reported strategy of treating X and Z as independently informative.

Absolute success reasoners. Subjects in this category attend mainly on the number of successes[13] (42%), significantly more than other groups ($p < 0.032$ for all pairwise comparisons), while placing relatively low weights on Benchmark X (9%) and Benchmark Z (22%). This pattern reflects their tendency to focus on raw success counts rather than relative frequencies.

Overall, these results confirm that our classification into Frequentist, Separate, and Absolute Success strategies captures meaningful heterogeneity in how participants incorporate information. Each group's distinct pattern of benchmark weighting explains a substantial portion of the variation in second-stage beliefs. It validates our use of self-reported reasoning to complement the quantitative measures of belief formation.

4.4.3 Path Dependence across Reasoning Type

Table C.4 reports separate regressions by reasoning type to assess whether stickiness differs by strategy. Overall, all three prominent reasoning types exhibit some degree of stickiness — a persistent influence of first-stage exposure to X. However, the extent of this effect varies.

For *Frequentist* and *Separate* subjects, the difference in loading on Benchmark X between the *X-first* and *Z-first* groups is about 8 and 6 percentage points, respectively. However, neither difference is statistically significant from zero, likely due to the smaller subsample sizes.

[13]Note that the absolute success approach does not explicitly pin down the denominator for forming beliefs. We circumvent this problem by ensuring that the number of successes is linear in the number of successes observed in the dataset for a given combination and that beliefs are centered around 50 percent to reflect the prior probability of successes. This makes the parameter estimates comparable to those of the benchmarks of X and Z.

Table 4 Path dependence across reasoning types

	Subjective success probability (pooled), by reasoning				
	All	Frequentist	Separate	Abs. success	Other
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.039	0.004	0.244***	−0.014	−0.008
	(0.032)	(0.040)	(0.088)	(0.062)	(0.062)
Benchmark X	0.087**	0.080	0.059	0.238***	0.056
× X first	(0.039)	(0.051)	(0.109)	(0.081)	(0.075)
Benchmark Z	0.383***	0.753***	0.331***	0.207***	0.191***
	(0.024)	(0.032)	(0.058)	(0.044)	(0.041)
Benchmark Z	0.035	0.015	0.046	0.035	0.005
× X first	(0.030)	(0.035)	(0.063)	(0.055)	(0.051)
Number	0.235***	0.136***	0.253***	0.417***	0.229***
of successes	(0.021)	(0.032)	(0.054)	(0.054)	(0.038)
Constant	47.348***	47.340***	47.189***	42.264***	49.711***
	(0.447)	(0.567)	(0.930)	(1.352)	(0.831)
Observations	3,136	968	528	508	1,132
R^2	0.375	0.743	0.407	0.343	0.162

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Columns 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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.

The most pronounced effect can be observed for *Absolute success* reasoners. In this group, *X-first* subjects incorporate Benchmark X to virtually the same extent as Z (a 24 percentage point difference relative to *Z-first* subjects, $p < 0.004$), indicating that their reliance on absolute counts reinforces model stickiness. This observed stickiness among the *Absolute Success* reasoning group is about three to four times larger than for *Frequentist* and *Separate* reasoners, though the differences are at most marginally significant ($p = 0.095$ and $p = 0.185$, respectively). In contrast, adherence to Benchmark Z does not differ significantly between treatments for any reasoning type.

These findings imply that while path dependence is (qualitatively) present across all three reasoning types, stickiness is strongest among subjects who only focus on the absolute number of successes in past data when forming conditional beliefs.

Result 3 (Reasoning types shape dynamic model formation): Subjects' self-reported reasoning predicts their model formation. In particular,

a.) there is substantial heterogeneity in reasoning across subjects;

b.) all reasoning types exhibit some degree of stickiness;

c.) the magnitude of stickiness varies, with *Absolute Success* reasoners showing the strongest effect.

4.4.4 Cognitive Effort, Reasoning Types, and Models

Table C.2 demonstrates a strong correlation between reasoning types and cognitive effort. More than 70% of subjects classified as *Frequentist* spend above-median time in stage two relative to stage one, suggesting that this reasoning type is associated with greater deliberation. In contrast, more than half of participants classified as *Separate* and *Absolute success* reasoners fall below the median, implying less cognitive effort.[14] Given these patterns, we investigate whether *cognitive effort* remains a significant predictor of model stickiness even after controlling for reasoning types.

Table 5 Path dependence by cognitive effort, controlling for reasoning types

	Subjective success probability (pooled), by effort		
	Baseline	Controlling for reasoning	Controlling for reasoning × treatment
	(1)	(2)	(3)
Benchmark X	0.043	0.059	0.064
× high effort	(0.063)	(0.065)	(0.068)
Benchmark X	−0.144*	−0.145*	−0.168**
×X-first× high effort	(0.079)	(0.077)	(0.082)
Benchmark Z	0.264***	0.135***	0.134***
× high effort	(0.047)	(0.045)	(0.046)
Benchmark Z	0.014	0.012	0.016
×X-first× high effort	(0.057)	(0.050)	(0.057)
Number	−0.060	−0.033	−0.033
of succ.× high effort	(0.043)	(0.044)	(0.045)
Not shown: Low effort baseline			
Addititional controls	–	Reasoning	Reasoning × X-first
Observations	3,136	3,136	3,136
R^2	0.396	0.457	0.458

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' exerted effort, controlling for a subject's reasoning type. It only reports the differences between high- and low-effort subjects, as measured by their relative time spent in stage 2. See appendix table A.13 for the full table. Column 1 reports the differences across high- and low-effort groups without controls. Column 2 repeats the analysis controlling for reasoning types, and column 3 controlling for reasoning types interacted with the treatment variable. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 5 presents a stepwise analysis: Column (2) examines the impact of cognitive effort on beliefs while controlling for reasoning type. Column (3) assesses whether cognitive effort remains a consistent driver of stickiness after accounting for differential stickiness across reasoning types.

[14]Note that we cannot disentangle the direction of causality between cognitive effort and reasoning sophistication and in principle, a simultaneous causal relationship in both directions is plausible.

As can be observed by comparing row 3 across columns, controlling for reasoning, the effect of exerting high effort on beliefs remains significant across reasoning types. On average, individuals exerting higher effort load more on Benchmark Z, though the effect size shrinks from about 26pp to about 13pp when controlling for reasoning types. This suggests that the initial effect is partially driven by *Frequentists* being overrepresented in the high-effort group.

Turning to stickiness, row 2 shows that the difference between low- and high-effort subjects becomes even more pronounced once controlling for differences across reasoning types. When controlling for reasoning-specific stickiness, high-effort subjects across all groups exhibit, on average, 16.8pp less stickiness (column 3, $p = 0.041$) than subjects exerting relatively less effort in the second stage.

Our analysis confirms that cognitive resource allocation across stages remains a primary driver of path dependence, even when accounting for reasoning types. This suggests that, although different reasoning types exhibit varying degrees of stickiness, effort allocation independently contributes to the persistence of initial models.[15]

Result 4 (Cognitive effort drives stickiness independent of reasoning type): Cognitive effort remains a robust driver of stickiness even when controlling for reasoning types. In particular, subjects who allocate less cognitive effort to stage two exhibit significantly higher stickiness compared to those who exert more effort.

5 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? In answering this question, we provide proof-of-concept evidence that models are 'sticky': many participants fail to sufficiently revise misspecified models even when correcting data becomes available.

Five design factors suggest that our empirical estimates of the effect are likely to represent a lower bound: (i) we ensure minimum attention to all potentially relevant variables, (ii) we provide all the relevant data to infer the relevant relationships throughout the revision stage, (iii) we limit the role of preference-based model revision by using minimal framing, (iv) we are transparent about the existence of the second variable from the beginning to mitigate demand effects and (v) we use elicitations that already guide people to think in the correct contingencies when revising their models. Besides

[15] Based on the regression, we still detect qualitatively different degrees of stickiness across reasoning types when controlling for cognitive effort: *Absolute Success* maintains the highest degree of stickiness. These results should be interpreted cautiously, however, as we are not sufficiently powered to establish significance.

the proposed cognitive mechanism, other biases may thus act to compound model stickiness in many economic situations.

Our insights on model stickiness and the role of cognitive effort and reasoning in explaining behavioral heterogeneity highlight the importance of better understanding individuals' reasoning processes and have implications for designing effective information provision policies. In particular, providing subjects with formerly missing data might lead them to recognize formerly overlooked factors of a model while failing to let go of conditionally irrelevant factors. In policy contexts where people should correct their mental models by removing factors (in addition to adding new ones), it may, therefore, be necessary to provide details on how to combine old and new pieces of information.

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. Further, it would be valuable to study whether the stickiness persists in the long run or to investigate under which conditions people realize that they have a misspecified model.

References

Ambuehl, Sandro and Heidi C. Thysen, "Choosing between causal interpretations: An experimental study," 2024.

Ba, Cuimin, "Robust Misspecified Models and Paradigm Shifts," 2023. arXiv:2106.12727 [econ].

_ , J. Aislinn Bohren, and Alex Imas, "Over- and Underreaction to Information," *SSRN Electronic Journal*, 2023.

Benjamin, Daniel J., "Errors in probabilistic reasoning and judgment biases," in "in" 2019.

Bohren, J Aislinn, Peter Hull, and Alex Imas, "Systemic Discrimination: Theory and Measurement," 2023.

Bordalo, Pedro, John Conlon, Nicola Gennaioli, Spencer Kwon, and Andrei Shleifer, "How People Use Statistics," 2024.

_ , 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.

Caplin, Andrew, 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.

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.

_ 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," 2023.

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.

Fréchette, Guillaume R, Emanuel Vespa, and Sevgi Yuksel, "Extracting Models From Data Sets: An Experiment," 2024.

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, "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.

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

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.

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

Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa, "Large Language Models are Zero-Shot Reasoners," 2023. arXiv:2205.11916.

Lanzani, Giacomo, "Dynamic Concern for Misspecification," 2024.

Liu, Manwei and Sili Zhang, "The Persistent Effect of Narratives: Evidence from an Online Experiment," 2024.

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

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.

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.

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," 2024.

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 contains details on our preregistration.

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

Section D includes the experimental instructions.

A Additional Tables

Table A.1 Summary statistics and balancing

Variable	ACS (2022)	All	X-first	Z-first	p-value H_0: *X-first = Z-first*
Gender					
Female	50%	50%	49%	51%	0.617
Age					
18-34	29%	43%	42%	44%	0.629
35-54	32%	44%	45%	42%	0.378
55+	38%	14%	13%	14%	0.565
Household net income					
Below 50k	34%	34%	35%	32%	0.38
50k-100k	29%	40%	38%	43%	0.116
Above 100k	37%	26%	27%	25%	0.417
Education					
Bachelor's degree or more	33%	59%	58%	60%	0.593
Region					
Northeast	17%	23%	25%	21%	0.126
Midwest	21%	20%	20%	21%	0.803
South	39%	37%	36%	37%	0.855
West	24%	19%	18%	21%	0.287
F-Stat (SUR)					0.689
Sample size	1,980,550	784	394	390	784

Notes: This table presents summary statistics for the demographics of the main experiment. It compares them to benchmark characteristics for the US adult population based on data from the American Community Survey 2022. Column 5 reports the p-value of a t-test, testing for equality between both treatment groups, as well as the statistic of a 'seemingly unrelated regressions' (SUR) F-test.

Table A.2 First stage beliefs

	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 the model formation in the first stage of the experiment. In each column, we regress subjects' subjective success probability for a project on the empirical first stage benchmark, that is, the empirical frequency of successes in the observed data conditional on the project characteristics, demeaned by the overall frequency of successes in the data. All columns pool the two beliefs each subject reports. Column 1 reports the data for the *X-first* treatment group, while column 2 reports the data for the *Z-first* treatment group. Standard errors clustered on the individual level are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.3 First stage beliefs by time spent in first stage

	All	Above median time S1	Below median time S1
	(1)	(2)	(3)
Benchmark First Stage Var.	0.717***	0.753***	0.681***
	(0.025)	(0.034)	(0.036)
Benchmark First Stage Var. $\times$ X first	0.093*	0.090	0.095
	(0.049)	(0.068)	(0.072)
Constant	51.420***	51.296***	51.543***
	(0.316)	(0.438)	(0.458)
Observations	1,568	784	784
R^2	0.532	0.569	0.495

Notes: This table analyzes the model formation in the first stage of the experiment. In each column, we regress subjects' subjective success probability for a project with either a high or a low first-stage variable on the empirical first-stage benchmark, that is, the empirical frequency of successes in the observed data conditional on the project characteristics, demeaned by overall frequency of successes in the data. All columns pool the two beliefs each subject reports. Column 1 reports the full data. Columns 2 and 3 report median splits by the total time spent on the first stage, with subjects that spend above (below) median time in column 2 (3). Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.4 Impact of learning order on subjective success probabilities

	Subjective success probability P(Y | ...), in percent			
	X=0, Z=0	X=1, Z=0	X=0, Z=1	X=1, Z=1
	(1)	(2)	(3)	(4)
X-first	−2.549*	−1.091	−1.563	2.178*
	(1.507)	(1.544)	(1.648)	(1.211)
Constant	33.028***	32.418***	57.741***	67.726***
	(1.082)	(1.126)	(1.147)	(0.853)
Observations	784	784	784	784
R^2	0.004	0.001	0.001	0.004

Notes: This table analyzes the treatment effects of the learning order on the subjective success probabilities in the second stage for projects with different features. Each column reports the success probability for one combination $(X = x, Z = z)$ of the two binary independent variables. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.5 Impact of learning order 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)
Observed 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 examines the treatment effects of learning order on choice probabilities in the second stage. The choice probability represents the proportion of subjects who selected one project over another. Columns 3–6 report choices between two projects that differ in exactly one of the two independent variables, while columns 1–2 present averages for choices that vary in either X or Z, respectively. 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 Impact of learning order on project valuation

	Willingness-to-pay (WTP) for preferred project					
	Intensive margin		Intensive margin - disaggregated			
	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)
Observed	0.146	0.142	−0.322	0.614	0.045	0.238
X first	(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 examines the treatment effects of learning order on willingness to pay (WTP) in the second stage. The WTP reflects the amount subjects are willing to pay for their chosen project. Columns 3–6 report WTP differences between two projects that vary in exactly one of the two independent variables, while columns 1–2 present averages for choices that differ in either X or Z, respectively. 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_2	0	No incorporation of Benchmark X
β_3	1	Full incorporation of Benchmark Z
β_4	0	No differential effect on X incorporation between treatment groups
β_5	0	No differential effect on Z incorporation between treatment groups

Notes: The rational benchmark assumes a frequentist approach that perfectly reflects 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 Summary statistics on time spent across stages

Variable	All	X-first	Z-first	p-value (KS) H_0: *X-first = Z-first*
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	784

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 treatment group. All columns report the mean, with the median in parentheses. Row 3 reports the average share allocated to stage 2 relative to time spent in both stages. Row 4 reports the share of subjects that spent above-median 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.9 Summary statistics on confidence across stages

Metric	**All**	**X-first**	**Z-first**	**p-value (KS)** H_0: *X-first = Z-first*
Confidence S1	67.82%	66.72%	68.93%	0.261
Confidence S2	60.22%	60.92%	59.52%	0.439
Above Median Confidence S1	56%	55%	58%	
Above Median Confidence S2	51%	52%	50%	
Sample size	784	394	390	

Notes: This table reports descriptive statistics on the self-reported confidence of subjects in the first and second stage of the experiment in percent by the assigned treatment group. All columns report the mean level. Row 3 reports share of subjects with spent above-median confidence in the first stage. Row 4 reports share of subjects with spent 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.10 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 analyzes the model formation in the second stage of the experiment. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2 presents the analysis for subjects who revisited the data table on the belief-elicitation page in the second stage, while Column 3 focuses on those who did not. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.11 Summary statistics on time spent across stages by reasoning type

Variable	**All**	**Frequentist**	**Separate**	**Absolute**	**Other**
Time spent S1	05:44	06:09	04:57	04:59	06:06
	(04:44)	(05:15)	(04:09)	(04:03)	(04:56)
Time spent S2	07:22	09:50	05:44	06:21	06:29
	(06:10)	(08:51)	(04:19)	(05:08)	(05:05)
Relative time S2/(S1+S2)	0.55	0.61	0.52	0.55	0.52
Above med. rel. time S2	50%	73%	36%	48%	38%
Sample size	784	242	132	127	283

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. Rows 1 and 2 report the mean, with the median in parentheses. Row 3 reports the average share allocated to stage 2 relative to time spent in both stages. Row 4 reports the share of subjects that spent above-median time in the second stage.

Table A.12 Belief about impact of independent variables across reasoning types

	Subjective success probability (pooled), by reasoning				
	All	Frequentist	Separate	Abs. success	Other
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.082***	0.047*	0.275***	0.090**	0.020
	(0.021)	(0.027)	(0.061)	(0.045)	(0.041)
Benchmark Z	0.401***	0.761***	0.355***	0.222***	0.194***
	(0.019)	(0.027)	(0.043)	(0.038)	(0.032)
Number of successes	0.235***	0.136***	0.253***	0.417***	0.229***
	(0.021)	(0.032)	(0.054)	(0.054)	(0.038)
Constant	47.348***	47.340***	47.189***	42.264***	49.711***
	(0.447)	(0.567)	(0.929)	(1.350)	(0.831)
Observations	3,136	968	528	508	1,132
R^2	0.374	0.742	0.406	0.338	0.162

Notes: This table analyzes the model formation in the second stage of the experiment. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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.13 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)
Benchmark X	0.018	0.008	0.007
	(0.044)	(0.044)	(0.044)
Benchmark X	0.163***	0.168***	0.180***
× X-first	(0.059)	(0.059)	(0.058)
Benchmark Z	0.260***	0.327***	0.328***
	(0.033)	(0.031)	(0.031)
Benchmark Z	0.010	0.005	0.003
× X-first	(0.039)	(0.036)	(0.037)
Number	0.265***	0.251***	0.251***
of successes	(0.033)	(0.033)	(0.033)
Benchmark X	0.043	0.059	0.064
× high effort	(0.063)	(0.065)	(0.068)
Benchmark X	−0.144*	−0.145*	−0.168**
×X-first× high effort	(0.079)	(0.077)	(0.082)
Benchmark Z	0.264***	0.135***	0.134***
× high effort	(0.047)	(0.045)	(0.046)
Benchmark Z	0.014	0.012	0.016
×X-first× high effort	(0.057)	(0.050)	(0.057)
Number	−0.060	−0.033	−0.033
of succ.× high effort	(0.043)	(0.044)	(0.045)
X-first × high effort	0.221	0.283	0.283
	(1.289)	(1.252)	(1.253)
high effort	−1.479	−1.284	−1.284
	(1.149)	(1.205)	(1.206)
Constant	48.028***	50.139***	50.139***
	(0.623)	(0.882)	(0.883)
Controls	–	Reasoning	Reasoning × X-first
Observations	3,136	3,136	3,136
R^2	0.396	0.457	0.458

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type and attention allocation. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. Rows 7 through 11 present interactions with an indicator variable, taking the value 1 for each subject that spent relatively more time on S2 than the median respondent. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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.14 Path dependence by reasoning type and cognitive effort

	All	Frequentist	Separate	Abs. success	Other
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.018	0.110	0.098	−0.007	−0.047
	(0.044)	(0.076)	(0.110)	(0.083)	(0.074)
Benchmark Z	0.260***	0.618***	0.286***	0.230***	0.123**
	(0.033)	(0.069)	(0.075)	(0.061)	(0.048)
Benchmark X	0.163***	0.110	0.151	0.325**	0.152
× X-first	(0.059)	(0.107)	(0.142)	(0.127)	(0.094)
Benchmark Z	0.010	−0.016	0.007	0.028	0.021
× X-first	(0.039)	(0.085)	(0.080)	(0.084)	(0.055)
Number	0.265***	0.190**	0.297***	0.392***	0.230***
of successes	(0.033)	(0.074)	(0.074)	(0.080)	(0.050)
Benchmark X	0.043	−0.152*	0.392**	−0.011	0.110
× high effort	(0.063)	(0.088)	(0.175)	(0.123)	(0.134)
Benchmark Z	0.264***	0.191**	0.122	−0.056	0.194**
× high effort	(0.047)	(0.076)	(0.117)	(0.089)	(0.086)
Benchmark X	−0.144*	−0.026	−0.241	−0.147	−0.252
×X-first× high effort	(0.079)	(0.120)	(0.211)	(0.163)	(0.156)
Benchmark Z	0.014	0.031	0.115	0.023	−0.062
×X-first× high effort	(0.057)	(0.091)	(0.126)	(0.114)	(0.109)
Number	−0.060	−0.074	−0.121	0.053	−0.003
of succ.× high effort	(0.043)	(0.082)	(0.104)	(0.109)	(0.075)
Constant	47.348***	47.340***	47.189***	42.264***	49.711***
	(0.447)	(0.569)	(0.935)	(1.359)	(0.833)
Observations	3,136	968	528	508	1,132
R^2	0.395	0.751	0.418	0.344	0.173

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type and attention allocation. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. Rows 7 through 11 present interactions with an indicator variable, taking the value 1 for each subject that spent relatively more time on S2 than the median respondent. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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.

B Research Transparency

Preregistration We preregistered our surveys and experiments at AsPredicted.org, registration number #174937. The preregistration includes details on the survey design, survey instructions, sampling process, planned sample size, exclusion criteria, and research questions. The pre-registration can be accessed here: `https://aspredicted.org/xqfv-9s2d.pdf`

Conflicting interests We declare that we have no conflicting interests.

C Empirical Approach

Table C.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 estimate 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.

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

Variable	All	Frequentist	Separate	Absolute	Other
Time spent S1	05:44	06:30	04:58	05:13	05:37
	(04:44)	(05:21)	(04:13)	(04:03)	(04:39)
Time spent S2	07:22	09:49	05:41	06:05	06:10
	(06:10)	(08:33)	(04:26)	(04:41)	(05:06)
Relative time S2/(S1+S2)	0.55	0.6	0.52	0.53	0.53
Above med. rel. time S2	50%	68%	38%	41%	40%
Sample size	784	285	182	87	230

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. Rows 1 and 2 report the mean, with the median in parentheses. Row 3 reports the average share allocated to stage 2 relative to time spent in both stages. Row 4 reports the share of subjects that spent above-median time in the second stage.

Table C.3 Belief about impact of independent variables across reasoning types (AI Codes)

	Subjective success probability (pooled), by reasoning				
	All	Frequentist	Separate	Abs. success	Other
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.082***	0.017	0.219***	0.101*	0.048
	(0.021)	(0.027)	(0.050)	(0.060)	(0.044)
Benchmark Z	0.401***	0.615***	0.415***	0.181***	0.206***
	(0.019)	(0.030)	(0.037)	(0.048)	(0.035)
Number of successes	0.235***	0.226***	0.248***	0.392***	0.176***
	(0.021)	(0.033)	(0.040)	(0.071)	(0.043)
Constant	47.348***	45.361***	46.769***	45.072***	51.129***
	(0.447)	(0.706)	(0.775)	(1.370)	(0.910)
Observations	3,136	1,140	728	348	920
R^2	0.374	0.599	0.446	0.310	0.148

Notes: This table analyzes the model formation in the second stage of the experiment. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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 C.4 Path dependence across reasoning types (AI Codes)

	Subjective success probability (pooled), by reasoning				
	All	Frequentist	Separate	Abs. success	Other
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.039	−0.032	0.168**	−0.015	0.046
	(0.032)	(0.037)	(0.075)	(0.088)	(0.067)
Benchmark X	0.087**	0.095*	0.100	0.225**	0.005
× X first	(0.039)	(0.049)	(0.093)	(0.110)	(0.080)
Benchmark Z	0.383***	0.573***	0.449***	0.205***	0.182***
	(0.024)	(0.038)	(0.048)	(0.062)	(0.043)
Benchmark Z	0.035	0.082**	−0.066	−0.047	0.052
× X first	(0.030)	(0.039)	(0.056)	(0.070)	(0.056)
Number	0.235***	0.226***	0.248***	0.392***	0.176***
of successes	(0.021)	(0.033)	(0.040)	(0.072)	(0.043)
Constant	47.348***	45.361***	46.769***	45.072***	51.129***
	(0.447)	(0.707)	(0.776)	(1.374)	(0.911)
Observations	3,136	1,140	728	348	920
R^2	0.375	0.602	0.448	0.316	0.149

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Columns 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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 C.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)
Benchmark X	0.018	0.009	0.005
	(0.044)	(0.044)	(0.043)
Benchmark X	0.163***	0.150**	0.158***
× X-first	(0.059)	(0.059)	(0.058)
Benchmark Z	0.260***	0.301***	0.297***
	(0.033)	(0.033)	(0.033)
Benchmark Z	0.010	−0.002	0.009
× X-first	(0.039)	(0.037)	(0.038)
Number	0.265***	0.266***	0.266***
of successes	(0.033)	(0.034)	(0.034)
Benchmark X	0.043	0.064	0.071
× high effort	(0.063)	(0.064)	(0.066)
Benchmark X	−0.144*	−0.128*	−0.143*
×X-first× high effort	(0.079)	(0.077)	(0.080)
Benchmark Z	0.264***	0.185***	0.195***
× high effort	(0.047)	(0.047)	(0.048)
Benchmark Z	0.014	0.030	0.007
×X-first× high effort	(0.057)	(0.052)	(0.056)
Number	−0.060	−0.062	−0.062
of succ.× high effort	(0.043)	(0.044)	(0.045)
X-first × high effort	0.221	0.268	0.268
	(1.289)	(1.286)	(1.287)
high effort	−1.479	−0.732	−0.732
	(1.149)	(1.193)	(1.194)
Constant	48.028***	51.364***	51.364***
	(0.623)	(0.958)	(0.959)
Controls	–	Reasoning	Reasoning × X-first
Observations	3,136	3,136	3,136
R^2	0.396	0.440	0.442

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type and attention allocation. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. Rows 7 through 11 present interactions with an indicator variable, taking the value 1 for each subject that spent relatively more time on S2 than the median respondent. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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 C.6 Path dependence by reasoning type and cognitive effort (AI Codes)

	All	Frequentist	Separate	Abs. success	Other
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.018	−0.025	0.065	0.037	0.009
	(0.044)	(0.063)	(0.090)	(0.126)	(0.082)
Benchmark Z	0.260***	0.340***	0.389***	0.199**	0.147***
	(0.033)	(0.070)	(0.066)	(0.085)	(0.050)
Benchmark X	0.163***	0.187**	0.206*	0.180	0.077
× X-first	(0.059)	(0.093)	(0.119)	(0.164)	(0.105)
Benchmark Z	0.010	0.116	−0.097	−0.093	0.031
× X-first	(0.039)	(0.073)	(0.074)	(0.078)	(0.067)
Number	0.265***	0.325***	0.284***	0.359***	0.175***
of successes	(0.033)	(0.069)	(0.056)	(0.107)	(0.057)
Benchmark X	0.043	−0.016	0.259*	−0.146	0.106
× high effort	(0.063)	(0.077)	(0.156)	(0.155)	(0.143)
Benchmark Z	0.264***	0.361***	0.149	0.028	0.099
× high effort	(0.047)	(0.081)	(0.093)	(0.124)	(0.094)
Benchmark X	−0.144*	−0.118	−0.268	0.126	−0.184
×X-first× high effort	(0.079)	(0.108)	(0.187)	(0.206)	(0.164)
Benchmark Z	0.014	−0.081	0.096	0.074	0.021
×X-first× high effort	(0.057)	(0.084)	(0.111)	(0.148)	(0.119)
Number	−0.060	−0.145*	−0.094	0.081	0.002
of succ.× high effort	(0.043)	(0.077)	(0.076)	(0.137)	(0.088)
Constant	47.348***	45.361***	46.769***	45.072***	51.129***
	(0.447)	(0.708)	(0.778)	(1.384)	(0.914)
Observations	3,136	1,140	728	348	920
R^2	0.395	0.619	0.459	0.322	0.154

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type and attention allocation. In each column, we regress subjects' subjective success probability for a project with variables $(X = x, Z = z)$ on the empirical benchmark for X and Z, that is the empirical frequency of successes in the observed data conditional on the project's variable $P_{emp}(Y|X = x)$ and $P_{emp}(Y|Z = z)$, demeaned by overall frequency of successes in the data $P_{emp}(Y)$. Rows 7 through 11 present interactions with an indicator variable, taking the value 1 for each subject that spent relatively more time on S2 than the median respondent. All columns pool the four subjective probabilities each subject reports. Column 1 reports the full data. Column 2-5 present the analysis for subjects based on their reasoning type. The "Other" category (column 5) comprises 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.

D Experimental Instructions

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).

[Table with 6 example projects]

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 feature can only take on two possible values (e.g. and), and the outcome is Success or Failure.

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

You have to answer all comprehension questions correctly in order to receive your payment.

According to the example data, which statement is correct?

- For the example data, approximately 33% of the projects with symbol are successful.
- 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.

Please select the correct statement.

- I can use the information about past projects to learn about the impact of features on project outcomes.
- I cannot 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 a single past project.
- In the past data, each row represents one thousand identical past project.
- In the past data, each row represents several past projects but I cannot tell the exact number.

Please select the correct statement.

- 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.
- I cannot obtain a bonus payment in this study.

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]

Stage 1: Choice Elicitation

Your 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:

- You choose the project you prefer to undertake.
- You indicate how much you prefer the chosen project.

Note for your bonus:

Both tasks are equally likely to determine your bonus.

Your preferred project: F1(♣) vs. F2(♦)

[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.

[Table: N° Card Outcome
F1 ♣ ?
F2 ♦ ?]

Click here to learn more about the bonus [If you are eligible for a bonus, you will receive the payoffs associated with the realized outcome of the selected project. The outcome will be drawn according to the true relationship between features and outcome. The project will pay you $10 if it is a success and will pay you $0 if it is a failure.]

Please select the project you prefer to undertake. Note: If you are indifferent you can select either of the two projects.

- F1 (♣)
- F2 (♦)

Project F2(♦) vs. F1(♣) [if previous choice = F2]

[Revisit past projects]

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

[If you are eligible for a bonus, the computer will randomly select a row and your choice in that row will determine your reward. The project you choose will pay you $10 if it is a success and will pay you $0 if it is a failure. If, for the selected row, you have chosen your less preferred project, you will receive the additional payment indicated in that row.]

[MPL: 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

[Dynamic Text: Based on your chosen row, you value project F2(♦) at least as much as project F1(♣) plus $X, but no more than project F1(♣) plus $(X+0.5).]

Stage 1: Model Elicitation (intensive margin)

Your assessment

[Revisit past projects]

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

[Table: N° Card Outcome

F3 ♣ ?
F4 ♦ ?

Click here to learn more about the bonus

[If this question is chosen to determine your bonus payment, we use the following formula to compute your payment: Probability of winning $10 (in percent) = 100 - 1/100(Estimate (in percent) - Truth)ˆ2, where Truth = 100 if the selected project is a Success, and 0 if it is a Failure. The outcome of the project will be drawn according to its true probability of success, which is determined by the feature of the project.]

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

- Slider from 0% (Never Successful) to 100% (Always successful)

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

- Slider from 0% (Never Successful) to 100% (Always successful)

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

- Slider from 0% (Not at all certain) to 100% (Always certain)

Stage 2

Part 2 (Example: *Z-first*)

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.

Part 2: 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]

Stage 2: Choice Elicitation

Your 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:

- You choose the project you prefer to undertake.
- You indicate how much you prefer the chosen project.

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

Your preferred project: P1(♦,🔵) vs. P2(♣,🔵) [Example]

[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.

[Table: N° Card Color Outcome
P1 ♦ 🔵 ?
P2 ♣ 🔵 ?]

Click here to learn more about the bonus [as in Stage 1]

Please select the project you prefer to undertake. Note: If you are indifferent you can select either of the two projects.

- P1(♦,🔵)
- P1(♣,🔵)

Your preferred project: P1(♦,●) vs. P2(♣,●) [Example]

[Revisit past projects]

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 [as in Stage 1]

[MPL: 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]

[Dynamic Text: Based on your chosen row, you value project P1(♦,●) at least as much as project P2(♣,●) plus $X, but no more than project P2(♣,●) plus $(X+0.5).]

Stage 2: Model Elicitation (intensive margin)

Your assessment

[Revisit past projects]

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

[Table: N° Card Color Outcome
P9 ♣ ● ?
P10 ♣ ● ?
P11 ♦ ● ?
P12 ♦ ● ?]

Click here to learn more about the bonus [as in Stage 1]

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

- Slider from 0% (Never Successful) to 100% (Always successful)

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

- Slider from 0% (Never Successful) to 100% (Always successful)

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

- Slider from 0% (Never Successful) to 100% (Always successful)

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

- Slider from 0% (Never Successful) to 100% (Always successful)

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

- Slider from 0% (Not at all certain) to 100% (Extremely certain)

Stage 2: Reasoning Elicitation

Important

You have now finished the main part of this survey.

On the next page, you will encounter an open question in which we will ask you to explain how you determined the success likelihood of a project.

From our experience, it can take about 2 minutes to complete this question.

Your responses are very valuable for this research project. Therefore, please take your time to respond carefully.

Your explanation

[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.

[Free form text box]

Stage 2: Model Elicitation (extensive margin)

Your model

[Revisit past projects]

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?

- Slider from 0% (Not at all certain) to 100% (Extremely certain)

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?

- Slider from 0% (Not at all certain) to 100% (Extremely certain)