Fragile Learning From Others*

Alisher Batmanov†
UC San Diego

Click here for the latest version

August 2, 2026

Abstract

Behavioral biases in decision-making are widespread and often persist even in the presence of feedback diagnostic of optimal behavior. This paper examines whether exposure to others' decisions can correct initial misconceptions and facilitate learning in an environment where departures from the theoretical benchmark arise from neglecting an informative signal in a worker hiring task. Using a laboratory experiment, I show that exposure to optimal behavior substantially improves decision quality. The analysis of the underlying mechanisms suggests that this improvement is not driven by mechanical imitation alone, as subjects respond asymmetrically to the quality of observed choices, nor entirely by the richer feedback environment that social exposure generates. I then evaluate the retention and transfer of these learning gains beyond the period of exposure and find that, for most subjects, the improvements are fragile. Comparing the effects of social exposure to those of explicit guidance further underscores the limits of observational learning as a policy tool. These findings highlight the dual role of social learning: while it can enhance decision-making, it can also generate imitation behavior that fails to generalize beyond the observed context.

Keywords: Learning, Exposure, Imitation, Retention, Transfer
JEL classification: C91, D83, D91, J71

*I am indebted to Emanuel Vespa and Isabel Treviño for invaluable guidance and support on this project. I thank Marina Agranov, Daniel Gottlieb, Aram Grigoryan, Ulrike Malmendier, Kirby Nielsen, Ryan Oprea, Frank Schilbach, Marta Serra-Garcia, Denis Shishkin, Joel Sobel, Lise Vesterlund, and Sevgi Yuksel for valuable conversations and constructive comments, as well as the audiences at Caltech CTESS Summer Program in Theory-Driven Experiments, Economic Science Association (ESA) North American Meeting in Columbus and World Meeting in Los Angeles, Behavioral & Experimental Economics Student Conference at Caltech, Los Angeles Experiments (LAX) Workshop at Caltech, American Economic Association Mentoring Conference in Chicago, Southwest Economic Theory Conference (SWET) at Loyola Marymount University, and UC San Diego Theory-Behavioral-Experimental Graduate Seminar. I am immensely grateful to undergraduate student members of my economic research lab at UC San Diego for outstanding research assistance and to all study participants for their time and patience. This research was funded by UC San Diego Economics Department research grant and the Institute for Humane Studies fellowship, and was approved by UC San Diego IRB under protocol #809339. This study was pre-registered under AEARCTR-0013607 on the AEA RCT Registry. All errors are my own.

†University of California San Diego, Department of Economics. Contact: abatmanov@ucsd.edu

1 Introduction

A large body of work in behavioral economics has shown that, due to cognitive difficulties in integrating important features of the environment into the decision-making process, agents often rely on simple yet incorrect mental models.[1] While one might expect that experience with the environment together with feedback informative of optimal behavior would correct such biases, initial misconceptions often inhibit learning by influencing how agents engage with this feedback and making them reluctant to adjust their behavior.[2] Even in the absence of explicit information-processing costs, other behavioral factors may hinder learning, for instance by inducing excessive confidence in initial responses. Although the literature has proposed several mechanisms that impede individual learning from feedback, relatively less is known about whether mere exposure to others' decisions can mitigate these limitations.[3] This raises the question of whether a different channel of social learning, namely observing others' actions, can improve behavior, and if so, whether the resulting gains reflect genuine understanding or merely context-specific imitation.

In this paper, I study three questions: (i) whether social learning through exposure to other agents' actions can facilitate optimal behavior in a statistical inference task; (ii) whether any resulting gains reflect mechanical imitation of what is observed or a selective response to it; and (iii) whether learning persists beyond the period of exposure, reflecting a deeper understanding of the optimal decision rule. I design a laboratory experiment with three key features. First, I consider a setting in which neglecting an informative signal leads to suboptimal decisions. In the control condition, participants complete this task independently for 80 rounds under fixed primitives (and 120 rounds in total), with transparent, immediate feedback. This provides a clear reference environment with substantial scope for learning. Second, I compare behavior in the control group to treatment conditions in which participants are additionally exposed to another participant's choices. By exogenously varying the quality of the observed choices, I can assess how

[1] For example, incorrect understanding of the environment can manifest itself in ignoring described correlations when making decisions (e.g., Enke & Zimmermann (2019); Eyster & Weizsacker (2012)), behaving as if one's vote is pivotal even when it is not (e.g., Esponda & Vespa (2014, 2024)), acting in a second-price auction as if it were a first-price auction (e.g., Cason & Plott (2014)), or having difficulty understanding sample selection (e.g., Araujo, Wang & Wilson (2021); Enke (2020); Esponda & Vespa (2018)).

[2] A growing theoretical and empirical literature on learning under misspecification studies the implications of adopting incorrect mental models (e.g., Esponda & Pouzo (2016); Hanna, Mullainathan & Schwartzstein (2014); Spiegler (2020)). Recent evidence in Esponda, Vespa & Yuksel (2024) demonstrates the persistence of suboptimal behavior even in the presence of feedback, suggesting that providing participants with feedback and experience may not always be sufficient to correct initial misconceptions.

[3] Recent evidence suggests that people systematically discount information obtained from others even when it is perfectly communicated (Conlon et al., 2026), raising the question of whether observing others' actions faces similar barriers.

exposure to optimal versus suboptimal behavior affects subjects' own decisions.[4] Third, I examine the robustness of learning gains by testing whether behavior persists once exposure ends and then when the primitives of the task change. This design allows me to study both the retention and the transfer of learning.

The experiment is built around a repeated inference task in which participants act as hiring managers choosing between two workers drawn from distinct groups. In each round, one worker is randomly selected from each group, and participants observe a noisy education signal that is informative about the worker's ability, while the underlying group-level ability distributions are known. The theoretical benchmark implies that, conditional on observing the same education signal for both workers, the optimal choice depends on how the signal differentially screens ability across groups. In particular, although one group has higher average ability ex ante, the education signal eliminates low-ability workers and favors the other group ex post, making it optimal to hire from that group. A common deviation from this benchmark arises when participants neglect the informational content of the signal and instead rely on prior group averages, leading them to hire suboptimally. After each decision, participants receive immediate feedback about the realized ability of the hired worker, allowing them to learn from experience over repeated rounds.

At baseline, when all participants complete the task independently, the incidence of optimal behavior is limited: even after substantial experience with the task and repeated opportunity to learn through feedback, only about 40% of choices are optimal.[5] In the subsequent social learning phase, the experiment introduces two main treatment conditions that vary the quality of social exposure. Subjects in Exposure observe choices made by another participant who consistently selects the optimal option, while subjects in NoExposure continue to make decisions independently with no social information. In both conditions, participants receive the same individual feedback after each round; the only difference is whether they additionally observe another participant's choice before making their own decision. Importantly, observed choices are not labeled as optimal, so participants must infer their quality from experience. By the end of this part, exposure to optimal decisions substantially improves performance. In the final round, the share of optimal choices in Exposure reaches 67%, compared to 46% in NoExposure (62% versus 49% over the last ten key rounds), and the difference is statistically significant.[6] These results show that observing optimal behavior

[4] Subjects are not explicitly informed about the optimality of the observed actions, nor are they instructed to use them when making their decisions. As a result, recognizing whether observed behavior is optimal is an integral part of the design.

[5] Throughout the paper, when assessing performance in each part of the experiment, I focus on the optimality rate in the last ten key rounds of each part. Thus, 40% means that on average subjects made 4.0 out of 10 optimal choices by the end of the baseline phase (Part 1).

[6] Statistical significance at the median is established using quantile regression. The difference is also significant

facilitates learning beyond what is achieved through feedback alone. Moreover, the improvement in Exposure reflects a broad shift in behavior: the distribution of optimal choice shares across subjects first-order stochastically dominates that in NoExposure.[7]

A natural question about the mechanism behind the main treatment effect is whether it is driven by mechanical imitation of observed choices. A diagnostic treatment, ExposureNeg, helps address this question. If subjects simply replicate observed behavior regardless of its content, then exposure to suboptimal decisions should produce a decline in performance comparable in magnitude to the improvement generated by exposure to optimal decisions. The data reject this prediction.[8] By the end of the exposure phase, the share of optimal choices in ExposureNeg is virtually identical to that of NoExposure, and the difference is not statistically significant. The effects of exposure are thus notably asymmetric, suggesting that the quality of observed choices matters. To organize the interpretation of these findings, I consider three candidate learning channels with distinct observable predictions — mechanical imitation, selective adoption, and model revision — and the asymmetry rules out the first. Direct evidence supports the second: examining transition behavior between consecutive rounds, I find that subjects in Exposure are over 80% more likely to switch to the optimal choice after their own suboptimal choice yields a low payoff than after a high one — a low payoff from one's own approach opens the window in which the observed alternative is tried, and favorable payoff realizations sustain it once adopted.[9] This evidence suggests that selective adoption is at play: subjects do not blindly copy observed choices but instead test them against their own payoff experience.

The selective adoption interpretation naturally raises a second question about mechanism: whether the treatment effects in Exposure could be partly driven by the richer feedback that accompanies optimal decisions, rather than by the social component of observing another participant's choices. A diagnostic treatment, Exposure (Full Info), addresses this by providing subjects with counterfactual payoff information on top of social exposure, thereby eliminating any informational advantage associated with observing optimal choices. Performance in this group is similar to that in Exposure; while the confidence interval on this comparison cannot rule out moderate effects of the added feedback, the evidence indicates that enriched feedback is not the

when comparing the fraction of subjects in each treatment who make eight or more optimal choices out of ten, a binary indicator of predominantly optimal behavior.

[7]In particular, Exposure includes a substantially larger share of subjects who choose optimally in all ten rounds and a smaller share of subjects who never choose optimally, relative to NoExposure.

[8]Detailed results for the ExposureNeg treatment are reported in Section 3.3.

[9]This pattern is consistent with a reinforcement-learning mechanism in which adoption of observed behavior is governed by the subject's own payoff experience. A simple Q-learning model, in which subjects update the value of each choice based on realized payoffs, captures the qualitative features of this transition behavior well and is presented as an illustration of the parallel between the observed choice dynamics and a numerical learning model (Appendix F).

primary driver of the main treatment effects (Appendix C.1). Further analysis of heterogeneity reveals that while the beneficial effects of exposure to optimal choices are stable across all levels of baseline confidence, the response to suboptimal choices is sharply moderated by baseline confidence, with point estimates indicating harm concentrated among the least confident subjects, whereas more confident subjects appear insulated (Appendix G). This pattern is consistent with the broader finding that learning from exposure is not a uniform process: its effectiveness depends on the interaction between the quality of observed behavior, the payoff signals it generates, and the confidence of the observer in her own approach.

To capture heterogeneity in learning gains from observing optimal decisions, I classify Exposure participants based on how their behavior changes between the individual learning and social exposure parts. Approximately 25% of subjects switch from predominantly suboptimal choices at baseline to predominantly optimal choices under exposure and are classified as *compliers*. Another 30% make optimal choices both before and after exposure (*always-takers*), while about 42% continue to choose suboptimally despite observing optimal decisions (*never-takers*). Because the classification conditions on realized outcomes within the treated group, it describes behavioral trajectories rather than causally attributing each individual crossing to exposure: a comparable trajectory occurs for 16% of control-group subjects through individual learning alone, against 25% under exposure. The compliers nevertheless represent the margin at which social exposure has the greatest potential to improve decisions and provide a natural focus for assessing the robustness of learning beyond the exposure phase.

Before turning to the stability of learning gains, it is useful to clarify what learning from observing others' decisions may represent in this setting. Improvements in performance following positive exposure may conceal several distinct learning patterns. One possibility is *pure imitation-based learning*, in which subjects follow observed choices without forming an explicit understanding of why those choices are optimal. A related possibility is *rule-based learning* that remains tightly linked to the original environment and leads subjects to choose optimally only as long as the task remains unchanged. Finally, exposure may induce a *deeper form of learning* in which subjects internalize the optimal decision rule in a way that transfers to a related task.[10] These distinct learning modes can generate similar behavior during exposure but imply different predictions once exposure ends or the environment changes, motivating a closer assessment of learning beyond the exposure phase.

[10]The sequential structure of the experiment is designed to distinguish among these learning modes. Removing social exposure while holding the environment fixed tests for pure imitation; changing the task primitives tests for rule-based learning that does not generalize; and introducing explicit advice tests whether guidance can reach subjects whom observation alone does not affect.

To assess whether learning gains persist once social exposure ends, the next part of the experiment returns subjects to independent decision-making while holding the task environment fixed. In this phase, participants face the same inference problem as at baseline but no longer observe others' choices, allowing learning driven by exposure to be separated from contemporaneous imitation. The results indicate that learning gains are only partially retained. Among Exposure compliers, nearly one-third revert to choosing the dominated option once exposure is removed, suggesting that a substantial share of the improvements observed during social learning reflects fragile behavior rather than durable understanding. Evidence on the intensive margin points to a similar pattern: although compliers reach an average optimality rate of 93% by the end of the exposure phase, this rate declines to roughly 82% once exposure is removed. Despite this drop, compliers continue to outperform the NoExposure group, whose average optimality rate is around 53%, but remain well below the always-takers group, whose performance stays extremely high at 97%.

To further examine the nature of learning induced by exposure, the analysis turns to behavior under a change in the environment's underlying primitives. The objective of this part is to assess whether learning under exposure reflects a genuine understanding of the optimal decision rule or instead relies on task-specific imitation. To this end, subjects continue to make decisions independently, but face an inference problem with different primitives.[11] Learning gains among Exposure compliers decrease sharply. Only about 41% of initial compliers continue to exhibit optimal behavior under the modified environment primitives, indicating limited transfer of earlier improvements to the new setting. Evidence on the intensive margin points to an even larger decline: the average optimality rate among compliers falls to roughly 60%, down from 82% under the original task in the previous part. Notably, this level of performance is slightly below that of the NoExposure group, whose average optimality rate is approximately 61% despite never having observed optimal decisions. This pattern is unlikely to be driven by an overall increase in task difficulty under the new primitives, as always-takers continue to perform at a relatively high level (89% optimality rate), while never-takers exhibit an improvement in performance.

Finally, the last part of the experiment introduces explicit written advice on how to approach the task, provided by the same participant whose choices subjects observed earlier. Under optimal advice, performance in Exposure increases sharply. Among compliers, the average optimal choice rate returns to the originally high level reaching 88% and exceeding NoExposure. Never-takers in Exposure, who showed little response to exposure alone, also improve substantially: their optimal choice rate increases to 70%, approaching that of NoExposure. As a result, the Exposure group

[11] To be precise, the underlying inference task is the same worker hiring setting, with only slight changes in parameters.

as a whole outperforms the control group under advice. In contrast, explicitly providing suboptimal advice in ExposureNeg does not generate a comparable decline in performance: optimal choice rates remain at 62%, while the NoExposure group continues to improve, leading to a gap of 12 percentage points that is marginally significant.

The results of this paper point to a distinction between learning *what* to do and understanding *why* it works, with implications for the design of interventions that aim to correct belief-driven mistakes. Passive exposure to optimal behavior reaches a meaningful but limited share of the treatable population: approximately 25% of Exposure subjects shift from suboptimal to optimal choices during the exposure phase, and among these compliers, roughly 70% retain the improvement once exposure ends. Adding explicit guidance substantially extends this reach. Under advice, compliers recover to near-peak performance levels, and never-takers, who constitute about 42% of the Exposure sample and show little response to observation alone, improve their optimal choice rate to 70%. This contrast suggests that while exposure to others' actions can nudge behavior in the short run, explicit guidance is more effective at reshaping how participants approach the underlying inference problem, particularly for agents whose initial misconceptions are not dislodged by observing payoff-relevant behavior alone.

Connections to the Literature

This paper connects to several strands of the literature. First, it contributes to the large literature on social learning. A classical strand studies sequential settings in which agents observe predecessors' actions and must infer the private information behind them, with theoretical foundations in A. V. Banerjee (1992), Bikhchandani, Hirshleifer & Welch (1992), and Smith & Sørensen (2000), and a substantial body of experimental work documenting deviations from Bayesian aggregation (Anderson & Holt, 1997; Duffy et al., 2019; Hung & Plott, 2001; Weizsäcker, 2010). Within this sequential paradigm, Eyster, Rabin & Weizsacker (2018) document that subjects imitate predecessors' actions even when those actions are redundant with public information, overweighting what they observe others do. A more recent strand studies repeated environments in which agents observe others over time and examines how observation and imitation shape learning when private information is weak or shared (Evdokimov & Garfagnini, 2020).[12]

[12] The theoretical foundations for imitation as a learning strategy include Schlag (1998), who shows that the proportional imitation rule is the unique rule that is improving and monotone under minimal information requirements, and related work on imitation dynamics in games (Björnerstedt & Weibull, 1996; Schlag, 2010; Vega-Redondo, 1997). Experimental evidence confirms that "imitate the best" is a robust behavioral tendency (Apesteguia, Huck & Oechssler, 2007; Huck, Normann & Oechssler, 1999; Offerman & Schotter, 2009). In psychology, Bandura's (1977) social learning theory identifies four processes—attention, retention, reproduction, and motivation—governing whether exposure translates into behavioral change, and Bandura, Ross & Ross (1963) show that imitation is modulated by

Particularly closely related is Agranov et al. (2024), who study learning through imitation in a repeated environment where others' actions carry no additional payoff-relevant information, showing that observation can nonetheless improve performance. This paper shares that feature but extends the approach in three directions: I exogenously vary the quality of observed choices, I test whether improvements persist once exposure ends, and I assess whether learning transfers to a modified environment. In a complementary direction, Conlon et al. (2026) document a distinct barrier to social learning: in a belief-updating task, participants systematically underweight a partner's signals relative to equivalent signals they generate themselves, even when the partner's information is perfectly communicated or directly observed.[13] While their paper studies how people weight others' *signals* when forming beliefs, I study whether observing others' *choices* can shift decision-making, and whether any resulting behavioral change is durable. The two sets of findings are complementary: people discount others' information when aggregating signals, and even when they do respond to others' behavior, the resulting learning is often shallow.

Second, this paper connects to a growing body of work on learning under misspecified mental models, which shows that agents who interpret feedback through a flawed representation of the environment converge to systematically biased steady states (Bohren & Hauser, 2021; Esponda & Pouzo, 2016; Fudenberg, Romanyuk & Strack, 2017; Heidhues, Kőszegi & Strack, 2018).[14] Feedback may have limited corrective power if agents remain confident in incorrect models, selectively attend to confirming information, or face cognitive uncertainty about the mapping from signals to optimal actions (Enke & Graeber, 2023; Hanna, Mullainathan & Schwartzstein, 2014; Spiegler, 2020). Most closely related is Esponda, Vespa & Yuksel (2024), who document persistent suboptimal behavior even under extensive feedback in a statistical inference task closely related to the one studied here.[15] I contribute to this line of work by studying a different channel through which mental models may be challenged: rather than providing explicit corrective information, I expose participants to the

the perceived payoff consequences of the modeled actions. In field settings, Conley & Udry (2010) show that farmers adjust inputs to align with those of successful peers, and Rendell et al. (2010) highlight that social learning strategies tend to outperform individual learning because demonstrators inadvertently filter information by displaying their highest-payoff behaviors.

[13]Their finding identifies a bias in information aggregation—people treat others' data as inherently less informative regardless of how it is transmitted—and rules out explanations based on communication frictions, mistrust, or presentation format.

[14]Esponda & Pouzo (2016) show that agents who maximize expected utility under a misspecified model converge to biased steady states; Heidhues, Kőszegi & Strack (2018) demonstrate that even small belief departures can sustain overconfident behavior indefinitely; Fudenberg, Romanyuk & Strack (2017) show that incorrect priors can prevent agents from experimenting sufficiently to encounter corrective data; and Bohren & Hauser (2021) extend these results to social settings where agents also draw inferences from others' behavior.

[15]In one of their diagnostic treatments, the authors directly challenge participants' confidence by informing them that their initial choice was incorrect, which alters subsequent behavior and sheds light on how confidence in the initial response can reduce engagement and result in slow and partial learning.

choices of another agent facing the same task, introducing an external reference point that may prompt reassessment of how agents engage with and interpret feedback.

Third, a distinctive feature of this paper is its focus on the durability and generalizability of learning gains. While most of the social learning literature documents contemporaneous effects of observation, substantially less is known about what happens once the model is removed. In psychology, Bandura's (1977) theory emphasizes that behavioral gains from observation may dissipate if retention relies on superficial features rather than deeper understanding of why the modeled behavior is effective. The question of whether learning transfers to novel settings dates back to Thorndike & Woodworth (1901), who proposed that transfer depends on shared elements between training and test environments.[16] In economics, field experiments on technology adoption provide some evidence on persistence: Dupas (2014) shows that subsidized exposure to health products generates lasting adoption, while Conley & Udry (2010) document persistent adjustments following observation of successful peers. This paper contributes controlled evidence on both retention and transfer, finding that the answer depends critically on the depth of understanding that exposure induces.

A related body of work on experience effects and reinforcement learning helps explain why observation-based learning gains may be inherently fragile. Personal experiences disproportionately shape beliefs and economic decisions, even when more complete data are readily available (Malmendier & Nagel, 2016), and experimental evidence demonstrates that people's beliefs and actions react more strongly to events that personally affect them than to equivalent events they merely observe (Merlo & Schotter, 2003; Simonsohn et al., 2008). The evidence in this paper is consistent with this interpretation: subjects in Exposure do not observe the payoffs of the participant they watch, and instead try the observed alternative primarily after their *own* choice delivers a low payoff, keeping it when their own payoff experience turns favorable. The fragility of learning gains once exposure ends is consistent with the removal of the external action cue around which this own-payoff experimentation was organized.

Fourth, this paper contributes to understanding the relative effectiveness of passive exposure to others' actions versus explicit guidance. In Part 5 of the experiment, subjects receive written advice from the same agent whose choices they previously observed, providing a within-design comparison of these channels. A related body of work studies intergenerational transmission of knowledge: Schotter & Sopher (2003) show that free-form advice can establish and sustain coordination con-

[16] See Barnett & Ceci (2002) for a taxonomy of near and far transfer, and Singley & Anderson (1989) for an extension to cognitive skills, arguing that transfer depends on the degree to which underlying production rules, rather than surface features, are shared across tasks.

ventions, and Chaudhuri, Graziano & Maitra (2006) find that the effectiveness of intergenerational advice depends on its publicity.[17] In this paper, advice produces improvements that are both faster and more broadly distributed than those achieved through exposure alone. In particular, never-takers, who show little response to observation, improve substantially under advice, suggesting that explicit guidance can reach agents whom passive exposure fails to move.

Finally, given that participants face a repeated inference task framed as a hiring decision, this paper also connects to the literature on inaccurate statistical discrimination. The classic theory of statistical discrimination assumes that decision-makers hold accurate beliefs about group-level productivity and condition on them rationally (Arrow, 1973; Phelps, 1972), and several empirical tests find behavior consistent with such accurate beliefs (Knowles, Persico & Todd, 2001; List, 2004). A more recent literature shows, however, that discrimination is often *inaccurate* — rooted in miscalibrated beliefs about groups that can persist even in the presence of outcome feedback (Bohren, Imas & Rosenberg, 2019). Recent experimental work has sharpened these findings: Campos-Mercade & Mengel (2024) attribute approximately 40% of the hiring gap in their experiment to non-Bayesian statistical discrimination driven by signal neglect, Esponda, Oprea & Yuksel (2023) document representative signal distortion, and Bohren et al. (2023) show that ignoring inaccurate beliefs can lead to erroneous attributions of taste-based discrimination. This paper contributes by examining whether exposure to another agent's optimal behavior can serve as a debiasing mechanism, finding that it partially corrects signal neglect but that the resulting improvements are fragile and do not fully transfer to modified environments.

2 Experimental Design

In this section, I provide a description of the overarching design framework. The experiment uses a between-subjects design, with participants exogenously assigned to observe either optimal choices, suboptimal choices, or no choices made by others. There are five parts and 120 independent rounds in total. In each round, participants make one binary choice and receive immediate feedback. Below, I describe the main inference task used in the experiment, the treatment design, and the main objective of each part.[18]

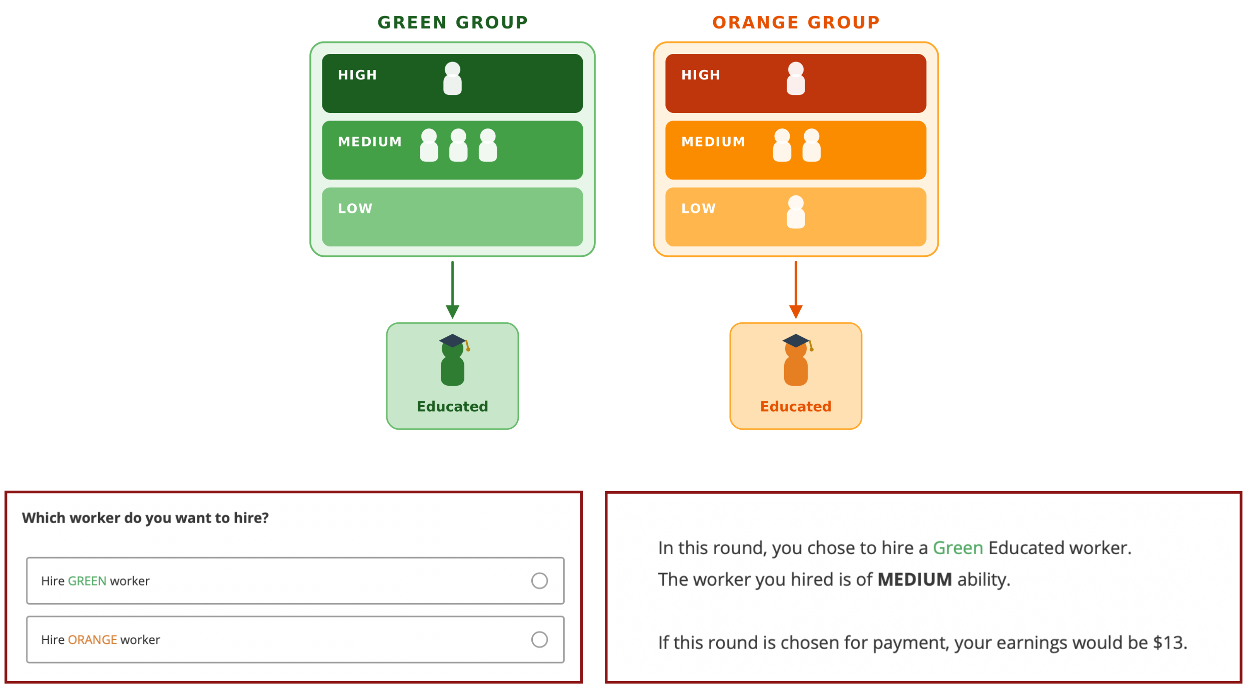
[17] Nyarko, Schotter & Sopher (2007) study experimental markets in which subjects can purchase either raw data about predecessors' choices or synthesized advice, finding that subjects bid significantly more for data. In field settings, A. Banerjee et al. (2013) show that the network position of initially informed individuals shapes diffusion through social ties; Cai, De Janvry & Sadoulet (2015) demonstrate that peer effects on insurance adoption operate through knowledge diffusion rather than imitation; and Duflo & Saez (2003) find that peer exposure to information sessions increases retirement plan enrollment among untreated colleagues.

[18] Experimental instructions can be found in Appendix H.

2.1 Inference Task

Throughout the experiment, all participants assume the role of a hiring manager who must choose between two workers: one from a group of Green workers ($k = green$) and one from a group of Orange workers ($k = orange$). Each group consists of four workers, and each worker can be of either low, medium, or high ability: $a \in \{l, m, h\}$. In every round, one worker is randomly drawn from each group, and participants must choose which of the two they would like to hire. Participants can observe the distribution of ability levels in both groups, but do not observe the abilities of the two selected workers.[19]

Figure 1: Group Composition and Decision Screen Interface



Instead, they observe the selected workers' binary education status (*educ* or *not educ*). Participants are also informed about the relationship between ability level and education status – they know that high-ability workers are educated with certainty, $\mathbb{P}(educ \mid h) = p_h = 1$, medium-ability workers are educated with 90% chance ($p_m = 0.9$), and low-ability workers are *not* educated with certainty ($p_l = 0$). The interface of the experiment is presented in Figure 1. After learning the education status of two selected workers, participants choose either to hire a Green worker or an Orange worker.

[19] In the experiment, one of two ability distributions is randomly selected in each round. One distribution is shown in Figure 1, and the other is nearly identical except that one medium-ability worker in each group is replaced with a high-ability worker. Theoretical predictions are identical for the two distributions.

Benchmark. Given the environment described above, it is straightforward to identify the theoretical benchmark. Consider the modal scenario in which both the green and the orange worker are educated. It is useful to contrast explicitly the two approaches a participant may take to this decision.

Suboptimal approach. A participant may neglect the informational content of the education signal — after all, both workers display the *same* signal, which makes it appear uninformative for the comparison — and instead rank the two groups by their *average* ability. Because the green group has higher average ability ex ante, this reasoning leads to hiring the **green** worker whenever both candidates are educated.

Optimal approach. The education signal, however, screens ability differently across the two groups. Since low-ability workers are never educated ($p_l = 0$), observing that the selected orange worker is educated *eliminates* the possibility that he is of low ability. Conditional on both workers being educated, the relevant comparison is therefore the ratio of high- to medium-ability workers, which is higher in the orange group – so the payoff-maximizing choice is to hire the **orange** worker, despite the green group's higher average ability.[20]

This logic is confirmed by the Bayesian posterior. Let k_a represent the proportion of workers of group identity $k \in \{green, orange\}$ who have ability $a \in \{l, m, h\}$. Using Bayes' rule, the posterior probability that an educated worker from group k is of high ability is given by:

$$\mathbb{P}(h \mid educ) = \frac{\mathbb{P}(educ \mid h) \times \mathbb{P}(h)}{\mathbb{P}(educ)} = \frac{p_h \times k_h}{p_h \times k_h + p_m \times k_m + p_l \times k_l} = \frac{k_h}{k_h + 0.9 k_m}$$

Noting that $green_h = 0.25, green_m = 0.75$, $orange_h = 0.25$ and $orange_m = 0.5$, we have $\mathbb{P}(h \mid educ) \approx 0.27$ for the green worker and $\mathbb{P}(h \mid educ) \approx 0.36$ for the orange worker. Importantly, participants only need to make a binary hiring choice, which does not require computing these conditional probabilities: recognizing that education rules out low ability, and comparing the resulting high-to-medium ratios, suffices. The focus in this paper is therefore on a type of learning that is more general than computing the correct Bayesian posterior belief: the key element is to not neglect the education signal but rather extract information from it, which is consistent with Bayesian reasoning.

[20] The inference task in this experiment is adapted from Campos-Mercade & Mengel (2024), who study how difficulties with Bayesian updating interfere with hiring decisions in a setting that speaks to the literature on statistical discrimination. While their focus is on documenting non-Bayesian behavior in individual decision-making, the present paper uses a similar environment to study whether exposure to others' choices can mitigate such biases.

Observation: *In the experiment, the payoff-maximizing choice conditional on both workers being educated is to hire the* ***Orange*** *worker. A naive participant, seeing that both workers are educated, would underinfer from the education signal and instead hire the* ***Green*** *worker, who comes from the group with higher average ability ex ante.*

2.2 Experiment Structure

Below I describe each of the five parts of the experiment and between-subject treatment design.

Part 1: Baseline Behavior (20 rounds). After each round, participants receive feedback about the ability of the worker they hired, along with the earnings for that round. Participants' earnings in the experiment depend on both the ability level and education status of the worker they hire. In particular, they earn more if the hired worker is educated, and conditional on an education status, they earn more if the hired worker is of higher ability.[21] Note that the ability distributions and the choice of environment primitives result in most rounds having both workers be educated, which is useful because I am specifically interested in subjects' behavior in these "rounds of interest," referred to as *key rounds* throughout the paper.[22] In the remaining rounds, in which at least one worker is not educated, one option is transparently dominant, and 96% of subjects' choices are optimal. The goal of Part 1 is to establish the share of optimal choices at baseline and to give subjects an opportunity to practice with the experimental interface.

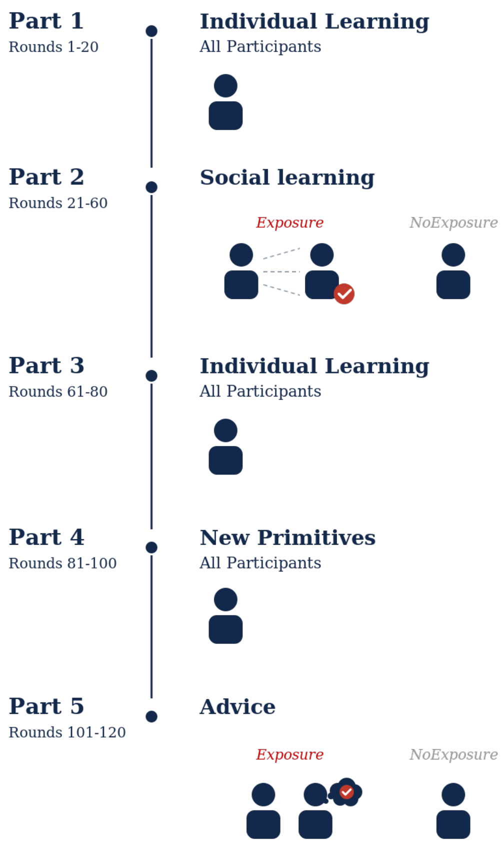
Part 2: Exposure to External Choices (40 rounds). Subjects continue making hiring decisions between a green worker and an orange worker, exactly as in Part 1. The central manipulation of the experiment occurs in this part: to exogenously vary exposure to others' choices, subjects are randomly assigned to one of two main treatment conditions, depicted in Figure 3.[23]

[21] Specifically, if they hire an Educated worker their payoff is either \$18 (high ability) or \$13 (medium ability). If they hire a Not educated worker, their payoff is either \$10 (medium ability) or \$5 (low ability).

[22] For the distribution of abilities in Figure 1, the probability that the worker from the green group is educated is $\mathbb{P}(educ) = \mathbb{P}(educ \mid h) \times \mathbb{P}(h) + \mathbb{P}(educ \mid m) \times \mathbb{P}(m) + \mathbb{P}(educ \mid l) \times \mathbb{P}(l) = 1 \times 0.25 + 0.9 \times 0.75 = 0.925$; the probability that the worker from the orange group is educated is $\mathbb{P}(educ) = 1 \times 0.25 + 0.9 \times 0.5 = 0.7$. Hence, the joint probability of both workers being educated is $0.925 \times 0.7 = 0.6475$, so by construction we expect that roughly two out of every three rounds would be rounds of interest with both workers being educated. In the experiment, around 71% were rounds of interest.

[23] The experimental design also included two diagnostic treatments. In ExposureNeg, subjects observed the choices of another participant who always made *suboptimal* decisions (i.e., hiring the green worker whenever both candidates were educated); outside the key rounds, this participant's choices coincided with the transparently dominant option, as do those of the optimal participant observed in Exposure, so the two observed choice streams differ only in the key rounds. In Exposure (Full Info), subjects received the same social exposure as in Exposure but additionally observed the payoff that would have resulted from the alternative hiring decision. For focus and clarity, the main text concentrates on the Exposure vs. NoExposure comparison. Detailed analysis of the ExposureNeg treatment is presented in Section 3.3, and the Exposure (Full Info) treatment is discussed in Appendix C.1.

Figure 2: Experiment Structure and Treatment Design



Subjects in NoExposure (**NoExp**) continue to make all decisions independently, in an environment identical to Part 1; this group serves as the reference condition. Subjects in Exposure (**Exp**) are additionally presented, in each round, with the hiring choice made by a participant from one of the previous study sessions who faced the exact same inference task and always made *optimal* decisions – that is, this participant chose to hire the orange worker whenever both candidates were educated. The instructions introduce this participant as follows:

> "For Part 2 of the experiment, in each round you will be provided with the *hiring choice of another study participant* who took part in the exact same experiment in one of the past sessions – we will call this participant 'participant X'. In each round, participant X saw exactly the same information as you will see."[24]

[24] Subjects are specifically informed: "If in round 30 you see that the abilities draw is Draw #1, the selected Green worker is educated, and the selected Orange worker is educated, it means that participant X in round 30 saw that the abilities draw is Draw #1, the selected Green worker is educated, and the selected Orange worker is educated. In rounds 21-60, you will be provided with this hiring choice and the interface will, for instance, say 'Participant X chose to hire the Green worker' or 'Participant X chose to hire the Orange worker' ".

Two clarifications about participant X's behavior are useful. First, the "always optimal" label refers to X's choices in the key rounds, in which both workers display the same education status — the margin on which optimal and suboptimal behavior are defined throughout the paper. Second, in the remaining rounds of Part 2, where one option is transparently dominant, participant X made the dominant choice, just as the overwhelming majority of subjects do; the informational content of the treatment is therefore concentrated in the key rounds.

Figure 3: Treatment Conditions in Part 2

Although subjects know that all the hiring choices they see were made by the same person, they are not told whether these choices are optimal or not – in fact, the experiment is designed to better understand subjects' ability to recognize the (sub)optimality of the choices they are exposed to. In particular, the hope is that subjects who initially make suboptimal hiring decisions will begin to reconsider their approach when presented with optimal choices. Notably, this environment differs from more common social learning settings in the literature, where others' choices are designed to convey payoff-relevant information and guide learning.[25] In contrast, in my setting, others' decisions are only useful to the extent that a subject does not already know the optimal approach - if they do, observing others should in theory provide no additional benefit.

Furthermore, note that the choices shown to subjects in the Exposure treatment were actual hiring decisions made by a participant who was originally part of the NoExposure group completing the tasks individually. The main hypothesis in Part 2 is that subjects in the Exposure treatment would exhibit more optimal behavior relative to the control condition with no exposure. Observing optimal choices could especially be helpful for participants who demonstrate higher subjective uncertainty about their approach to the task.[26]

Part 3: Learning Retention (20 rounds). This part follows the same structure as Part 1: all participants, regardless of their treatment group, make decisions independently. The goal of including this part is to examine whether the participants have retained what they learned from

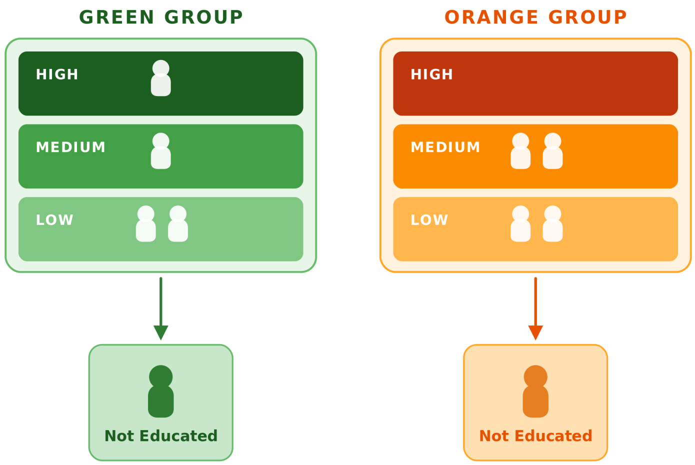
[25] For example, by observing someone's action, a subject can often infer the private signal that person received - information that can then inform their own decision.

[26] Presumably, feeling certain or confident about one's own approach is associated with lower susceptibility to being influenced by the observed choices of another subject, regardless of whether one's own choices are optimal. I test this prediction in a later section.

being exposed to another subject's choices in Part 2.

Specifically, if they revert to their original decision patterns from Part 1, it would suggest that their earlier behavior was a result of short-term imitation rather than genuine learning. There is, however, an important caveat: participants may have simply memorized the choices they observed in Part 2, and even without continued exposure in Part 3, they might still replicate those choices from memory. To separate genuine learning of the decision rule from memory-based replication, it is instructive to subject participants to a slightly different environment.

Figure 4: Distribution of Abilities in Parts 4 and 5



Part 4: New Environment Primitives (20 rounds). This part is designed to better understand whether subjects have truly learned from seeing optimal choices in the EXPOSURE treatment. Subjects in all treatments complete 20 rounds independently, similar to Part 1 and Part 3. The main difference is that in Part 4 the inference task uses a *different distribution of abilities* depicted in Figure 4. In this new setup, the green group consists of two low-ability workers, one medium-ability worker, and one high-ability worker, while the orange group has two low-ability and two medium-ability workers. Subjects are also informed that a medium-ability worker is educated with a 10% chance ($p_m = 0.1$) under the new primitives of the environment. The remaining features of the hiring task stay intact: high-ability workers are always educated, low-ability workers are never educated, and the green group has a higher average ability.

Crucially, most of the rounds will now feature both workers being *Not Educated*, and conditional on this, the optimal choice remains to be *hiring an orange worker*.[27] The objective of Part 4 is

[27]If both workers are not educated, then neither can be high-ability since high-ability workers are always educated. The relevant posterior is therefore $\mathbb{P}(m \mid \textit{not educ})$: since $p_m = 0.1$, a medium-ability worker is *not* educated with probability 0.9. For the green group, $\mathbb{P}(m \mid \textit{not educ}) = \frac{0.9 \times green_m}{0.9 \times green_m + 1 \times green_l} = \frac{0.9 \times 0.25}{0.9 \times 0.25 + 0.5} \approx 0.31$. For the orange

straightforward: to establish whether subjects would exhibit similar treatment effects under the new environment primitives, i.e., whether they would continue making significantly more optimal choices after exposure to optimal choices. In particular, comparing the prevalence of optimal behavior among Exposure subjects across the two environments helps separate context-bound adoption from behavior consistent with deeper learning, with the caveats on interpreting the latter discussed in Section 4.

Figure 5: Advice in Exposure

Part 5: Direct Advice (20 rounds). In the last part of the experiment, before subjects complete 20 more rounds under the new environment primitives, they are shown a written recommendation by the same participant whose choices they were observing in Part 2. In particular, subjects in the Exposure treatment read advice written by the participant making 100% optimal choices, that explains how to optimally approach the task and extract useful information from education signals (see Figure 5).[28] The goal here is to see how exposure to an elaborate written recommendation to make a particular choice performs relative to observing others' choices.

Implementation Details

The experiment was conducted in person at the University of California San Diego Economics Laboratory (EconLab). Prior to the main part of the experiment, all subjects were required to

group, $\mathbb{P}(m \mid not\ educ) = \frac{0.9\times0.5}{0.9\times0.5+0.5} \approx 0.47$. Since the orange worker is about one and a half times as likely to be of medium ability conditional on being not educated, the optimal choice is again to hire the orange worker. The stakes on this margin are, if anything, larger than in Parts 1–3: the expected per-round payoff difference between the two options in the key rounds is \$0.82 in Parts 4–5, compared with \$0.43 in Parts 1–3, so the payoff feedback available for learning the optimal choice is stronger in the new environment. As in Parts 1–3, subjects need only make a binary hiring choice rather than compute these posterior probabilities explicitly.

[28] The numerical rationale in the participant-written advice is approximate: for the both-not-educated scenario it computes the chance of a medium-ability worker from worker counts after excluding high-ability workers (2 of 4 for orange; 1 of 3 for green), rather than from the full posterior that also weights education probabilities ($\approx$ 47% versus $\approx$ 31%; see the derivation in the Part 4 description above). The recommendation and the ranking of the two options are correct. If anything, this feature reinforces the interpretation developed later: the advice that moves subjects is an accessible, directionally correct rationale rather than exact Bayesian arithmetic.

correctly answer six comprehension questions to proceed. After that, participants could move at their own pace but they were informed that they would not be able to leave early. A total of 303 subjects participated in the two main treatments: 223 in Exposure and 80 in NoExposure.[29] On average, subjects earned \$14.4 in the experiment. Final payment was based on the outcome of one randomly selected round, together with a separately drawn confidence-elicitation round that determined the chance of winning an additional \$2 bonus (Appendix G). At the end of the experiment, subjects completed a short demographic survey.

Preregistration and empirical strategy. The study was pre-registered in the AEA RCT Registry under AEARCTR-0013607. The registered primary outcome is the round-level binary indicator of whether a subject's choice coincides with the Bayesian benchmark, to be used both for round-by-round dynamics and aggregated across rounds and treatment groups; the registered secondary outcomes are the incentivized confidence elicitations, response times, and subjects' end-of-experiment strategy descriptions. Each subject was randomly assigned to one of the treatment groups; recruitment materials and sign-up procedures were identical across conditions, so subjects had no information about treatment when enrolling. Because participants made all decisions autonomously — with no communication, no shared feedback, and no group-level payoffs — there is no mechanical channel for within-session correlation of errors, and standard errors are accordingly clustered at the subject level throughout. All subjects were required to answer every comprehension question correctly in order to proceed to the incentivized rounds; no subject was excluded on this basis, and the analysis includes all 400 participants who completed the experiment: 223 in Exposure and 80 in NoExposure (303 subjects in the two main arms), together with 53 in ExposureNeg and 44 in Exposure (Full Info). Sessions took place in May and December 2024, May–June 2025, and May 2026; all four treatments were fielded from the first wave in May 2024 onward, consistent with the preregistration, and the instructions and procedures remained identical across waves. The later sessions served to increase the overall sample size and, in particular, to over-recruit the Exposure arm: because much of the analysis characterizes heterogeneity in individual responses within this group, and the subgroup of central interest makes up roughly a quarter of it, a larger sample in this arm was needed for adequately powered subgroup comparisons. In implementing the registered aggregate comparisons, the paper reports the share of optimal choices in the last ten key rounds of each part as its principal summary, alongside the last-round indicator and a binary indicator for making at least eight of ten optimal choices; where these endpoints differ in statistical precision, all are reported. The randomized treatment comparisons of Sections 3.2 and 3.3 constitute the confirmatory analyses, while the learning-type taxonomy,

[29] An additional 53 and 44 subjects were assigned to the ExposureNeg and Exposure (Full Info) treatments described earlier; they are analyzed separately in Section 3.3 and Appendix C.1, respectively.

complier subtypes, confidence interactions, and simulation exercises are exploratory analyses that organize the heterogeneity behind them.[30]

3 Results

I organize the main experimental results as follows: In Section 3.1, I report the incidence of optimal behavior at baseline and show that a majority of subjects choose a suboptimal option. In Section 3.2, I then present the main treatment effects of exposure to external choices in Part 2 of the experiment, establishing that exposure to optimal decisions facilitates learning. Throughout the analysis, I focus on the key rounds: rounds in which both workers are Educated in Parts 1–3, and rounds in which both workers are Not Educated in Parts 4–5.[31] I postpone the analysis of learning robustness and retention to the subsequent section.

3.1 Baseline Suboptimal Behavior

In the initial rounds of the experiment in Part 1, all subjects complete the tasks individually regardless of their treatment assignment and receive feedback after each round. At baseline, fewer than two in five of subjects' choices in the key rounds of Part 1 are optimal (37%), and the share remains at only 40% in the last ten rounds; that is, when both the Green and the Orange worker are educated, most subjects fail to hire the Orange worker.[32] In fact, even after some experience with the environment more than half of subjects stick to the suboptimal Green worker option in the last round of Part 1.

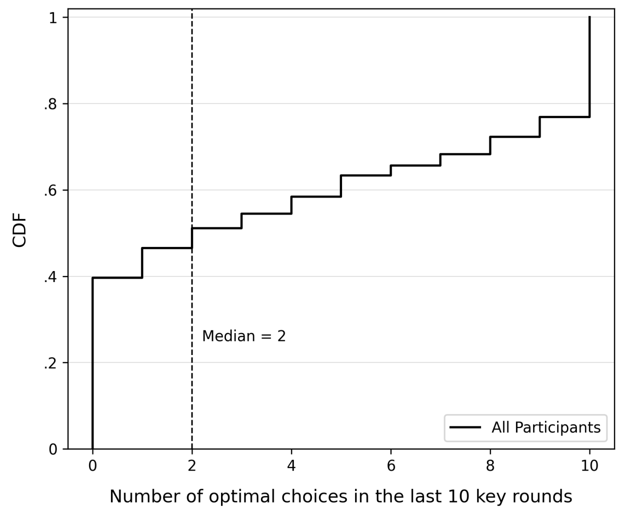
The bimodal nature of baseline behavior is further illustrated in Figure 6, which plots the cumulative distribution of optimal choices in the last ten rounds of Part 1 pooled across the Exposure and NoExposure groups. It can be seen that the median subject makes only two optimal choices out of ten. While approximately 23% of subjects are fully optimal with all ten choices correct, 40% of subjects make zero optimal choices. The overall shape confirms that subjects are not clustered around an intermediate performance level; rather, the distribution is spread across the entire range, with pronounced concentrations at the two extremes.

[30] In particular, the 80% threshold used in the learning-type classification was not part of the pre-registration. The appendix on alternative complier specifications therefore examines a range of alternative thresholds and classification criteria and shows that the composition of learning types and their dynamics are qualitatively unchanged.

[31] Out of 120 rounds across five parts of the experiment, 85 constitute rounds of interest (71%). In the remaining 35 rounds, which fall outside these signal configurations, 96% of subjects' choices are optimal and there is no statistically significant difference between treatments based on a joint test ($p = 0.755$).

[32] Departure from the optimal benchmark is costly. The gap in earnings for the average subject is statistically significant ($p < 0.001$), demonstrating that suboptimal behavior translates into measurable monetary losses.

Figure 6: Cumulative Distribution of Optimal Choices in Part 1



Notes: This figure plots the cumulative distribution of the number of optimal choices in the last ten rounds of Part 1 for all 303 participants, pooled across the Exposure and NoExposure groups. All subjects complete Part 1 individually, so no treatment differences are expected at baseline. The vertical dashed line denotes the median.

Given that all subjects complete the tasks individually in Part 1, irrespective of treatment assignment, no differences across treatment groups are expected at baseline. Consistent with this prediction, a test of equality of baseline optimality rates across the Exposure and NoExposure groups fails to reject the null hypothesis ($p = 0.88$). Throughout the paper, NoExposure serves as the reference group and represents the counterfactual of completing the inference task individually in every part of the experiment. Because all subjects receive immediate feedback on the outcome of their hiring choice in each round, it is natural to expect gradual learning even in the absence of social exposure.[33]

While details on the distribution of subject types are presented in subsequent sections, the key takeaway from this section is that the study's baseline condition holds: in the first part of the experiment, most subjects choose a dominated option, indicating substantial scope for learning. This result is summarized below:

Result #1: *At baseline, fewer than two in five choices are optimal, and about 40% of participants make zero optimal choices in the last ten rounds of Part 1.*

Next, I study subjects' behavior in Part 2 to evaluate to what extent exposure to others' choices can influence own decisions. The goal of the next section is, therefore, to assess whether observing decisions made by another subject can accelerate learning relative to the individual-learning control

[33] Consistent with this intuition, the share of optimal choices in the NoExposure group increases modestly over time, reaching 49% in the last ten rounds of Part 2.

group.

3.2 Learning from Exposure to Others' Decisions

In Part 2 of the experiment, which lasts for 40 rounds, subjects in the Exposure group additionally observe choices made by another participant who approached the task optimally. Given that subjects are not informed whether the observed choices are optimal or not, the goal of Part 2 is to assess whether participants can recognize the quality of these decisions and learn from them.

By the end of Part 2, optimal behavior in the Exposure group is substantially higher than in NoExposure. In the last round of Part 2, 67% of Exposure subjects make the optimal choice, compared with 46% in NoExposure, a 21 percentage point difference that is statistically significant ($p = 0.001$). The treatment effect is especially pronounced at the median: a quantile regression reveals that the median Exposure subject makes 8 out of 10 optimal choices in the last ten rounds, compared with 5 out of 10 in NoExposure ($p < 0.001$).[34]

To explore the dynamics of these treatment effects, I deploy the following measure that captures the notion of switching to optimal behavior: for each treatment group and for each of the 40 rounds in Part 2, I compute the cumulative percentage of subjects who choose the optimal alternative in a given round and continue to do so in at least 80% of remaining rounds in this part.[35] As Figure 7a shows, about 40% of NoExposure subjects choose optimally in most of the remaining rounds from the start, and this number rises only to 43% by round 55, the end of the displayed horizon. In contrast, the share of Exposure subjects switching to optimal behavior starts to grow steadily from round 35 onward, gradually widening the gap between the two groups.[36]

The cumulative distribution of optimal choices in Part 2, plotted in Figure 7b, confirms that the improvement in Exposure reflects a broad shift in behavior: the distribution of optimal choice shares across subjects first-order stochastically dominates that in NoExposure.[37] Another

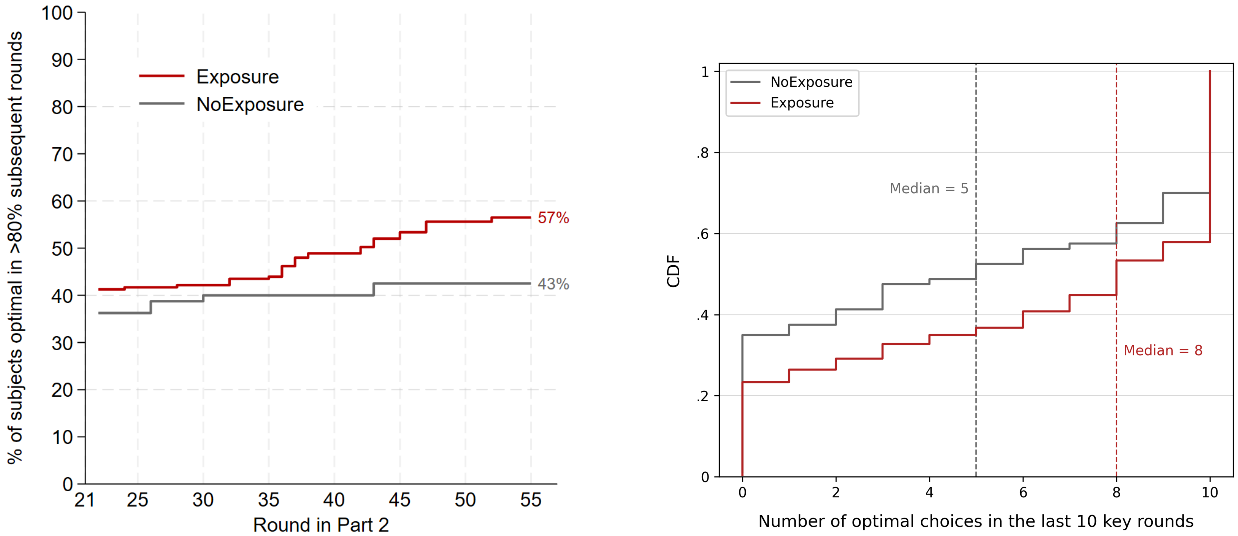
[34] Detailed statistical tests at multiple time horizons are reported in Table 6. The treatment effect estimates over the last ten rounds are somewhat attenuated relative to the last-round estimates, reflecting the fact that many Exposure subjects switch to optimal behavior gradually rather than all at once. The treatment effect remains significant at multiple time horizons and under alternative specifications reported in the appendix.

[35] This measure requires that a subject who chooses the optimal option in a given round continues to behave mostly optimally thereafter, while allowing for occasional mistakes. The threshold is justified by the empirical transition probability from optimal choice in round $t - 1$ to suboptimal choice in round t of around 0.13, estimated for the Exposure group. In later sections of the paper, I continue to use this threshold for other types of analyses; qualitatively, the results are robust to alternative specifications.

[36] In Section 3.3, I provide further evidence of empirical patterns consistent with payoff-driven selective adoption: subjects in the Exposure group are over 80% more likely to switch to the optimal choice after their own suboptimal choice yields a low payoff than after a high one.

[37] I test for first-order stochastic dominance using the two-stage procedure in Barrett & Donald (2003). I first test

Figure 7: Effects from Exposure to Optimal Choices



(a) Dynamics of Optimal Behavior (b) Cumulative Distribution of Optimal Choices

Notes: The red line corresponds to Exposure and the gray line corresponds to NoExposure. Panel (a) displays the percentage of subjects choosing the optimal alternative in a given round of Part 2 and then continuing to do so in at least 80% of subsequent rounds; the lines end at round 55, beyond which too few rounds remain for the criterion to be meaningful, and the numbers displayed next to each line report the value of this measure at round 55. Panel (b) plots the cumulative distribution of the number of optimal choices in the last ten rounds of Part 2 for each group. The vertical dashed lines denote the median for each group. The distribution of optimal choices in Exposure first-order stochastically dominates that in NoExposure; details of statistical tests are provided in the text.

notable feature is that the Exposure group tends to split into three broad clusters: roughly 42% of subjects make all 10 optimal choices, 23% make none, and the remaining 35% fall between these two extremes.

The results on learning under exposure to external decisions are summarized as follows:

Result #2: *By the end of Part 2, there are substantially more optimal choices in Exposure than in NoExposure. The distribution of optimal choices in Exposure first-order stochastically dominates that in NoExposure, with statistically significant differences at both the mean and the median.*

A natural concern about this result is that the treatment effect may partly reflect richer feedback rather than observation per se: because the observed choices are optimal, the payoff feedback that accompanies them is systematically more favorable than what a suboptimal chooser generates alone. A diagnostic treatment, Exposure (Full Info), addresses this concern by additionally

the null hypothesis that the distribution in the NoExposure treatment either first-order stochastically dominates or is equal to the distribution in the Exposure treatment. I reject this null hypothesis at the 1 percent level ($p < 0.001$). I then test the null hypothesis that the distribution in the Exposure treatment first-order stochastically dominates the distribution in the NoExposure treatment. The null in this case is not rejected ($p = 0.539$).

revealing the counterfactual payoff of the alternative choice, thereby eliminating any informational advantage of observing optimal behavior. Behavior in this group is statistically indistinguishable from Exposure (70% versus 62% optimal choices over the last ten rounds, $p = 0.204$; 66% versus 67% in the last round, $p = 0.863$), indicating that richer feedback adds little beyond the observation of choices itself; Appendix C.1 reports the full analysis and the associated caveats.

3.3 Understanding Potential Learning Mechanisms

The results so far show that subjects perform substantially better when exposed to optimal behavior, even though the observed choices are never labeled as such. A natural question is what drives this positive effect. A higher optimality rate relative to the individual-learning control group is consistent with several distinct underlying processes. On the one hand, it could mean that subjects blindly imitate the observed behavior without internalizing why it is optimal to do so; on the other hand, it could mean that exposure to an alternative approach challenges subjects to reconsider what the optimal decision rule is and why hiring the orange worker is indeed the payoff-maximizing option. In this and the subsequent section, I show that on aggregate it is neither, and that the operative channel lies between these two extremes. There is, of course, heterogeneity across subjects — some are pure imitators and some are deep learners, as the classification in Section 3.4 makes precise — but the goal here is to explain what drives the treatment effect on aggregate. To this end, I consider three potential learning channels, each with a distinct observable prediction:

1. **Mechanical imitation**

   ↪ subjects copy observed behavior, regardless of its quality

   Predicts: symmetric effects — bad examples hurt as much as good ones help

2. **Selective adoption**

   ↪ subjects test the observed alternative against their own payoff experience

   Predicts: asymmetric, gradual adoption that tracks own payoffs; gains tied to the original context

3. **Model revision**

   ↪ subjects infer *why* the observed choice is optimal

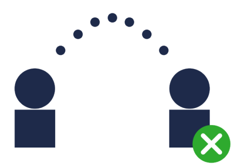
   Predicts: gains persist and transfer to related environments

This subsection evaluates the first two channels: a diagnostic treatment with exposure to *sub-optimal* choices tests mechanical imitation, and the timing of subjects' switching behavior provides direct evidence on selective adoption. The third channel, model revision, requires observing behavior once exposure ends and the environment changes; it is the subject of Section 4.

Exposure to Suboptimal Behavior

If subjects simply replicate observed behavior regardless of its content, then exposure to suboptimal decisions should produce a decline in performance comparable in magnitude to the improvement generated by exposure to optimal decisions. To test this prediction, the experimental design includes an additional treatment, ExposureNeg (**ExpNeg**), depicted in Figure 8. A total of 53 subjects were assigned to this treatment. In Part 2, ExposureNeg subjects observe the hiring choices of a participant from a prior session who consistently made *suboptimal* decisions — always hiring the green worker whenever both candidates were educated. As in Exposure, the observed choices are never labeled, and the instructions are identical up to the identity of participant X. In Part 5, subjects in this group read the written recommendations of the same participant. In all other parts, ExposureNeg subjects make decisions individually, identical to all other participants.

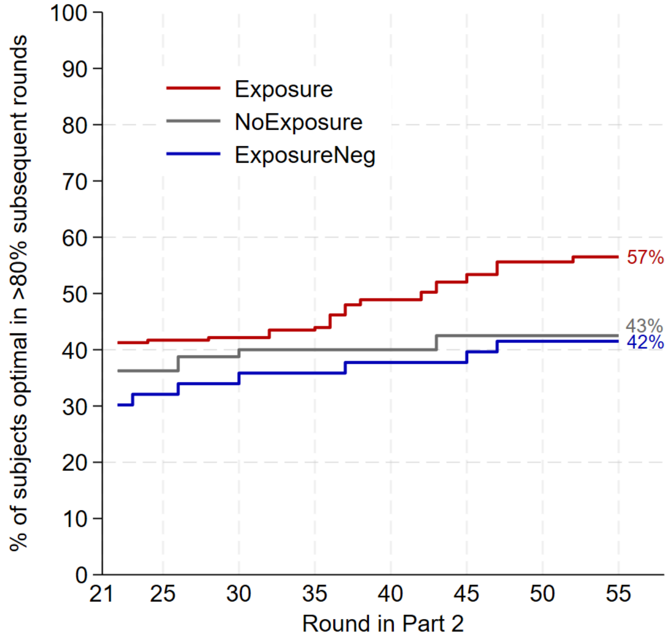
Figure 8: The ExposureNeg Treatment



The data reject the mechanical imitation hypothesis. In the last ten rounds of Part 2, the share of optimal choices in ExposureNeg is 48%, compared with 49% in NoExposure, a difference that is not statistically significant ($p = 0.935$). By the final round of Part 2, 47% of ExposureNeg subjects make the optimal choice, again indistinguishable from the control group ($p = 0.918$). In contrast, the Exposure group reaches 62% in the last ten rounds and 67% in the last round. The gap between Exposure and ExposureNeg is statistically significant in the last ten rounds ($p = 0.046$) and especially pronounced in the last round ($p = 0.008$), confirming that observing optimal choices generates substantially larger behavioral gains than observing suboptimal ones. The null for ExposureNeg is also informative about magnitudes: with the baseline control, the 95% confidence interval for its effect on the last-ten optimal choice share spans -0.10 to $+0.09$ (Table 6), so the data rule out deterioration larger than about ten percentage points, though not more moderate harm — and the confidence heterogeneity documented below indicates where such harm concentrates.

Figure 9 illustrates the dynamics of these treatment effects. As in Figure 7a, the figure plots the cumulative percentage of subjects who choose the optimal alternative in a given round and

Figure 9: Dynamics of Optimal Behavior: Exposure to Optimal and Suboptimal Choices



Notes: Red, blue, and gray lines correspond to Exposure ($N = 223$), ExposureNeg ($N = 53$), and NoExposure ($N = 80$), respectively. The figure displays the percentage of subjects choosing the optimal alternative in a given round of Part 2 and then continuing to do so in at least 80% of subsequent rounds; the lines end at round 55, beyond which too few rounds remain for the criterion to be meaningful, and the numbers displayed next to each line report the value of this measure at round 55.

continue to do so in at least 80% of subsequent rounds. The Exposure line rises steadily from round 35 onward, reaching 57% by round 55, while the ExposureNeg trajectory closely tracks the NoExposure control group throughout (42% versus 43% at round 55). The same ordering holds for the simple last-round choice rates: in the final round of Part 2, 67% of Exposure subjects choose optimally, compared with 46% in NoExposure and 47% in ExposureNeg. The visual pattern confirms that the treatment effects documented in the main analysis are specific to exposure to optimal choices and cannot be attributed to a generic effect of observing any external decisions. Regression evidence underlying these patterns is reported in Table 1.

The asymmetry is sharpest when subjects are split by their own baseline behavior. Two groups in the design observe a peer who systematically *contradicts* their own choices: baseline-suboptimal subjects in Exposure, whose observed peer always chooses optimally, and baseline-optimal subjects in ExposureNeg, whose observed peer always chooses suboptimally. Mechanical imitation predicts that both groups should drift toward the observed behavior. Table 2 shows that only the first group moves: baseline-suboptimal Exposure subjects raise their share of optimal choices from 0.13 to 0.46 (+33 percentage points, against +14 for comparable NoExposure subjects), while baseline-optimal ExposureNeg subjects do not move at all (0.94 to 0.94, against −2 points for

Table 1: Exposure to Optimal vs. Suboptimal Choices

	Part 1 (Individual)		**Part 2 (Social)**	
	Last 10	**Last 1**	**Last 10**	**Last 1**
	(1)	(2)	(3)	(4)
Exposure	0.008 (0.053)	0.023 (0.065)	0.128** (0.056)	0.210*** (0.064)
ExposureNeg	0.004 (0.073)	0.028 (0.088)	-0.006 (0.078)	0.009 (0.089)
NoExposure Mean	0.398	0.425	0.491	0.463
Observations	3,560	356	3,560	356
H_0: Exp = ExpNeg	$p = 0.957$	$p = 0.954$	$p = 0.046$	$p = 0.008$

Notes: Each column reports a separate OLS regression of a round-level optimal choice indicator on treatment dummies (Exposure and ExposureNeg), with the NoExposure group as the reference category. In Part 1, all subjects ($N = 356$) make individual choices, so no treatment differences are expected. In Part 2, Exposure subjects ($N = 223$) observe another participant's optimal choices, while ExposureNeg subjects ($N = 53$) observe suboptimal choices; NoExposure subjects ($N = 80$) continue independently. Standard errors are clustered at the subject level and shown in parentheses. NoExp Mean denotes the average share of optimal choices among NoExposure subjects. The bottom row reports the p-value from a test of equality between the Exposure and ExposureNeg coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2: Who Moves? Baseline Behavior and Contradicting Observations

	Part 1	**Part 2**	**Change**
Subjects suboptimal at baseline ...			
... observing **optimal** choices (Exposure)	0.13	0.46	+33 pp
... observing no one (NoExposure)	0.18	0.32	+14 pp
Subjects optimal at baseline ...			
... observing **suboptimal** choices (ExposureNeg)	0.94	0.94	0 pp
... observing no one (NoExposure)	0.94	0.92	−2 pp

Notes: Entries are shares of optimal choices in the last ten key rounds of Parts 1 and 2. Subjects are optimal at baseline if they make at least eight optimal choices in the last ten key rounds of Part 1 ($y_1 \geq 0.8$), the threshold used throughout the paper. Group sizes: 150 (Exposure, suboptimal), 57 (NoExposure, suboptimal), 18 (ExposureNeg, optimal), 23 (NoExposure, optimal).

comparable control subjects). Subjects adopt observed behavior only when it outperforms their own approach — a first indication of the selective adoption channel developed next. The baseline split also locates the causal effect where randomization identifies it: within the baseline-suboptimal stratum, regressing the Part-2 outcome on treatment with a baseline control yields an Exposure effect of 17.9 percentage points ($p < 0.001$), free of the post-treatment selection that complicates within-group classifications.

Baseline confidence provides a complementary susceptibility margin. While exposure to suboptimal choices has no effect on average, its interaction with baseline confidence is large and statistically significant: point estimates indicate harm concentrated among the least confident subjects (a marginal effect of -21 percentage points at the bottom quartile of confidence, $p = 0.053$), whereas more confident subjects appear insulated. The beneficial effect of exposure to optimal choices, by contrast, is stable across the entire confidence distribution. This pattern is consistent with confident subjects evaluating and filtering the behavior they observe rather than copying it — another indication that adoption is a discriminating, not a mechanical, process. The full analysis is reported in Appendix G.

Payoff-Driven Selective Adoption

The asymmetry documented above rules out mechanical imitation but does not by itself establish how exposure to optimal choices produces learning. This part provides direct evidence on the selective adoption channel by characterizing *when* subjects switch to the observed alternative: transitions between consecutive choices reveal what triggers the decision to try the observed choice, and what sustains it once adopted.

Table 3 reports, separately by treatment, the probability of choosing optimally in round t conditional on the subject's own choice in round $t-1$ and the payoff that choice delivered. Aggregate behavior exhibits substantial inertia: in Exposure, subjects who chose optimally in round $t-1$ continue to do so with probability 0.87, and suboptimal choices persist at a similar rate of 0.82. The informative variation lies in how transitions respond to the subject's own realized payoff, on two margins. The first margin is *trying* the observed alternative. Conditional on a suboptimal choice in round $t-1$, subjects in Exposure switch to the optimal choice with probability 0.22 when their own choice yielded a low payoff (\$13), but only 0.12 when it yielded a high payoff (\$18): a low own payoff raises the propensity to try the observed alternative by over 80%.[38] A weaker version of the same response is present in NoExposure (0.18 versus 0.13), reflecting exploration driven by own feedback alone, and it is entirely absent in ExposureNeg (0.13 versus 0.13), where the observed peer *agrees* with the subject's own suboptimal choice. The second margin is *keeping* the adopted choice. Conditional on an optimal choice in round $t-1$, Exposure subjects persist with probability 0.90 after a high payoff and 0.84 after a low one, whereas persistence in NoExposure is payoff-insensitive (0.83 versus 0.83). This contrast is consistent with the composition of the two groups: optimal choosers in the control group predominantly follow a rule they understand,

[38] Throughout, I describe this as a response to a *low payoff* rather than to "disappointment": the data record payoffs, not emotions, though it is natural to conjecture that a low realized payoff is accompanied by disappointment about not having earned more.

Table 3: Choice Transitions by Own Payoff Experience, Part 2

	Probability of choosing optimally in round t			
	Trying: switch to optimal (suboptimal at $t-1$)		**Keeping**: stay optimal (optimal at $t-1$)	
Own payoff at $t-1$:	High ($18)	Low ($13)	High ($18)	Low ($13)
Exposure	0.12	**0.22**	**0.90**	0.84
NoExposure	0.13	0.18	0.83	0.83
ExposureNeg	0.13	0.13	0.86	0.84

Notes: The table reports transition probabilities between consecutive rounds of interest in Part 2, conditional on the subject's own choice in round $t-1$ and the payoff it delivered. In rounds of interest both workers are educated, so realized payoffs equal $18 if the hired worker is of high ability and $13 if of medium ability. All conditioning variables are observable to the subject. Trying-margin cells pool 1,029 (high) and 1,701 (low) transitions from 145 and 164 subjects in Exposure, 442 and 699 transitions from 59 and 62 subjects in NoExposure, and 313 and 492 transitions from 38 and 42 subjects in ExposureNeg.

while the Exposure group includes many recent adopters whose new behavior is still sustained by favorable payoff realizations.[39]

These contrasts are statistically distinguishable, which separates selective adoption from generic win-stay/lose-shift exploration. Estimating a linear probability model for switching to the optimal choice on the trying margin — with indicators for a low own payoff in round $t-1$, treatment, and their interactions, and standard errors clustered at the subject level — yields a low-payoff response of 10.8 percentage points within Exposure ($p < 0.001$), of which 5.9 points exceed the own-feedback exploration response in NoExposure ($p = 0.013$ for the interaction). In ExposureNeg, where the observed peer agrees with the subject's own suboptimal choice, the response is absent (-0.6 points, $p = 0.76$) and significantly smaller than in Exposure ($p < 0.001$). A low payoff on one's own choice thus triggers movement specifically toward an observed alternative that contradicts the subject's current approach.

These patterns characterize the *timing* of adoption. Disappointment with one's own choice opens the window in which the observed alternative is tried: subjects act on the contrast between their own approach and the observed one primarily after their own approach lets them down. Favorable payoff realizations then sustain the new behavior once adopted. The comparison across treatments underscores that both ingredients matter: absent an observed alternative (NoExposure),

[39] The within-Exposure type split, using the classification introduced in Section 3.4, confirms this reading: always-takers are payoff-insensitive on the keeping margin (0.98 versus 0.97), while compliers are strongly payoff-sensitive on both margins (0.92 versus 0.84 for keeping; 0.47 versus 0.26 for trying).

disappointment produces only weak undirected exploration; and when the observed peer confirms the subject's own suboptimal choice (ExposureNeg), the disappointment response disappears altogether. In Appendix F, I show that a simple Q-learning model in which observation expands the state space reproduces these dynamics.

The evidence in this subsection thus points to selective adoption as the operative channel on aggregate: observation supplies a concrete alternative, and subjects' own payoff experience governs whether and when it is adopted and retained. The first two candidate channels are summarized in the following result:

Result #3: *The effects of exposure are asymmetric in the quality of observed choices: exposure to suboptimal behavior leaves performance statistically indistinguishable from the control group, ruling out mechanical imitation. Consistent with selective adoption, subjects in Exposure are over 80% more likely to switch to the optimal choice after their own suboptimal choice yields a low payoff than after a high one — a response to one's own low payoff that is weaker without exposure and absent under exposure to suboptimal choices.*

What this evidence cannot yet establish is the depth of the resulting learning — whether adoption remains tied to the context in which it was reinforced, or whether some subjects go further and revise their model of the task. Distinguishing these possibilities is the purpose of the classification introduced next and of the analysis in Section 4.

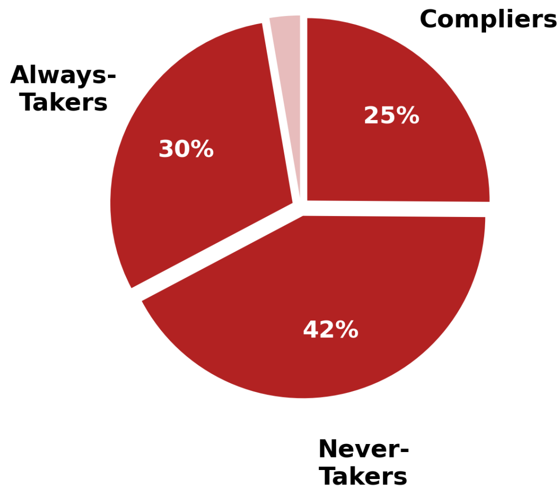
3.4 Classification of Exposure Subjects By Learning Effects

To assess learning gains from observing others' optimal decisions, it is helpful to compare subjects' baseline behavior, when they confronted the task independently, with their choices at the end of the social-learning component. In principle, three learning trajectories may arise. The first consists of subjects who start with a low share of optimal choices but reach high optimality rates once they observe others' decisions; these subjects are referred to as *compliers.* The second consists of subjects who make optimal choices regardless of social exposure; these are labeled *always-takers.* The third consists of subjects who show little or no response to exposure and continue to choose suboptimally even when presented with optimal decisions; these are labeled *never-takers.*[40] To operationalize these learning types, I rely on the following performance measure:

Definition: *Let $y_t \in [0, 1]$ denote the share of optimal choices a subject makes in the final ten*

[40] There is also a fourth possibility in which subjects begin with a high share of optimal choices but later move to a low optimality rate after observing others' optimal decisions. These subjects can be viewed as *defiers.* In this sample, only six out of 223 subjects fall into this category, so I exclude them from subsequent analysis and focus on the three primary groups.

Figure 10: EXPOSURE Learning Type Classification



Notes: The EXPOSURE treatment includes 223 participants. Among them, 56 are classified as Compliers, 67 as Always-Takers, and 94 as Never-Takers. Compliers are participants who performed suboptimally at baseline in Part 1 and optimally in Part 2 when exposed to optimal choices; Always-Takers are those who performed optimally in both parts; and Never-Takers are those who performed suboptimally in both parts. The remaining 6 subjects performed optimally in Part 1 and suboptimally in Part 2 (defiers); they are excluded from the analysis.

key *rounds of Part t. A subject is said to exhibit* ***optimal behavior*** *in Part t if* $y_t \geq 0.8$.

Using this criterion, EXPOSURE subjects can be grouped into three learning types based on how their choices respond to observing optimal decisions: *compliers* satisfy $y_1 < 0.8$ and $y_2 \geq 0.8$, *always-takers* satisfy $y_1 \geq 0.8$ and $y_2 \geq 0.8$, and *never-takers* satisfy $y_1 < 0.8$ and $y_2 < 0.8$. This classification provides a direct mapping between subjects' observed choices and the learning trajectories described above.[41] As Figure 10 shows, approximately 25% of EXPOSURE subjects are classified as compliers, 30% as always-takers, and 42% as never-takers. The main focus of the analysis in the remainder of the paper is on the roughly one-quarter of subjects classified as compliers, whose realized trajectories shift from suboptimal to optimal behavior during the exposure phase, while the remaining subjects either were already making optimal choices or did not cross the threshold. Because the classification conditions on realized Part-2 outcomes within the treated group, it describes realized trajectories rather than a causally identified stratum: in NOEXPOSURE, 13 of 80 subjects (16%) cross the same threshold through individual learning alone, compared with 25% under exposure. The causal statements in this paper accordingly rest on the randomized comparisons of Sections 3.2 and 3.3, with the taxonomy serving as a lens on the heterogeneity

[41] The threshold of 0.8 aligns with the median optimality rate of 80% for EXPOSURE subjects in Part 2 (y_2). Importantly, the qualitative conclusions from the complier analysis in the next section are robust to using alternative thresholds.

behind them.

Result #4: *Exogenously varying exposure to optimal behavior has its largest impact on the one quarter of participants who comply with the treatment and shift from suboptimal to optimal choices. From the perspective of the target subpopulation that behaves suboptimally at baseline (compliers and never-takers), over one-third reach optimal behavior during exposure — well above the 16% that do so without it in* NoExposure.

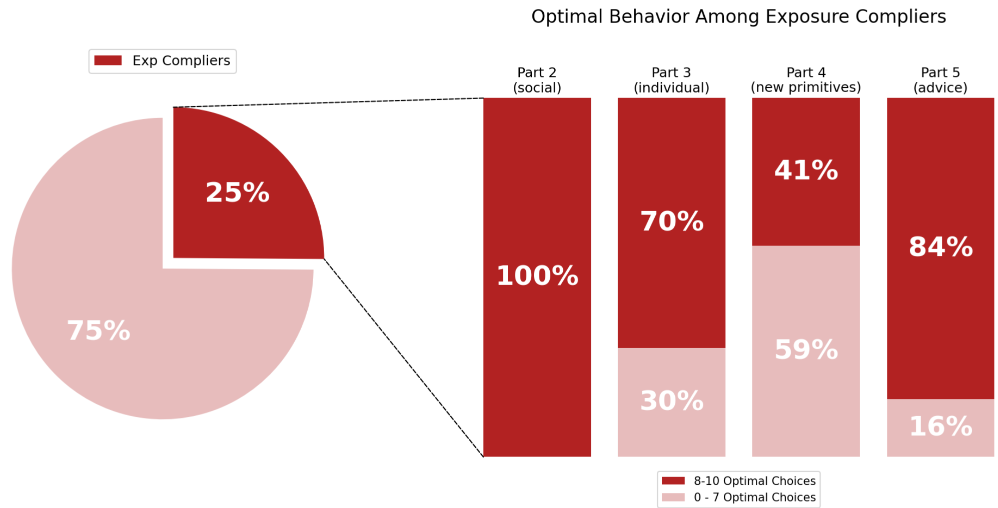
The goal of the next section is to evaluate how much of this treatment effect persists once social exposure is removed, providing insight into the robustness of the learning gains.

4 Learning Gains Beyond Exposure

The analysis so far shows that the positive effect of exposure is not driven by pure imitation: subjects distinguish the quality of the choices they observe, successfully adopting them when exposed to optimal behavior while disregarding external choices when exposed to suboptimal behavior. The dynamics of their behavior point in the same direction — the payoffs from subjects' own choices predict the likelihood of switching to the observed alternative, and that likelihood differs sharply across treatment groups — indicating that subjects engage in selective adoption of observed behavior. The question I tackle next is whether exposure can accomplish more than that: is it powerful enough that participants internalize not only the behavioral pattern but the underlying decision rule, in the sense that their learning gains would be retained beyond exposure and transfer to similar but new decision environments, consistent with the third channel proposed earlier — model revision?

The learning induced by exposure may range from shallow imitation of observed choices, through rules that remain bound to the context in which they were observed, to genuine revision of the underlying model. The main goal of this section is to provide further insights into what subjects actually learn from exposure and to assess whether this learning can be transferred when the primitives of the environment change. In Section 4.1, I examine subjects' decisions once they return to making choices independently in Part 3 of the experiment. In Section 4.2, I study whether subjects continue to behave optimally when they encounter a new hiring task in Part 4, in which the primitives of the environment differ from those they experienced previously. Based on the classification described above, I estimate that roughly 30% of the learning observed among Exposure compliers reflects shallow imitation, another 30% reflects context-bound imitation, and the remaining 40% reflects deeper learning.

Figure 11: Share of Exposure Compliers Exhibiting Optimal Behavior



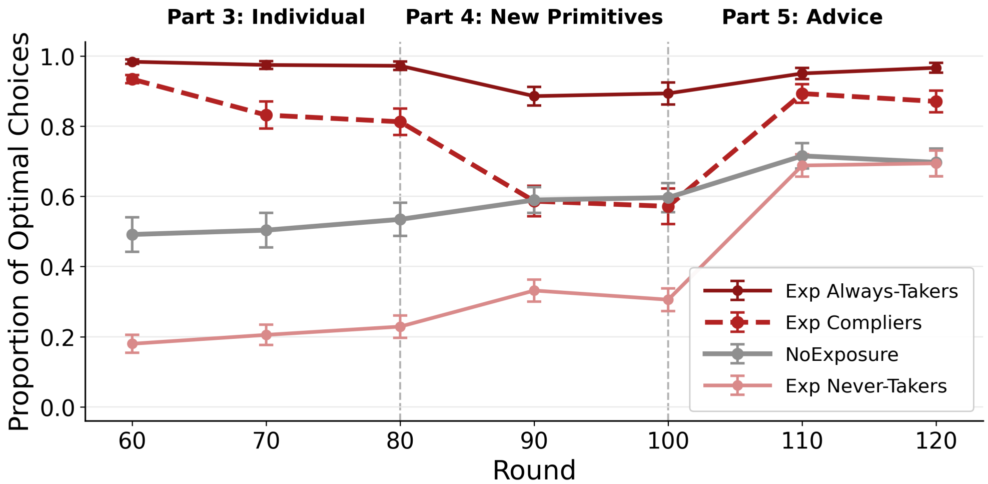
Notes. Compliers in the Exposure group are subjects who performed suboptimally in Part 1 but switched to optimal decisions in Part 2, where optimality is defined as making at least eight optimal choices in the final ten key rounds of a given part. Out of 223 subjects in this group, 56 (25%) meet this criterion. Each bar reports the percentage of these original compliers who continue to exhibit optimal behavior in subsequent parts. Part 3 (Individual) returns subjects to independent decision-making; Part 4 (New Primitives) introduces a change in the environment's underlying primitives; and Part 5 (Advice) provides subjects with written guidance.

4.1 Retention of Learning

The simplest test of the stability of learning gains is to keep the task environment fixed and remove only the exposure to others' decisions. This is precisely what subjects face in Part 3, where they return to making choices independently under the same conditions as in Part 1. For compliers in Exposure, Figure 11 reports the share who continue to follow the optimal decision rule. By construction, all compliers choose optimally by the end of the social-learning phase in Part 2. Once external choices are removed, however, 17 of the 56 compliers (30%) revert to the dominated option. This evidence suggests that once exposure ends, a substantial share of the learning gains dissipates. Because compliers are selected for crossing a threshold in a short window, part of the measured reversion reflects sampling noise rather than behavior: the null benchmarks in Appendix D.4 attribute roughly one fifth of observed reverters to chance threshold-crossing, so the 30% figure is best read as an upper bound on shallow imitation. For most of these compliers, however, the improvements observed during social learning appear to rest on imitation of others' choices, leading to gains that are sizable but ultimately fragile.

To shed light on the intensive margin, Figure 12 plots the share of optimal choices for each Exposure subgroup alongside the NoExposure benchmark. Although compliers reached an average optimality rate of 93% at the end of Part 2 under exposure, this falls to roughly 82% by the

Figure 12: Incidence of Optimal Behavior Beyond Exposure



Notes. This figure displays the evolution of the average proportion of optimal choices made by subjects across Parts 3–5 for the three subgroups within the EXPOSURE treatment, as well as for the NOEXPOSURE control group. Each line tracks the average performance from the last 10 rounds by aggregating individual choices across subjects in a given group. The initial level corresponding to round 60 represents the proportion of optimal choices in the last ten rounds of Part 2 (not shown in the graph). Compliers are subjects who performed suboptimally in Part 1 but switched to optimal decisions in Part 2; Always-Takers are those who performed optimally in both parts; and Never-Takers are those who performed suboptimally in both parts. Part 3 (Individual) returns subjects to independent decision-making; Part 4 (New Primitives) introduces a change in the environment's underlying primitives; and Part 5 (Advice) provides subjects with written guidance. Markers denote group means at each ten-round mark; error bars indicate ± 1 standard error around each estimate.

end of Part 3. Despite this decline, their performance remains well above that of the control group (53%), though markedly below the level observed among always-takers, whose average remains extremely high at 97% ($p < 0.001$, see Table 4). A further noteworthy pattern in Figure 12 is that while EXPOSURE never-takers are defined as having fewer than 80% optimal choices in Parts 1 and 2, their actual average optimality rate in Parts 1–2 is below 20%. For this group, returning to independent decision-making produces little change, consistent with the fact that they did not respond to social exposure in the first place.

In summary, a simple test of learning retention, implemented by returning subjects to individual decision-making once exposure is removed in Part 3, reveals a notable pattern. Among compliers in the EXPOSURE group, who initially switched to optimal behavior, 30% (17 of 56) revert to suboptimal behavior. Correspondingly, their share of optimal choices declines from 93% to 82%. This group is the only one to experience such a decline, while optimal choice proportions among EXPOSURE always-takers and never-takers, as well as NOEXPOSURE subjects, remain essentially unchanged. This suggests that for a substantial subset of subjects, improved performance depends on shallow imitation through continued exposure, and that once this exposure is removed, these

Table 4: Incidence of Optimal Behavior by Exposure Learning Type

	Part 3 (Individual)		**Part 4** (New Primitives)		**Part 5** (Advice)	
	Last 10	**Last 1**	**Last 10**	**Last 1**	**Last 10**	**Last 1**
Compliers	0.289***	0.237***	-0.014	0.025	0.145***	0.286***
	(0.059)	(0.081)	(0.060)	(0.085)	(0.045)	(0.067)
Always-Takers	0.443***	0.443***	0.283***	0.296***	0.226***	0.375***
	(0.048)	(0.062)	(0.048)	(0.067)	(0.038)	(0.054)
Never-Takers	-0.304***	-0.310***	-0.280***	-0.291***	-0.033	0.024
	(0.056)	(0.070)	(0.049)	(0.073)	(0.048)	(0.074)
NoExposure Mean	0.529	0.512	0.610	0.600	0.735	0.625
Observations	2,970	297	2,970	297	2,970	297
H_0: Compl. = AT	$p < 0.001$	$p = 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.009$	$p = 0.021$
H_0: Compl. = NT	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

Notes: Each column reports a separate OLS regression of a round-level optimal choice indicator on Exposure behavior type dummies, with the NoExposure group ($N = 80$) as the reference category. The regression sample consists of 297 subjects: 80 from NoExposure and 217 from Exposure ($223 - 6$ defiers). Compliers are Exposure subjects who performed suboptimally in Part 1 but switched to optimal decisions in Part 2, where optimality is defined as making at least 80% optimal choices in the final ten rounds; Always-Takers performed optimally in both parts; and Never-Takers performed suboptimally in both parts. The Exposure group is classified into 56 Compliers, 67 Always-Takers, 94 Never-Takers, and 6 Defiers (excluded). Part 3 (Individual) returns subjects to independent decision-making; Part 4 (New Primitives) introduces a change in the environment's underlying primitives; and Part 5 (Advice) provides subjects with written guidance. The reported tests assess whether mean optimal choice rates differ across behavioral types within each part. Standard errors are clustered at the subject level and are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

subjects revert toward suboptimal choices rather than internalizing the optimal strategy.

4.2 Transfer of Learning

The results in the previous section show that 70% of Exposure compliers retain their learning gains even after exposure to optimal behavior ends. However, this evidence does not distinguish between two underlying forms of learning. One possibility is that subjects learn to recognize the optimality of observed choices and continue to imitate them during and after exposure, as long as the task environment remains unchanged. A second possibility is that subjects acquire a deeper understanding of why those choices are optimal, allowing them to apply what they have learned to a related task with different underlying primitives. In this section, I examine whether the learning gains observed in the Exposure group transfer to a related but distinct environment, in which subjects face new primitives and therefore cannot directly rely on previously accumulated experience.

Beginning in Part 4 of the experiment, subjects switch to a new hiring task with a different set of primitives. The ability distributions change so that the green group now includes two low-, one medium-, and one high-ability worker, while the orange group includes two low- and two medium-ability workers. In addition, the probability that a medium-ability worker is educated, p_m, falls from 0.9 to 0.1. Together, these changes flip the environment: whereas in earlier parts both workers were educated in most rounds, in Part 4 both workers are now *not* educated in most rounds.[42] All other features remain the same: $p_h = 1$ and $p_l = 0$, and the green group still has higher average ability. Because a not-educated worker cannot be of high ability, the relevant comparison is the posterior probability of medium ability conditional on not being educated, which is higher for the orange group ($\mathbb{P}(m \mid \textit{not educ}) \approx 0.47$ versus ≈ 0.31 for green; see Section 2.2). As a result, conditional on the modal signal, the optimal choice is still to hire an orange worker.

One can see from Figure 11 that among the original EXPOSURE compliers, only 41% continue to exhibit optimal behavior under the new task primitives in Part 4. Their share of optimal choices drops further to 60%, falling below the level observed in NOEXPOSURE, as Figure 12 shows. In other words, despite having reached near-fully optimal performance under exposure, the compliers' optimal choice rate reverts to that of the control group, whose members were never exposed to others' decisions. Because compliers are selected on realized Part-2 outcomes while NOEXPOSURE is an unselected mixture of types, this equality is a descriptive benchmark rather than a complier-specific causal contrast; the causally identified counterpart is the disappearance of the aggregate treatment advantage: taking the EXPOSURE group as a whole, its share of optimal choices in the last ten rounds of Part 4 (57%) is statistically indistinguishable from NOEXPOSURE (61%, $p = 0.344$).

Could the decline simply reflect a harder environment? The structure of the task argues otherwise: the decision problem is unchanged — conditional on the modal signal, the same one-step inference identifies the optimal group — and the new primitives are close to a mirror image of the original ones. More importantly, the groups that never relied on observed choices show no deterioration at all. NOEXPOSURE subjects *improve* from 53% to 61% across the change in primitives (Table 4), never-takers improve from 22% to 33%, and always-takers dip only modestly, from 97% to 89%. A general increase in difficulty would depress all of these trajectories; instead, the sharp decline is specific to compliers. The always-takers' modest dip does suggest that re-mapping a familiar rule onto new signals is not entirely costless, and the possibility that such adjustment costs bind more tightly for recent adopters cannot be fully excluded; but adjustment costs alone cannot explain why compliers fall all the way to the level of a control group that is simultaneously improv-

[42] The analysis for Parts 4 and 5 focuses on these key rounds where both workers are not educated. In the remaining rounds, where at least one worker is educated, nearly all subjects choose the dominant option.

ing. Nor can weaker feedback explain it: the expected per-round payoff difference between the two options is nearly twice as large in Part 4 as in Parts 1–3 ($0.82 versus $0.43; see Section 2.2). On balance, the evidence indicates that learning gains from exposure are largely tied to the context in which the observed choices were adopted, such that the average treatment effect dissipates once the environment primitives change.

Below, I summarize the main result from the analysis of learning gains beyond direct exposure using learning retention and learning transfer as natural benchmarks:

Result #5: *In Part 3, 70% of Exposure compliers continue to behave optimally when returning to individual decision-making, and their average optimality rate remains high at 82%. In Part 4, under new task primitives, only 41% continue to behave optimally, and their optimality rate falls to 60% — statistically indistinguishable from NoExposure subjects, who were never exposed to external choices.*

4.3 Direct Advice

To contrast the effect of exposure to others' decisions with the explicit provision of recommendations, Part 5 of the experiment keeps the new environment primitives but introduces written advice on how to approach the task. This advice is provided by the same participant whose choices subjects observed in Part 2. The purpose of this section is twofold. First, since learning gains among Exposure compliers disappear under the new primitives, as shown in Section 4.2, this part examines whether explicit advice can restore optimal choice rates to previously high levels. Second, the largest subgroup within Exposure consists of never-takers (94 out of 223 subjects), who did not respond to exposure to optimal behavior and maintained optimality rates below 40%. This section therefore examines whether explicit recommendations have a stronger effect than exposure alone for this subgroup.

Recall what this part entails. Before the final 20 rounds, each treated subject reads a written recommendation composed by participant X — the same individual whose round-by-round choices she observed in Part 2. For Exposure subjects, the advice explains how to approach the task and how to extract information from the education signals; it communicates the *why* that observing choices alone could not.[43] The comparison between passive exposure and explicit guidance also

[43] ExposureNeg subjects instead read the recommendation of the participant whose *suboptimal* choices they observed. Explicitly suboptimal advice does not produce a symmetric decline: the ExposureNeg optimality rate in Part 5 (62%) is comparable to its Part 4 level (58%), while NoExposure continues to improve from 61% to 74%, generating a marginally significant 12 percentage point gap ($p = 0.056$) — the only part of the experiment in which the two groups diverge. Suboptimal advice thus operates by failing to support the continued individual learning observed in the control group, rather than by steering subjects toward the dominated choice. In all other post-exposure

anticipates the policy discussion in Section 4.4: the two interventions differ not only in how strongly they move behavior, but in *whom* they reach.

For the Exposure group, the provision of explicit optimal guidance has a substantial effect on both margins. Among compliers, a large fraction respond to the advice: 84% exhibit optimal behavior by making at least eight optimal choices in the last ten rounds of Part 5, and their optimal choice rate rises to 88%, significantly above the NoExposure level, as Table 4 shows ($p = 0.001$).[44] The response among Exposure never-takers is even more striking. On the extensive margin, the share of never-takers exhibiting optimal behavior rises from 0% in Part 2 to 53% in Part 5. On the intensive margin, their optimality rate — which never exceeded 33% in the first four parts of the experiment — more than doubles to 70%, statistically indistinguishable from the control group level, as Figure 12 illustrates. Following the provision of optimal advice, the Exposure group as a whole achieves a 9 percentage point higher optimal choice rate than the control group ($p = 0.020$).[45]

The picture that emerges is consistent across subgroups: never-takers, whom mere exposure to optimal behavior left unmoved, respond strongly once the justification behind the optimal choice is provided. Two features of the design qualify how these gains should be read. Advice arrives in Part 5 for every treated subject — there is no continuation arm that withholds it — and NoExposure itself improves from 61% to 74% over the same rounds, so the Part-5 changes identify the effect of adding advice to each group's accumulated history rather than the standalone effect of advice. The never-takers' jump from at most 33% in earlier parts to 70% is nonetheless far larger than any natural trajectory observed in the data, consistent with a substantial contribution of the explanation itself. The results on explicit guidance are summarized as follows:

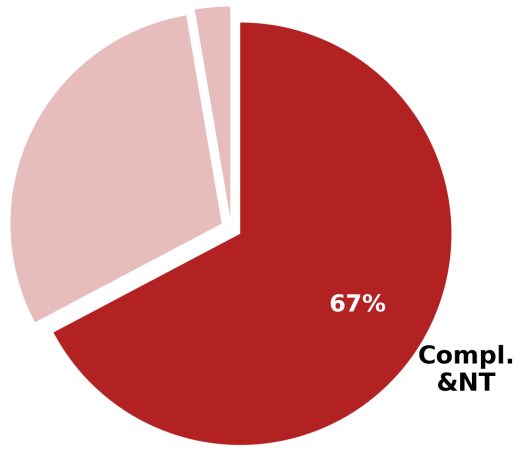
Result #6: *Following explicit advice in Part 5, the optimal choice rate among Exposure compliers returns to its original high level under exposure. The largest response comes from Exposure never-takers: for this subgroup, whom exposure to optimal behavior alone failed to move over 40 rounds of exposure and the 40 rounds that followed, a direct explanation of the optimal approach more than doubles the share of optimal choices to 70%, statistically indistinguishable from the No-Exposure level.*

parts, ExposureNeg behavior remains statistically indistinguishable from NoExposure (Part 3: $p = 0.837$; Part 4: $p = 0.602$).

[44] Always-takers also respond to the advice, reaching a 96% optimality rate in Part 5 compared to 89% in Part 4.

[45] The overall effect of positive recommendation is relatively moderate despite improvements across all Exposure subgroups. This is partly driven by a gradual increase in the NoExposure group to a 74% optimality rate in Part 5, even though these subjects continue to face the task individually and learn only from feedback.

Figure 13: Treatable Sample within the Exposure Group



Notes. This figure highlights the treatable sample within the Exposure group ($N = 223$). Always-takers (67 subjects) are those who performed optimally before exposure and are excluded from the policy-relevant population. The treatable sample consists of 150 subjects: 56 compliers, who switched from suboptimal to optimal behavior under exposure, and 94 never-takers, who did not respond to exposure. Optimality is defined as making at least eight optimal choices in the final ten key rounds of a given part.

4.4 Implications for Policy

The preceding analysis reveals substantial heterogeneity in how subjects respond to exposure and advice. In this section, I synthesize these findings to draw out their implications for intervention design. I begin by defining the population that is plausibly amenable to policy and then develop a taxonomy of learning types that distinguishes subjects by the depth and durability of the behavioral change they exhibit. I then examine how each type responds across the sequence of interventions and discuss what these patterns imply for the design of policies aimed at reducing statistical discrimination.

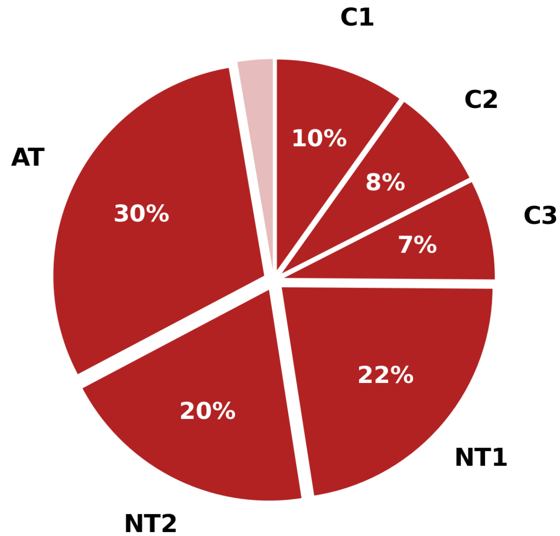
Of the 223 subjects in the Exposure group, 67 are always-takers who already behave optimally before any social exposure, and 6 are defiers who move in the opposite direction; both groups are set aside, the former because they do not require intervention and the latter because of their negligible number. The remaining 150 subjects, 56 compliers and 94 never-takers, constitute the *treatable sample*: individuals who begin the experiment with suboptimal decision-making and are therefore plausible targets of policy interventions (Figure 13). This group exhibits considerable heterogeneity in the depth of learning from exposure and in its responsiveness to explicit guidance, which I now characterize in detail.

Complier subtypes. Among the 56 compliers, the experimental design allows me to distinguish three subtypes based on how their behavior evolves across Parts 3–5. The subtypes are the behavioral footprints of the three learning channels introduced in Section 3.3: deep learners correspond to model revision, context-bound learners to selective adoption that remains anchored to the environment in which it was reinforced, and shallow imitators to adoption sustained only by contemporaneous observation. The first subtype, which I label *Deep Learners* (C1), consists of subjects who retain optimal behavior both after exposure ends in Part 3 and when the environment primitives change in Part 4. This pattern is consistent with having internalized the underlying decision rule, namely, that hiring from the minority group maximizes expected payoff conditional on the modal signal, rather than merely replicating the observed choices. Because the optimal action remains Orange in the key rounds of the new environment, however, continued optimal behavior is also compatible with a simpler carried-over rule, such as choosing Orange whenever the two signals coincide; the Deep Learner label is therefore a provisional name for the strongest observed behavioral pattern, with the response to advice and subjects' own reflections providing complementary, though not decisive, evidence of understanding. With this qualification, Deep Learners represent the strongest form of behavioral change induced by exposure. Of the 56 compliers, 22 subjects (39%) fall into this category.

The second subtype, *Context-Bound Learners* (C2), comprises subjects whose learning persists when the task environment is unchanged (Part 3) but who revert to suboptimal behavior when the underlying primitives change (Part 4). These subjects appear to have learned a strategy that works well in the original context but remains anchored to the specific features of that environment. When the environment shifts, their understanding does not transfer. This group contains 17 subjects (30% of compliers).

The third subtype, *Shallow Imitators* (C3), consists of subjects who revert to suboptimal behavior as soon as direct exposure ends in Part 3. For these individuals, the improvement observed during social learning appears to rest entirely on contemporaneous reinforcement of observed choices, with no durable change in decision-making. This group contains 17 subjects (30% of compliers). Given the size of these subgroups, the subtype-level analysis is best read as a descriptive characterization of the heterogeneity behind the average effects, rather than as a basis for formal between-subtype inference; Appendix D shows that the composition of the taxonomy is stable across alternative classification thresholds and quantifies how much of it could arise from noise alone.

Figure 14: Full Taxonomy of Learning Types in the Exposure Group



Notes. This figure displays the complete classification of subjects in the Exposure group ($N = 223$) into six learning types. Always-Takers (67 subjects) performed optimally before exposure. The remaining 150 subjects form the treatable sample, further classified by their response to exposure and advice. Among compliers (56 subjects): C1 Deep Learners retain optimal behavior across all parts including under new primitives; C2 Context-Bound Learners retain optimality in the original context but revert under new primitives; C3 Shallow Imitators revert as soon as exposure ends. Among never-takers (94 subjects): NT1 Eventual Responders reach optimality only after receiving explicit advice in Part 5; NT2 Non-Responders remain suboptimal throughout. Optimality is defined as making at least eight optimal choices in the final ten key rounds of a given part.

Never-taker subtypes. Among the 94 never-takers, the introduction of explicit advice in Part 5 reveals a further distinction. *Eventual Responders* (NT1) are subjects who do not respond to passive exposure but reach optimality once they receive explicit guidance explaining the rationale behind the optimal strategy. This group comprises 50 subjects (53% of never-takers). Notably, while NT1 subjects are classified as never-takers because they do not cross the optimality threshold in Part 2, they show gradual signs of improvement even before advice is introduced: 12% exhibit optimal behavior in Part 3 and 24% in Part 4. The large jump to 100% optimality in Part 5, however, occurs only after explicit guidance is provided, suggesting that these subjects benefit from a combination of accumulated experience and a clearly articulated rationale. The remaining 44 subjects are *Non-Responders* (NT2), who remain suboptimal throughout all five parts of the experiment despite receiving both exposure and advice. Among NT2 subjects, the share exhibiting optimal behavior never exceeds 2% in any part. These subjects may require fundamentally different forms of intervention, such as changes to the choice architecture itself.

Figure 14 summarizes the full taxonomy. To visualize how each learning type responds across the sequence of experimental interventions, Figure 15 plots the share of subjects exhibiting optimal behavior in each part, separately for each subtype within the treatable sample. The heatmap reveals

Figure 15: Response to Interventions by Learning Type

Notes. This figure displays the share of subjects exhibiting optimal behavior in each part of the experiment, separately for each learning type within the treatable sample of the Exposure group. Rows correspond to the five subtypes: C1 Deep Learners, C2 Context-Bound Learners, and C3 Shallow Imitators among compliers; NT1 Eventual Responders and NT2 Non-Responders among never-takers. Columns correspond to the five experimental parts: Part 1 (Individual Learning), Part 2 (Social Exposure), Part 3 (Individual Learning), Part 4 (New Primitives), and Part 5 (Direct Advice). Cell color reflects the share behaving optimally, ranging from dark red (0%) to dark green (100%). Optimality is defined as making at least eight optimal choices in the final ten key rounds of a given part.

sharply divergent trajectories. Deep Learners (C1) reach 100% optimality in Part 2 and maintain it throughout all subsequent parts, including under the new primitives in Part 4. Context-Bound Learners (C2) follow the same path through Part 3 but drop to 0% under new primitives, before recovering to 82% once explicit advice is provided. Shallow Imitators (C3) revert immediately in Part 3 and remain at 0% until advice restores two-thirds of them to optimality. Among never-takers, Eventual Responders (NT1) show a gradual trajectory, from 0% in Parts 1 and 2, to 12% in Part 3 and 24% in Part 4, culminating in 100% optimality once advice is introduced. Non-Responders (NT2), by contrast, remain at or near 0% throughout. These patterns underscore that exposure and advice operate through distinct channels and that the same intervention can produce qualitatively different responses depending on the depth of a subject's underlying learning process.

Open-ended reflections by subtype. At the end of the experiment, subjects answered two unincentivized open-ended questions: the first asked them to describe the strategy they employed and whether it evolved over the 120 rounds; the second asked specifically how they used participant X's hiring choices in Rounds 21–60 and X's recommendations and explanations in Rounds 101–120 when making their own decisions. The second question is particularly informative, and reading these responses by learning type reveals a striking correspondence between what subjects *say* they did with the observed choices and how their behavior classifies them.

Deep Learners (C1) describe observation as a prompt for understanding, using the language

Table 5: Cumulative Reach of Interventions in the Treatable Sample

Intervention	**Reach**	**Composition**
Exposure alone	37% (56/150)	Compliers: C1 + C2 + C3
Exposure + Advice	65% (97/150)	Adds all NT1, most C2 & C3
Not optimal in Part 5	35% (53/150)	NT2 + some C2, C3

Notes: This table summarizes the cumulative share of the treatable sample (150 subjects: 56 compliers and 94 never-takers) that exhibits optimal behavior under each intervention. "Exposure alone" counts all subjects who reach optimality during the social learning phase in Part 2 (i.e., all compliers). "Exposure + Advice" counts all subjects who exhibit optimal behavior in Part 5, after receiving both exposure and explicit guidance. "Not optimal in Part 5" counts subjects who do not exhibit optimal behavior in the final part: all 44 NT2 subjects, together with 9 compliers who reached optimality earlier but do not return to it under advice. Optimality is defined as making at least eight optimal choices in the final ten key rounds of a given part.

of insight rather than compliance: one writes that following X's choices "*made me realize how the probability worked, and that it was better to choose out of three educated vs four educated*," and another that X's choices "*made me realize that when the given information says educated, it rules out low skill workers, and that when the given information says uneducated, it rules out high skill workers*" — a complete statement of the screening logic at the heart of the task. Their answers to the strategy question tell the same story: one deep learner articulates the full contingent rule, "*When worker is educated I chose the color with more workers in high and medium abilities. When worker was uneducated I chose the color that had least workers as low ability.*" Context-Bound Learners (C2) describe adopting what they saw without extracting a general principle; one reports switching away from green "*after studying participant X's choices*," yet continuing to select green workers under the new primitives "*in the scenario of uneducated workers for both*" until X's written explanation corrected them — precisely the on-support adoption without transfer that defines this subtype. Shallow Imitators (C3) use the language of deference: "*I follow the suggestions*," "*it was the same as participant X's*," or "*I went with participant X's choices and recommendations a lot since it seems like they know the strategy.*" Among never-takers, Eventual Responders (NT1) report discounting the observed actions but being persuaded by the explanation: "*I use X's recommendation more in 101–120 as the explanation make sense but not a lot in 21–60 rounds. Evidence can persuade people more.*" Non-Responders (NT2), finally, express deliberate disregard for both channels: "*I kept my same strategy of picking green workers no matter what*," or "*I looked at participant X's choices and it didn't seem right so I just continued with my strategies.*"

While these responses are self-reported and collected after all decisions were made, their alignment with the behavioral classification is reassuring: the depth of learning inferred from choices alone is mirrored in how subjects themselves describe what they took from observing — and later being advised by — participant X.

Aggregate reach of interventions. Table 5 summarizes the cumulative reach of the two interventions studied in this experiment. Under exposure to the optimal choices of others, roughly 37% of the treatable sample reaches optimal behavior (56 out of 150 subjects), corresponding to all complier subtypes combined (C1, C2, and C3). However, as the analysis above demonstrates, fewer than half of these subjects exhibit learning that is both durable and transferable. When exposure is supplemented with explicit advice in Part 5, the share of the treatable sample exhibiting optimal behavior rises to nearly two-thirds (97 out of 150, or 65%) — a net increase of 41 subjects, as 50 never-takers newly reach optimality while 9 compliers fall back. Counting subjects who reach optimality at any point during exposure or advice, the two interventions together reach 106 of 150 (71%). The largest gains come from two sources: all 50 Eventual Responders (NT1) — by construction the group that responds to advice but not to exposure, so the informative quantity is its size, over half of all never-takers — and the majority of Context-Bound Learners and Shallow Imitators (C2 and C3), who had reverted under changed conditions but return to optimality once guidance is provided. The remaining 53 subjects (35% of the treatable sample) — all 44 Non-Responders (NT2), together with 9 compliers who do not return to optimality under advice — are not at optimality at the end of the experiment.

Implications for intervention design. These findings carry practical implications for policies aimed at improving the quality of decision-making through observational learning. One important application concerns inaccurate statistical discrimination. Consider, for example, a hiring manager who systematically favors candidates from a particular educational background, even when applicants are equally qualified, thereby making a mistake by failing to leverage useful information about candidate quality *conditional on qualifications.* The results suggest that exposure to optimal behavior can serve as an effective, low-cost first intervention, similar to letting a manager observe the hiring decisions of a better-performing peer. However, the gains from such exposure are often fragile and context-specific: a manager may revert to old hiring practices once the peer observation ends, or struggle to apply the lesson when encountering a different pool of applicants. Supplementing peer exposure with explicit guidance substantially extends the reach of the intervention. In the experimental setting, explaining *why* a particular choice is optimal — how conditioning on the common signal reverses the apparent ranking of the two groups — raises the share of the treatable sample at optimality by 27 percentage points, largely by reaching never-takers whom exposure alone had left unmoved; whether advice-induced gains would themselves survive a further change of environment is a question this design leaves open. At the same time, a nontrivial fraction of the population (35%) is not moved to optimality by either passive exposure or explicit recommendations, pointing to the potential value of complementary approaches, such as decision aids or algorithmic evaluation, that embed the conditional reasoning rather than relying on individual learning.

5 Discussion

Observational learning is widespread in workplaces: managers observe how colleagues screen the same job applicants, farmers watch how neighbors cultivate the same crop, doctors see how peers treat the same condition. Organizations rely on it because it is cost-effective and scalable, yet we know surprisingly little about what it actually teaches. If observers internalize *why* the observed decisions are correct, that knowledge travels with them to new situations; if they only learn *what* to do in a particular setting, the gains can vanish as soon as the context changes. The two are indistinguishable while the environment is fixed. This paper separates them experimentally, in a setting where individual feedback alone demonstrably fails to correct a systematic bias in statistical inference.

The experiment delivers three main findings. First, exposure to others' optimal choices substantially improves the quality of subjects' own decisions — a 21 percentage point gap relative to the individual-learning control group by the end of the social learning phase — even though the observed choices are never labeled as optimal. Second, the mechanism is selective adoption rather than imitation: exposure to suboptimal choices leaves behavior statistically indistinguishable from the control group, and within Exposure, subjects try the observed alternative primarily after their own choice delivers a low payoff and retain it when it pays off. Subjects filter what they observe through their own payoff experience, adopting the observed behavior selectively rather than copying it. Third, the resulting gains are fragile. The quarter of subjects who respond to exposure reach near-fully optimal behavior, yet when the environment's primitives change — the same underlying logic, different parameters — their performance reverts to the level of a control group that was never exposed to optimal behavior at all. On aggregate, exposure changes what subjects do, not their model of why it works. Because the design measures choices rather than beliefs, this conclusion about subjects' models is necessarily indirect — inferred from transfer behavior, the response to advice, and subjects' own end-of-experiment reflections rather than from elicited posteriors; belief elicitation at multiple points of such an experiment is a natural extension. What reaches the subjects that observation leaves unmoved is explanation: a short written rationale from the same observed peer — added, in this design, on top of each group's accumulated experience — restores compliers to their earlier high performance and more than doubles the optimality rate of never-takers.

Several additional results sharpen this interpretation. Providing counterfactual payoffs on top of exposure — Exposure (Full Info) — yields no significant additional improvement, indicating that richer feedback adds little once the observed choices are available, though the design bounds only this incremental effect rather than fully separating observation from information. Confidence governs susceptibility rather than gains: exposure to optimal choices helps at all confidence levels, while harm from exposure to suboptimal choices concentrates among less confident subjects, consistent with confidence moderating how subjects evaluate and filter what they observe. Notably, treatment changes behavior without detectably changing stated confidence: subjects who switch to optimal choices report no greater certainty that their approach is correct, another indication that adoption runs ahead of understanding. The heterogeneity across subjects maps cleanly onto the three learning channels: deep learners, whose gains transfer, are the footprint of model revision; context-bound learners and shallow imitators embody selective adoption that never progressed to an updated model of the task; and their own end-of-experiment reflections mirror this classification almost verbatim.

These findings carry a practical message for organizations and policies that lean on observational learning — shadowing, onboarding, sharing of best practices. Such interventions are attractive because they are cheap, and this paper confirms that they work: observing a better-performing peer moves over a third of the population that starts out behaving suboptimally. But the same evidence implies that these gains may not survive a changing environment, precisely because observation spreads *what* to do far more readily than *why* it works. In the context of inaccurate statistical discrimination, for instance, letting a hiring manager observe a better-performing peer is a sensible first intervention, but explaining why the peer's decisions are correct — how conditioning on qualifications reverses the apparent ranking of backgrounds — is what extends the reach to decision-makers whom observation alone leaves unmoved. Whether explanation also makes the improvement portable to new applicant pools is a natural conjecture that this design does not test: advice arrives in the final part, and its gains are never confronted with a further change of environment. For the remainder who respond to neither channel, complementary approaches such as decision aids or algorithmic evaluation that embed the conditional reasoning may be required. Observation, in short, is a powerful way to spread behavior; it is a far weaker way to spread understanding — and, on the evidence here, it is understanding that travels.

References

Agranov, M., Lopez-Moctezuma, G., Strack, P. & Tamuz, O. (2024). Learning Through Imitation: an Experiment. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4083318

Anderson, L. R., & Holt, C. A. (1997). Information Cascades in the Laboratory. *American Economic Review*, *87*(5).

Apesteguia, J., Huck, S. & Oechssler, J. (2007). Imitation—Theory and Experimental Evidence. *Journal of Economic Theory*, *136*(1), 217–235. doi: 10.1016/j.jet.2006.07.006

Araujo, F. A., Wang, S. W. & Wilson, A. J. (2021). The Times They Are A-Changing: Experimenting with Dynamic Adverse Selection. *American Economic Journal: Microeconomics*, *13*(4). doi: 10.1257/mic.20190088

Arrow, K. J. (1973, 1). The theory of discrimination. In *Discrimination in labor markets* (pp. 3–33). Princeton University Press. doi: 10.1515/9781400867066-003

Bandura, A. (1977). *Social Learning Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A., Ross, D. & Ross, S. A. (1963). Vicarious Reinforcement and Imitative Learning. *The Journal of Abnormal and Social Psychology*, *67*(6), 601–607. doi: 10.1037/h0043728

Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. (2013). The Diffusion of Microfinance. *Science*, *341*(6144), 1236498. doi: 10.1126/science.1236498

Banerjee, A. V. (1992). A simple model of herd behavior - JSTOR: The Quarterly Journal of Economics, Vol. 107, No. 3 (Aug., 1992), pp. 797-817. *The Quarterly Journal of Economics*, *107*(3).

Barnett, S. M., & Ceci, S. J. (2002). When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer. *Psychological Bulletin*, *128*(4), 612–637. doi: 10.1037/0033-2909.128.4.612

Barrett, G. F., & Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, *71*(1), 71–104. doi: 10.1111/1468-0262.00390

Bikhchandani, S., Hirshleifer, D. & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5). doi: 10.1086/261849

Björnerstedt, J., & Weibull, J. W. (1996). Nash Equilibrium and Evolution by Imitation. In K. J. Arrow, E. Colombatto, M. Perlman & C. Schmidt (Eds.), *The rational foundations of economic behavior.* Macmillan.

Bohren, J. A., Haggag, K., Imas, A. & Pope, D. (2023). Inaccurate Statistical Discrimination: An Identification Problem. *Review of Economics and Statistics*.

Bohren, J. A., & Hauser, D. N. (2021). Learning With Heterogeneous Misspecified Models: Characterization and Robustness. *Econometrica*, *89*(6). doi: 10.3982/ecta15318

Bohren, J. A., Imas, A. & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review*, *109*(10). doi: 10.1257/aer.20171829

Cai, J., De Janvry, A. & Sadoulet, E. (2015). Social Networks and the Decision to Insure. *American Economic Journal: Applied Economics*, *7*(2), 81–108. doi: 10.1257/app.20130442

Campos-Mercade, P., & Mengel, F. (2024). Non-Bayesian Statistical Discrimination. *Management*

Science, *70*(4). doi: 10.1287/mnsc.2023.4824

Cason, T. N., & Plott, C. R. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, *122*(6). doi: 10.1086/677254

Chaudhuri, A., Graziano, S. & Maitra, P. (2006). Social Learning and Norms in a Public Goods Experiment with Inter-Generational Advice. *Review of Economic Studies*, *73*(2), 357–380. doi: 10.1111/j.1467-937X.2006.0376.x

Conley, T. G., & Udry, C. R. (2010). Learning about a New Technology: Pineapple in Ghana. *American Economic Review*, *100*(1), 35–69. doi: 10.1257/aer.100.1.35

Conlon, J. J., Mani, M., Rao, G., Ridley, M. & Schilbach, F. (2026). Not Learning from Others. *Journal of Political Economy*. (Forthcoming)

Duffy, J., Hopkins, E., Kornienko, T. & Ma, M. (2019). Information choice in a social learning experiment. *Games and Economic Behavior*, *118*. doi: 10.1016/j.geb.2019.06.008

Duflo, E., & Saez, E. (2003). The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, *118*(3), 815–842. doi: 10.1162/00335530360698432

Dupas, P. (2014). Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment. *Econometrica*, *82*(1), 197–228. doi: 10.3982/ECTA9508

Enke, B. (2020). What You See is All There is. *Quarterly Journal of Economics*, *135*(3). doi: 10.1093/qje/qjaa012

Enke, B., & Graeber, T. (2023). Cognitive Uncertainty. *The Quarterly Journal of Economics*, *138*(4), 2021–2067. doi: 10.1093/qje/qjad025

Enke, B., & Zimmermann, F. (2019). Correlation Neglect in Belief Formation. *Review of Economic Studies*, *86*(1). doi: 10.1093/restud/rdx081

Esponda, I., Oprea, R. & Yuksel, S. (2023). Seeing What is Representative. *The Quarterly Journal of Economics*. doi: 10.1093/qje/qjad020

Esponda, I., & Pouzo, D. (2016). Berk-Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models. *Econometrica*, *84*(3). doi: 10.3982/ecta12609

Esponda, I., & Vespa, E. (2014). Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics*, *6*(4). doi: 10.1257/mic.6.4.180

Esponda, I., & Vespa, E. (2018). Endogenous sample selection: A laboratory study. *Quantitative Economics*, *9*(1). doi: 10.3982/qe650

Esponda, I., & Vespa, E. (2024). Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory. *Review of Economic Studies*, *91*(5). doi: 10.1093/restud/rdad102

Esponda, I., Vespa, E. & Yuksel, S. (2024). Mental Models and Learning: The Case of Base-Rate Neglect. *American Economic Review*, *114*(3). doi: 10.1257/aer.20201004

Evdokimov, P., & Garfagnini, U. (2020). Individual vs. Social Learning: An Experiment. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3521347

Eyster, E., Rabin, M. & Weizsacker, G. (2018). An Experiment on Social Mislearning. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2704746

Eyster, E., & Weizsacker, G. (2012). Correlation Neglect in Financial Decision-Making. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1735339

Fudenberg, D., Romanyuk, G. & Strack, P. (2017). Active learning with a misspecified prior. *Theoretical Economics*, *12*(3). doi: 10.3982/te2480

Hanna, R., Mullainathan, S. & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *Quarterly Journal of Economics*, *129*(3). doi: 10.1093/qje/qju015

Heidhues, P., Kőszegi, B. & Strack, P. (2018). Unrealistic Expectations and Misguided Learning. *Econometrica*, *86*(4). doi: 10.3982/ecta14084

Huck, S., Normann, H.-T. & Oechssler, J. (1999). Learning in Cournot Oligopoly—An Experiment. *The Economic Journal*, *109*(454), C80–C95. doi: 10.1111/1468-0297.00418

Hung, A. A., & Plott, C. R. (2001). Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions. *American Economic Review*, *91*(5). doi: 10.1257/aer.91.5.1508

Knowles, J., Persico, N. & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, *109*(1). doi: 10.1086/318603

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *Quarterly Journal of Economics*, *119*(1). doi: 10.1162/003355304772839524

Malmendier, U., & Nagel, S. (2016). Learning from Inflation Experiences. *The Quarterly Journal of Economics*, *131*(1), 53–87. doi: 10.1093/qje/qjv037

Merlo, A., & Schotter, A. (2003). Learning in Elite Tournaments. *Econometrica*, *71*(4), 1443–1489.

Nyarko, Y., Schotter, A. & Sopher, B. (2007). On the Informational Content of Advice: A Theoretical and Experimental Study. *Economic Theory*, *33*(3), 519–536. doi: 10.1007/s00199-005-0037-7

Offerman, T., & Schotter, A. (2009). Imitation and Luck: An Experimental Study on Social Sampling. *Games and Economic Behavior*, *65*(2), 461–502. doi: 10.1016/j.geb.2008.03.004

Phelps, E. S. (1972). The Statistical theory of Racism and Sexism. *American Economic Review*, *62*(4).

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., . . . Laland,

K. N. (2010). Why Copy Others? Insights from the Social Learning Strategies Tournament. *Science*, *328*(5971), 208–213. doi: 10.1126/science.1184719

Schlag, K. H. (1998). Why Imitate, and If So, How?: A Boundedly Rational Approach to Multi-armed Bandits. *Journal of Economic Theory*, *78*(1), 130–156. doi: 10.1006/jeth.1997.2347

Schlag, K. H. (2010). Imitation and Social Learning. In *Encyclopedia of the sciences of learning.* Springer.

Schotter, A., & Sopher, B. (2003). Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study. *Journal of Political Economy*, *111*(3), 498–529. doi: 10.1086/374187

Simonsohn, U., Karlsson, N., Loewenstein, G. & Ariely, D. (2008). Direct Risk Aversion: Evidence from Risky Prospects Valued Below Their Worst Outcome. *Psychological Science*, *19*(4), 320–325.

Singley, M. K., & Anderson, J. R. (1989). *The Transfer of Cognitive Skill.* Cambridge, MA: Harvard University Press.

Smith, L., & Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, *68*(2). doi: 10.1111/1468-0262.00113

Spiegler, R. (2020). Behavioral implications of causal misperceptions. *Annual Review of Economics*, *12*. doi: 10.1146/annurev-economics-072219-111921

Thorndike, E. L., & Woodworth, R. S. (1901). The Influence of Improvement in One Mental Function upon the Efficiency of Other Functions. *Psychological Review*, *8*(3), 247–261. doi: 10.1037/h0074898

Vega-Redondo, F. (1997). The Evolution of Walrasian Behavior. *Econometrica*, *65*(2), 375–384. doi: 10.2307/2171898

Weizsäcker, G. (2010). Do we follow others when we should? A simple test of rational expectations. *American Economic Review*, *100*(5). doi: 10.1257/aer.100.5.2340

Online Appendix for

Fragile Learning From Others

Alisher Batmanov
UC San Diego

Contents:

A. Additional Analysis: Treatment Differences

This appendix provides a comprehensive set of statistical tests complementing the main treatment comparisons discussed in Section 3.2. I present results for all four treatment groups, with NoExposure ($N = 80$) as the omitted category, alongside Exposure ($N = 223$), ExposureNeg ($N = 53$), and Exp+F ($N = 44$). Throughout, I focus on optimal choice behavior in Part 2 of the experiment, measured over the last 10 rounds of interest and in the last round.[46]

A.1. OLS Regression Estimates

Table 6 reports OLS estimates of treatment effects on optimal choice in Part 2 under several specifications. Columns (1)–(3) use the round-level binary optimal choice indicator pooled across the last 10 rounds of Part 2, with standard errors clustered at the subject level; Columns (4)–(6) use the subject-level binary optimal choice in the last round, with heteroskedasticity-robust standard errors.

The baseline specification in Column (1) confirms the main finding from Section 3.2: Exposure subjects make 12.8 percentage points more optimal choices than NoExposure subjects ($p = 0.022$), while ExposureNeg subjects perform indistinguishably from NoExposure ($p = 0.935$). The Exp+F group shows an even larger improvement of 20.6 percentage points ($p = 0.005$), although the difference between Exposure and Exp+F is not statistically significant ($p = 0.203$), consistent with the interpretation that additional feedback does not substantially augment the effects of social exposure alone.

These results are robust to controlling for baseline behavior: adding the Part 1 optimal choice share in Column (2) explains substantial additional variation (R^2 increases from 0.021 to 0.334) but leaves the treatment effect estimates virtually unchanged. The strong predictive power of baseline behavior indicates substantial persistence in individual performance — whether rooted in prior understanding, stable heuristics, or attention — but it does not confound the treatment comparisons. Adding demographic controls in Column (3) for gender, age, and major further confirms this stability: none of the demographic variables are individually significant, and the treatment coefficients remain nearly identical.

[46] The four treatment groups are balanced on observable pre-treatment characteristics. Baseline performance in Part 1 is nearly identical across groups: 39.8% (NoExposure), 40.5% (Exposure), 40.2% (ExposureNeg), and 39.5% (Exp+F); a Kruskal-Wallis test fails to reject equality ($p = 0.989$). The groups are also balanced on gender ($p = 0.392$) and major ($p = 0.584$). Age shows a marginally significant difference ($p = 0.028$), driven by a slightly higher mean in Exp+F (20.6 years) relative to the other groups (19.7–19.9 years); however, as shown in Table 6, adding age as a control does not affect the treatment effect estimates.

Table 6: OLS Estimates of Treatment Effects on Optimal Choice in Part 2

	Last 10 Rounds			Last Round		
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure	0.128**	0.123***	0.122***	0.210***	0.205***	0.203***
	(0.056)	(0.038)	(0.038)	(0.064)	(0.051)	(0.051)
ExposureNeg	−0.006	−0.009	−0.007	0.009	0.006	0.011
	(0.078)	(0.048)	(0.048)	(0.089)	(0.063)	(0.063)
Exposure+Feedback	0.206***	0.208***	0.212***	0.197**	0.198***	0.209***
	(0.073)	(0.056)	(0.057)	(0.091)	(0.076)	(0.076)
Part 1 Optimal Share		0.659***	0.650***		0.660***	0.655***
		(0.033)	(0.035)		(0.039)	(0.040)
Female			0.011			0.050
			(0.032)			(0.040)
Age			−0.008			−0.013
			(0.007)			(0.010)
Econ Major			−0.044			−0.043
			(0.032)			(0.040)
Constant	0.491***	0.229***	0.416***	0.463***	0.200***	0.451**
	(0.049)	(0.036)	(0.145)	(0.056)	(0.046)	(0.201)
R^2	0.021	0.334	0.337	0.039	0.357	0.364
Observations	4,000	4,000	4,000	400	400	400
Equality Tests (p-values)						
Exp = Exp+F	0.203	0.108	0.092	0.863	0.919	0.926
ExpNeg = NoExp	0.935	0.849	0.891	0.918	0.921	0.860
Exp = ExpNeg	0.046	0.003	0.004	0.008	<0.001	<0.001

Notes: This table reports OLS estimates of the effect of treatment assignment on the share of optimal choices in Part 2 of the experiment, with NoExposure as the omitted category. In Columns (1)–(3), the dependent variable is the round-level binary optimal choice indicator pooled across the last 10 rounds of Part 2, with standard errors clustered at the subject level. In Columns (4)–(6), the dependent variable is the binary optimal choice in the last round of Part 2, with heteroskedasticity-robust (HC1) standard errors. *Part 1 Optimal Share* is the proportion of optimal choices in the last 10 rounds of Part 1. *Econ Major* equals one if the subject's major is Economics, Business Economics, or Joint Mathematics-Economics. The equality tests report p-values from F-tests of the stated null hypotheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The last-round estimates in Columns (4)–(6) are qualitatively similar but larger in magnitude, reflecting the fact that many Exposure subjects switch to optimal behavior gradually over Part 2, so that the last-round comparison captures the full cumulative treatment effect. The treatment effect of Exposure in the last round is 21.0 percentage points ($p = 0.001$), and the effect is robust to the inclusion of baseline and demographic controls. The equality test confirms that Exposure and ExposureNeg produce significantly different outcomes ($p = 0.008$), while Exposure and Exp+F do not ($p = 0.863$).

A.2. Alternative Specifications and Nonparametric Tests

Table 7 examines whether the main findings are robust to alternative definitions of the dependent variable and to distributional considerations. Columns (1)–(2) use a binary indicator that equals one if a subject makes at least 8 optimal choices out of 10 in the last 10 rounds of Part 2, capturing whether a subject achieves a high level of optimality rather than the continuous share. Under this binary measure, 55.2% of Exposure subjects achieve high optimality compared with 42.5% in NoExposure, a 12.7 percentage point difference ($p = 0.051$). The ExposureNeg rate (41.5%) is virtually identical to NoExposure ($p = 0.910$), and Exp+F (54.5%) is indistinguishable from Exposure ($p = 0.941$). These patterns are unchanged when controlling for baseline behavior in Column (2).

Column (3) reports median regression estimates of the subject-level share of optimal choices in the last 10 rounds of Part 2. The median regression confirms that Exposure shifts the median optimal choice share from 0.50 to 0.80 ($p < 0.001$), while Exp+F shifts it further to 0.90 ($p < 0.001$). The ExposureNeg effect on the median is negative but not significant (-0.10, $p = 0.340$). The equality test for Exposure vs. ExposureNeg is highly significant ($p < 0.001$), confirming that the distributional shift is concentrated in the positive exposure treatments.

The parametric results are corroborated by distribution-free tests.[47] The first-order stochastic dominance of Exposure over NoExposure, documented using the Barrett & Donald (2003) procedure in Section 3.2, further supports the conclusion that the treatment shifts the entire distribution of optimal behavior rather than affecting only a particular quantile.

[47] The Wilcoxon rank-sum test (Mann-Whitney U) confirms that Exposure differs significantly from NoExposure in the last 10 rounds ($p = 0.022$) and in the last round ($p = 0.001$), while ExposureNeg does not ($p = 0.914$ and $p = 0.920$, respectively). Fisher's exact test on the binary optimality indicator yields consistent results for the last 10 rounds (Exposure vs. NoExposure, $p = 0.068$; ExposureNeg vs. NoExposure, $p = 1.000$) and for the last round (Exposure vs. NoExposure, $p = 0.001$; ExposureNeg vs. NoExposure, $p = 1.000$; Exposure vs. ExposureNeg, $p = 0.011$). All pairwise comparisons between Exposure and Exp+F are insignificant under both tests.

Table 7: Alternative Specifications: Binary Outcome and Median Regression

	Binary Optimality		Median
	(1)	(2)	(3)
Exposure	0.127*	0.121**	0.300***
	(0.065)	(0.048)	(0.077)
ExposureNeg	−0.010	−0.013	−0.100
	(0.088)	(0.060)	(0.105)
Exposure+Feedback	0.120	0.122	0.400***
	(0.094)	(0.076)	(0.111)
Part 1 Optimal Share		0.704***	
		(0.040)	
Constant	0.425***	0.145***	0.500***
	(0.056)	(0.042)	(0.066)
Observations	400	400	400
Equality Tests (p-values)			
Exp = Exp+F	0.941	0.990	0.306
ExpNeg = NoExp	0.910	0.828	0.340
Exp = ExpNeg	0.073	0.014	<0.001

Notes: Columns (1)–(2) report OLS estimates where the dependent variable equals one if a subject makes at least 8 optimal choices out of 10 in the last 10 rounds of Part 2, with heteroskedasticity-robust (HC1) standard errors and NoExposure as the omitted category. Column (3) reports median (50th percentile) regression estimates of the subject-level share of optimal choices in the last 10 rounds of Part 2. The equality tests for Columns (1)–(2) report p-values from F-tests; Column (3) reports p-values from Wald tests based on the quantile regression variance-covariance matrix.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

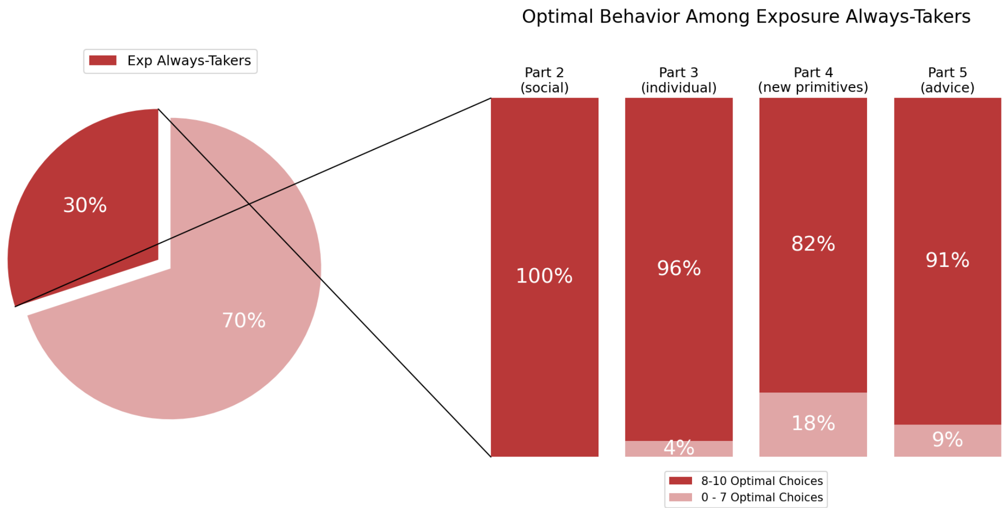
B. Additional Analysis: Always-Takers and Never-Takers

Section 3.4 classifies Exposure subjects into compliers, always-takers, never-takers, and defiers based on whether their behavior changes between the baseline and social learning phases. The main text documents complier dynamics in detail; this appendix completes the decomposition by examining always-takers and never-takers in both the Exposure ($N = 223$) and Exp+F ($N = 44$) treatments. These two types are of substantive interest because they represent the boundaries of what social exposure can achieve: always-takers reveal that a substantial fraction of subjects already behave optimally before treatment, while never-takers identify the subjects for whom passive observation is insufficient.

Always-takers. In the Exposure group, 67 subjects (30%) are classified as always-takers. These subjects performed optimally in both Part 1 and Part 2, meaning they had already internalized the correct decision rule prior to social exposure. Figure 16 displays their dynamics across subsequent parts. The key finding is that always-takers maintain remarkably high optimality rates throughout: 96% in Part 3 (individual decision-making without exposure), 82% in Part 4 (new environment primitives), and 91% in Part 5 (advice). The decline to 82% in Part 4 is notable because it indicates that even subjects with strong prior understanding face difficulty when the task environment changes. The fact that over four-fifths of always-takers remain optimal under the new primitives is consistent with comprehension of the underlying statistical logic, though threshold-based choices alone cannot rule out alternatives such as a stable Orange-favoring heuristic combined with feedback-based relearning. The recovery to 91% under advice in Part 5 is likewise consistent with the Part 4 decline reflecting the difficulty of adjusting to the new environment rather than a permanent loss of understanding.

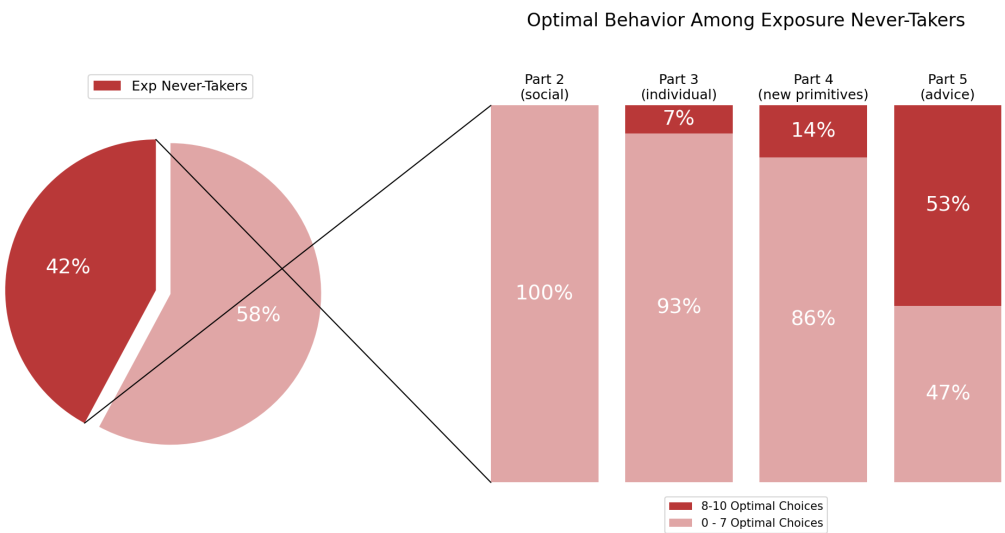
Never-takers. In the Exposure group, 94 subjects (42%) are classified as never-takers — the largest learning type. Despite observing participant X's optimal choices throughout Part 2, these subjects did not switch to optimal behavior. Figure 17 reveals a striking pattern in their subsequent performance. Never-takers remain largely suboptimal in Part 3 (7%) and Part 4 (14%), confirming that they did not internalize the optimal rule during the social learning phase. However, their behavior changes dramatically in Part 5: 53% of never-takers achieve optimal performance once explicit written advice is provided. This finding carries two important implications. First, never-takers are not fundamentally incapable of learning the optimal rule; rather, passive observation of another agent's choices is insufficient to shift their behavior, while direct written guidance is effective. Second, the fact that nearly half of never-takers still fail to achieve optimality even under explicit advice (47% remain suboptimal in Part 5) suggests substantial heterogeneity in the barriers

Figure 16: Exposure Always-Taker Classification and Dynamics



Notes. Always-takers in the Exposure group ($N = 67$, 30% of Exposure) are subjects who performed optimally in both Part 1 (baseline) and Part 2 (social learning), where optimality is defined as making at least 8 optimal choices in the final 10 key rounds. Each bar reports the percentage of always-takers who exhibit optimal behavior in subsequent parts.

Figure 17: Exposure Never-Taker Classification and Dynamics



Notes. Never-takers in the Exposure group ($N = 94$, 42% of Exposure) are subjects who performed suboptimally in both Part 1 and Part 2. Each bar reports the percentage who exhibit optimal behavior in subsequent parts.

to learning: for some subjects, the barrier may lie not in the mode of information transmission but in processing the statistical reasoning itself, though inattention to or distrust of the advice cannot be ruled out on this evidence.

C. Additional Analysis: Exposure (Full Info)

C.1. Aggregate Treatment Effects

A natural question that arises in relation to the learning mechanism in the Exposure treatment is that learning gains may not be solely attributable to observing another participant's choices. Subjects never observe participant X's payoffs; however, because exposure induces some subjects to adopt the observed optimal action, the payoff feedback they generate for *themselves* becomes endogenously richer: a subject who tries the optimal action experiences its payoff distribution directly, which a suboptimal chooser learning alone encounters less often. As a result, the improvements documented in Section 3.2 could reflect a combination of social learning and this endogenously enhanced feedback, making it difficult to isolate the pure effect of exposure.

This concern relates to a broader issue studied in the literature on learning from endogenously selected samples. As Esponda & Vespa (2018) demonstrate, when agents' own choices determine the data they observe, learning can be severely hindered because the resulting feedback is not representative of what a different choice would have produced. In the present setting, subjects in Exposure face a version of this problem in reverse: because exposure shifts choices toward the optimal action, the feedback stream subjects generate for themselves is systematically more favorable than what a suboptimal chooser would encounter alone. To the extent that this enhanced feedback, rather than the act of observing optimal choices per se, drives the treatment effect, the interpretation of the main results would shift from social learning toward a form of information provision.

To disentangle these channels, the experimental design includes an additional treatment group, Exposure (Full Info) (**Exp+F**), depicted in Figure 18. Whereas subjects in Exposure receive exposure *to choices* only, subjects in Exp+F receive exposure *to choices and payoffs*: they observe the same optimal choices of participant X and are additionally provided with payoff-relevant information about the counterfactual outcome. Specifically, regardless of whether a subject's own choice is optimal or suboptimal, subjects in this treatment observe the payoff that would have resulted from the alternative hiring decision. This design eliminates the endogenous feedback channel: if the main treatment effect in Exposure is driven by richer payoff information rather than by exposure to others' choices, then providing all subjects with this information directly should lead to substantially larger effects in Exp+F relative to Exposure.

Figure 18: The Exposure (Full Info) Treatment

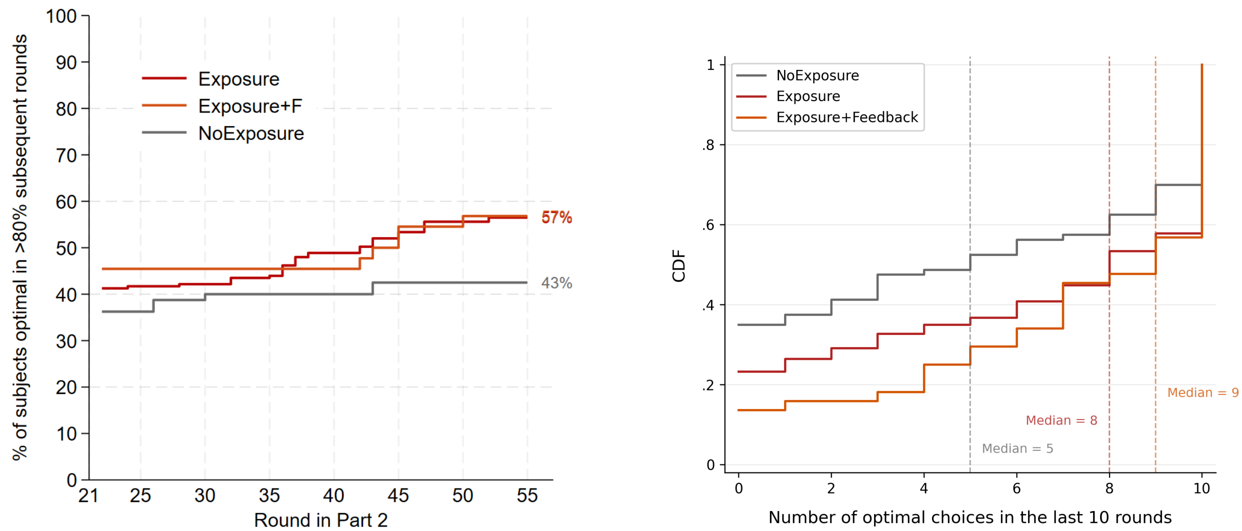
Figure 19 plots the cumulative distribution of optimal choices in the last ten rounds of Part 2 for the three relevant treatment groups. Over the last ten rounds of Part 2, the mean optimal choice rate in Exp+F is 70%, compared with 62% in Exposure and 49% in NoExposure. While Exp+F subjects perform slightly better on average, the difference between Exp+F and Exposure is not statistically significant (7.8 percentage points, 95% confidence interval $[-4.3, +19.9]$, $p = 0.204$); the comparison thus rules out large incremental gains from counterfactual information conditional on observation, though not moderate ones. By the final round of Part 2, the two exposure groups are virtually indistinguishable: 67% of Exposure subjects and 66% of Exp+F subjects make the optimal choice ($p = 0.863$). In contrast, both exposure treatments produce higher optimal choice rates than NoExposure, with the difference between Exp+F and NoExposure reaching significance in the last ten rounds ($p = 0.005$) and the difference between Exposure and NoExposure reaching significance in the last round ($p = 0.001$).[48]

These results suggest that the endogenous feedback channel plays a limited role in explaining the treatment effects documented in the main analysis. Providing subjects with richer counterfactual information does not substantially increase the rate of optimal choices beyond the level achieved through exposure alone. The finding that both Exposure and Exp+F produce similar improvements relative to NoExposure is consistent with the interpretation that the main mechanism underlying the treatment effect is the observation of optimal choices, rather than the informativeness of the associated feedback.

An additional piece of evidence comes from the complier classification applied to the Exp+F group. Using the same 80% threshold as in Section 3.4, the distribution of learning types in Exp+F closely mirrors that in Exposure. As Figure 20 shows, the share of compliers, always-takers, and never-takers is comparable across the two treatments, and the trajectory of complier

[48]The difference between Exposure and NoExposure in the last ten rounds has $p = 0.022$, while the difference between Exp+F and NoExposure in the last round has $p = 0.031$.

Figure 19: Effects of the Exposure (Full Info) Treatment



(a) Dynamics of Optimal Behavior

(b) Cumulative Distribution of Optimal Choices

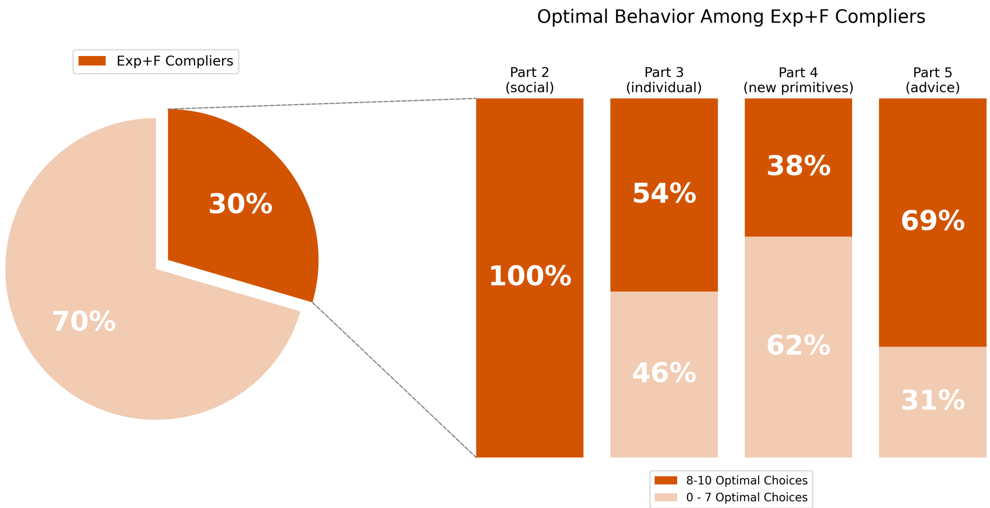
Notes: The red, orange, and gray lines correspond to Exposure ($N = 223$), Exp+F ($N = 44$), and NoExposure ($N = 80$), respectively. Panel (a) displays the percentage of subjects choosing the optimal alternative in a given round of Part 2 and then continuing to do so in at least 80% of subsequent rounds. Panel (b) plots the cumulative distribution of the number of optimal choices in the last ten rounds of Part 2 for each group. The vertical dashed lines denote the median for each group. In Exp+F, subjects receive the same social exposure as in Exposure but additionally observe the payoff that would have resulted from the alternative hiring decision.

behavior in Parts 3–5 follows a similar pattern to that documented for Exposure in Section 4: the share of compliers who retain optimal behavior declines once exposure is removed and falls further under the new primitives, before recovering under advice.[49] This parallel suggests that the type of learning induced by exposure is qualitatively similar regardless of whether counterfactual feedback is additionally provided.

The evidence from Exp+F thus serves as a useful robustness check for the main results. The fact that providing additional counterfactual feedback does not meaningfully improve performance beyond what exposure to optimal choices already achieves suggests that the learning gains in Exposure are primarily driven by the social component of the treatment, namely the observation of another participant's optimal decisions, rather than by the informativeness of the feedback environment. One caveat is in order: the design does not include a feedback-only cell in which subjects receive counterfactual payoff information *without* observing another participant's choices, so the comparison bounds the incremental effect of richer information conditional on observation rather than separating the two channels entirely. Two features of the data nevertheless speak against a pure information-provision account: the observed choices are what identify the alternative worth

[49]The detailed classification of Exp+F always-takers and never-takers is provided in the next subsection.

Figure 20: Exp+F Complier Classification and Dynamics



Notes: Compliers in the Exp+F group ($N = 44$) are subjects who performed suboptimally in Part 1 but switched to optimal decisions in Part 2, where optimality is defined as making at least eight optimal choices in the final ten key rounds of a given part. The pie chart shows the distribution of learning types. Each bar reports the percentage of these original compliers who continue to exhibit optimal behavior in subsequent parts. In the Exp+F treatment, subjects receive the same social exposure as in Exposure but additionally observe the payoff that would have resulted from the alternative hiring decision.

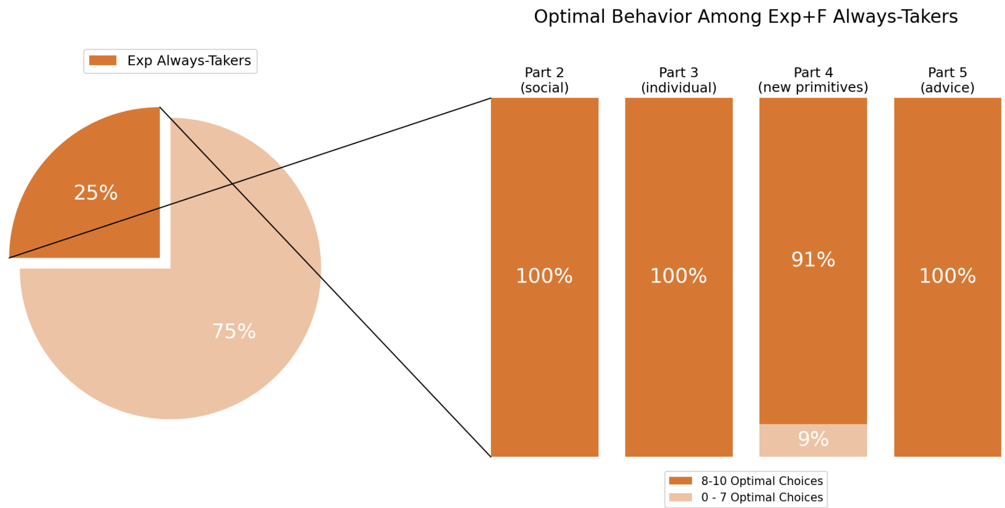
trying — the asymmetry by baseline behavior in Table 2 shows that subjects move only when the observed behavior contradicts and outperforms their own — and the transition evidence shows that adoption is timed by subjects' own payoff realizations rather than by information about the alternative, which is constant across rounds.

Additional Result (Exposure Full Info): *Providing counterfactual payoffs on top of exposure does not significantly improve behavior relative to exposure alone, indicating that richer feedback adds little once observed choices are available; absent a feedback-only cell, the comparison bounds this incremental effect rather than separating observation from information entirely.*

C.2. Learning Types

The Exp+F treatment provides subjects with the same social exposure as Exposure but additionally reveals the payoff that would have resulted from the alternative hiring decision. If counterfactual feedback accelerates learning, we would expect a higher complier share and different dynamics for the remaining types. The classification of Exp+F subjects yields a distribution that is broadly similar to Exposure: 30% compliers ($N = 13$), 25% always-takers ($N = 11$), 45% never-takers ($N = 20$), and no defiers, although the smaller sample size ($N = 44$) limits the precision of these comparisons.

Figure 21: Exp+F Always-Taker Classification and Dynamics



Notes. Always-takers in the Exp+F group ($N = 11$, 25% of Exp+F) are subjects who performed optimally in both Part 1 and Part 2. Each bar reports the percentage who exhibit optimal behavior in subsequent parts.

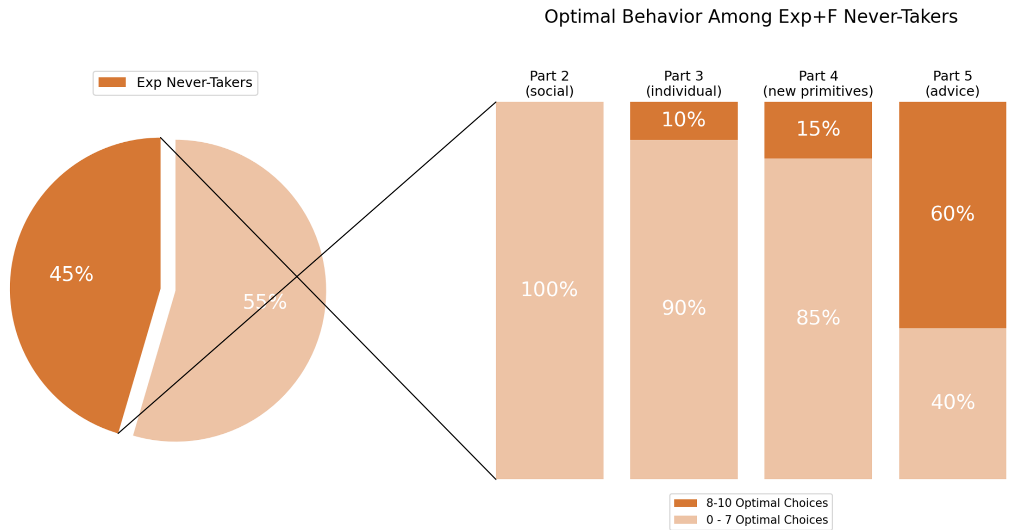
Always-takers. Figure 21 shows the dynamics of the 11 always-takers in Exp+F. Their performance is remarkably stable: 100% remain optimal in Parts 3 and 5, with only a slight decline to 91% in Part 4 under the new primitives. This stability closely parallels the Exposure always-takers (96%, 82%, 91%), with the Exp+F group performing slightly better in each part. While the small sample size prevents strong conclusions, the pattern is consistent with the view that counterfactual feedback may reinforce the understanding of subjects who already behave optimally, helping them maintain performance even when the environment changes. Alternatively, the difference may simply reflect sampling variation in a small group.

Never-takers. Figure 22 displays the dynamics of the 20 never-takers in Exp+F. The trajectory is qualitatively similar to Exposure: low optimality in Part 3 (10%) and Part 4 (15%), rising to 60% in Part 5 under advice. The Part 5 recovery rate is somewhat higher than for Exposure never-takers (60% vs. 53%), which could suggest that prior exposure to counterfactual feedback primes these subjects to be more receptive to advice. However, the difference is modest and the small sample size precludes definitive conclusions. The key qualitative finding is consistent across both treatments: never-takers are not permanently resistant to learning, but they require a more direct form of information transmission than passive social observation.

C.3. Summary

In sum, the always-taker and never-taker analyses of this and the preceding appendix reinforce and extend the main findings in several ways. First, the existence of a large always-taker group

Figure 22: Exp+F Never-Taker Classification and Dynamics



Notes. Never-takers in the Exp+F group ($N = 20$, 45% of Exp+F) are subjects who performed suboptimally in both Part 1 and Part 2. Each bar reports the percentage who exhibit optimal behavior in subsequent parts.

(30% in Exposure, 25% in Exp+F) confirms that a substantial fraction of the population can solve the task without any social learning, consistent with the theoretical prediction that Bayesian reasoning is accessible to some subjects even in the absence of external information. Second, always-takers' high persistence rates across Parts 3–5 are consistent with an understanding of the underlying structure: these subjects maintain optimal behavior even when the environment changes (Part 4) and advice is provided (Part 5). Third, the never-taker response to advice in Part 5 (53% in Exposure, 60% in Exp+F) demonstrates that this bundled intervention reaches many subjects whom passive observation did not; because the advice combines an explanation with explicit scenario-specific recommendations and arrives after substantial additional experience, the response does not by itself distinguish comprehension from reduced computational demands or simple compliance. Finally, the qualitative similarity of these patterns across Exposure and Exp+F reinforces the conclusion from the aggregate analysis above that additional counterfactual feedback does not fundamentally alter the nature of learning induced by social exposure.

D. Additional Analysis: Alternative Complier Specifications

The main analysis classifies Exposure subjects as compliers, always-takers, or never-takers using a threshold of 0.8 (at least eight optimal choices in the final ten key rounds of a given part). While this threshold is standard in the experimental literature on binary repeated choices, the classification and resulting dynamics could in principle be sensitive to its exact value. This appendix examines the robustness of the complier classification under four alternative specifications: a more lenient threshold of 0.6, a stricter threshold of 1.0, and two criteria that combine a level requirement with a minimum improvement relative to baseline. The central question is whether the patterns documented in Section 4 — declining optimality when the environment changes, followed by recovery under advice — are a robust feature of complier behavior or an artifact of the particular threshold used.

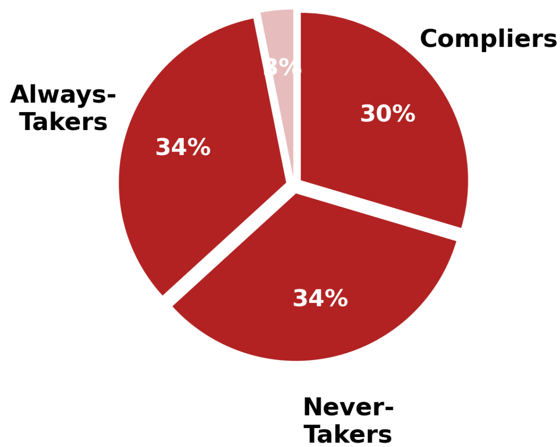
D.1. Threshold 0.6

Under the 0.6 threshold, a subject is classified as exhibiting optimal behavior if at least 6 of 10 choices in the final ten key rounds are optimal. This more lenient criterion expands the set of subjects counted as having "learned" and correspondingly changes the composition of each type. Figure 23 displays the resulting distribution: 30% compliers ($N = 66$), 34% never-takers ($N = 75$), 34% always-takers ($N = 75$), and 3% defiers ($N = 7$). Compared with the baseline 0.8 classification (25% compliers, 42% never-takers, 30% always-takers), the lower threshold reclassifies some never-takers as compliers (the complier share rises from 25% to 30%) and reclassifies some subjects from never-takers to always-takers (the always-taker share rises from 30% to 34%), reflecting subjects who achieved 6–7 optimal choices in both parts.

The critical test is whether the complier dynamics remain qualitatively similar. Figure 24 shows that compliers under the 0.6 threshold exhibit the same U-shaped pattern documented in the main text: 74% remain optimal in Part 3, declining to 50% in Part 4, and recovering to 94% in Part 5 under advice. The Part 4 decline is somewhat less severe than under the 0.8 threshold (50% vs. 41%), which is expected: the broader criterion includes subjects who barely crossed the threshold in Part 2 and are therefore closer to suboptimal, but it also includes subjects who were just below the 0.8 bar and may have still partially learned. The Part 5 recovery rate (94%) is higher than the baseline (84%), though because lowering the cutoff changes both the composition of the complier group and the criterion applied to later behavior, such cross-threshold differences should not be attributed to the marginal subjects themselves.

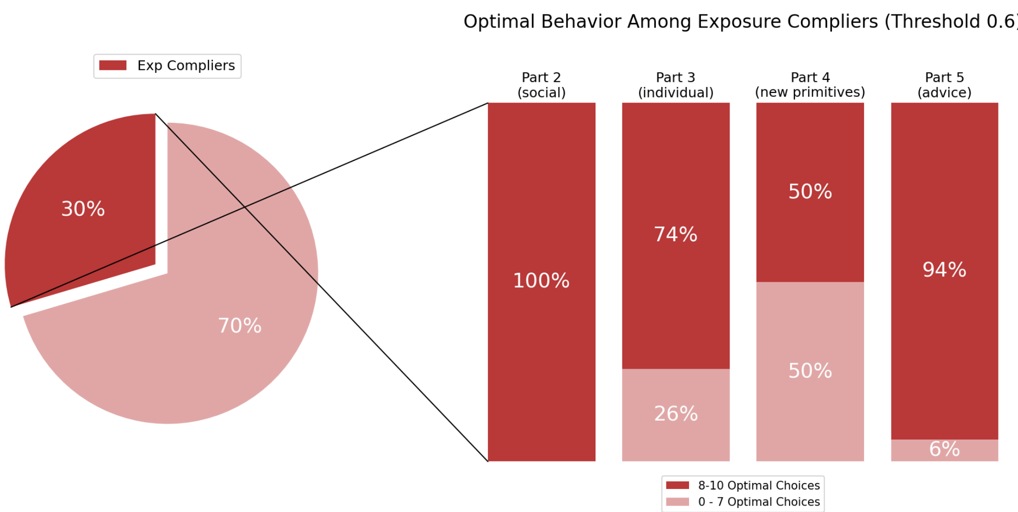
Figure 25 presents the round-by-round evolution of optimal choice rates across Parts 3–5 for

Figure 23: Exposure Learning Type Classification under Threshold 0.6



Notes. Classification of Exposure subjects ($N = 223$) into learning types using an alternative optimality threshold of 0.6. A subject exhibits optimal behavior if at least 6 of 10 choices in the final ten key rounds are optimal. Compliers switched from suboptimal in Part 1 to optimal in Part 2; always-takers performed optimally in both; never-takers performed suboptimally in both.

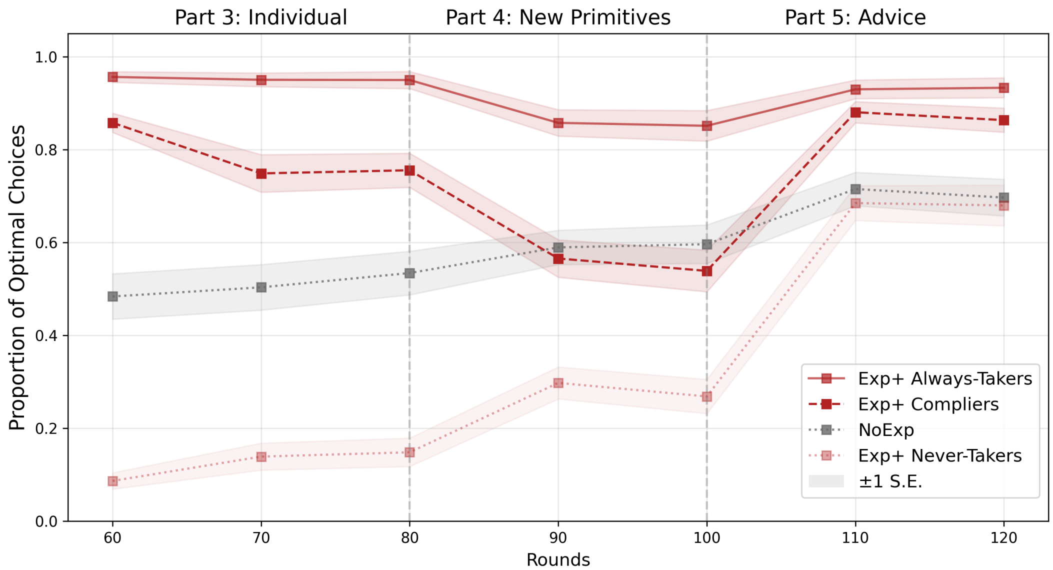
Figure 24: Exposure Complier Analysis under Threshold 0.6



Notes. Complier dynamics using an alternative optimality threshold of 0.6 ($N = 66$, 30% of Exposure). Each bar reports the percentage of compliers who exhibit optimal behavior in subsequent parts.

all learning types under the 0.6 threshold. The intensive-margin dynamics replicate the baseline findings: always-takers maintain consistently high performance; compliers show a characteristic decline in Part 4 followed by recovery in Part 5; and never-takers start below the NoExposure control group and converge toward the complier trajectory only when advice is provided.

Figure 25: Evolution of Optimal Behavior under Threshold 0.6



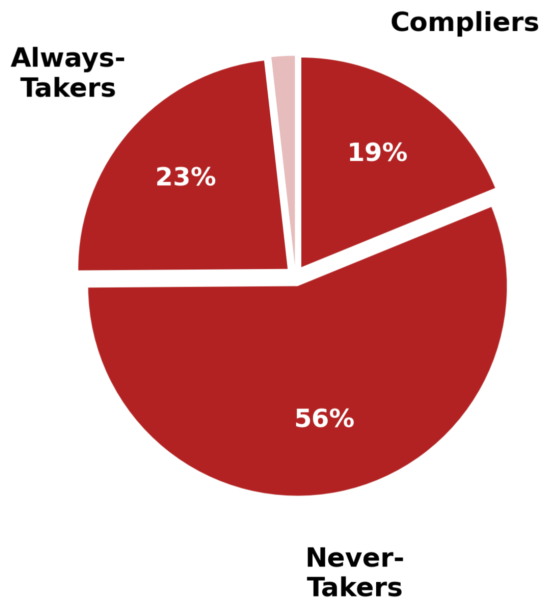
Notes. Each line tracks the average proportion of optimal choices across Parts 3–5 for a given learning type under the 0.6 threshold. Shaded regions indicate ±1 standard error bands.

D.2. Threshold 1.0

Under the 1.0 threshold, a subject is classified as exhibiting optimal behavior only if all ten choices in the final ten key rounds are optimal. This is the most stringent possible criterion and admits only subjects who achieved perfect accuracy. Figure 26 shows that the resulting distribution shifts dramatically toward never-takers: 19% compliers ($N = 42$), 56% never-takers ($N = 125$), 23% always-takers ($N = 52$), and 2% defiers ($N = 4$). The complier share falls from 25% to 19%, reflecting the fact that some subjects who made 8–9 optimal choices in Part 2 but not a perfect 10 are reclassified as never-takers. The always-taker share declines from 30% to 23% for the same reason: subjects who performed well but not perfectly in Part 1 are no longer counted as having pre-treatment mastery.

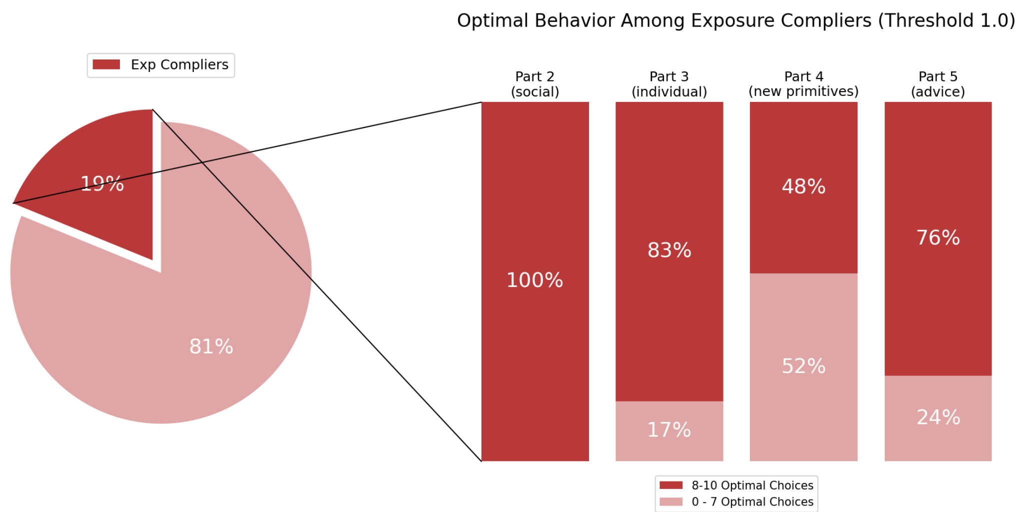
Despite the substantially smaller and more select complier group, the qualitative dynamics persist. Figure 27 shows that compliers under the 1.0 threshold exhibit an even steeper decline than under the baseline: 83% optimal in Part 3, dropping to 48% in Part 4, and recovering to 76% in Part 5. The sharper Part 4 decline (a drop of 35 percentage points from Part 3, versus 29 at the baseline threshold) is consistent with the interpretation that perfect accuracy in Part 2 does not guarantee perfect transfer to the new primitives — even subjects who achieved 100% optimal choices under the original environment face genuine difficulty when the task parameters change. The somewhat lower Part 5 recovery (76% vs. 84%) may reflect the stringency of the threshold:

Figure 26: EXPOSURE Learning Type Classification under Threshold 1.0



Notes. Classification of EXPOSURE subjects ($N = 223$) into learning types using an alternative optimality threshold of 1.0. A subject exhibits optimal behavior only if all 10 choices in the final ten key rounds are optimal.

Figure 27: EXPOSURE Complier Analysis under Threshold 1.0

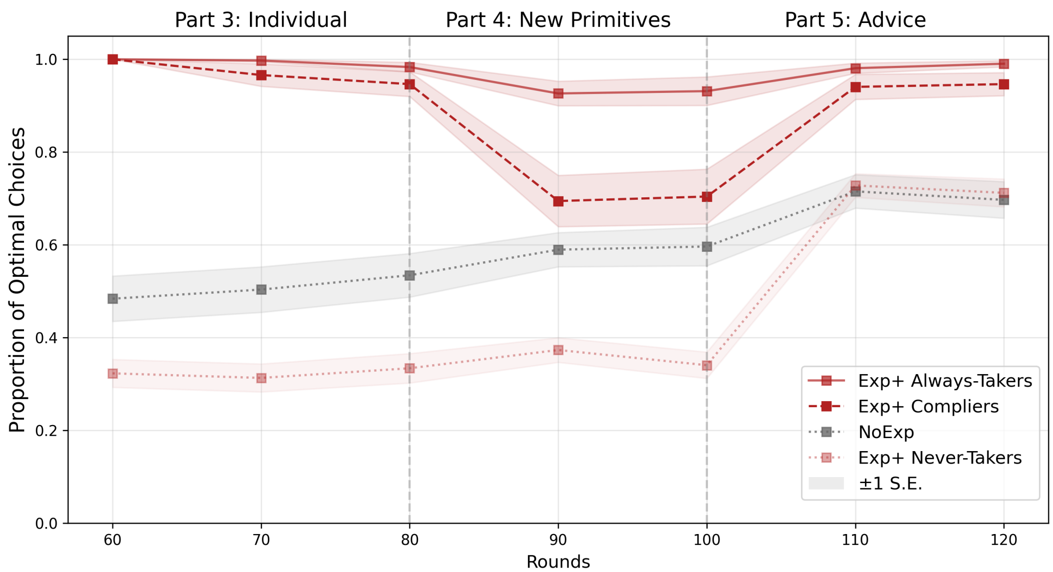


Notes. Compliers under threshold 1.0 ($N = 42$, 19% of EXPOSURE) achieved 100% optimal choices in Part 2 but not in Part 1. Each bar reports the percentage who exhibit optimal behavior in subsequent parts.

achieving perfect accuracy under advice is harder than achieving 80% accuracy.

Figure 28 confirms that the intensive-margin dynamics under the 1.0 threshold replicate the baseline findings across all learning types. Always-takers remain near the ceiling throughout. The complier trajectory shows the same characteristic pattern of decline and recovery; because the figure

Figure 28: Evolution of Optimal Behavior under Threshold 1.0



Notes. Each line tracks the average proportion of optimal choices across Parts 3–5 for a given learning type under the 1.0 threshold. Shaded regions indicate ±1 standard error bands.

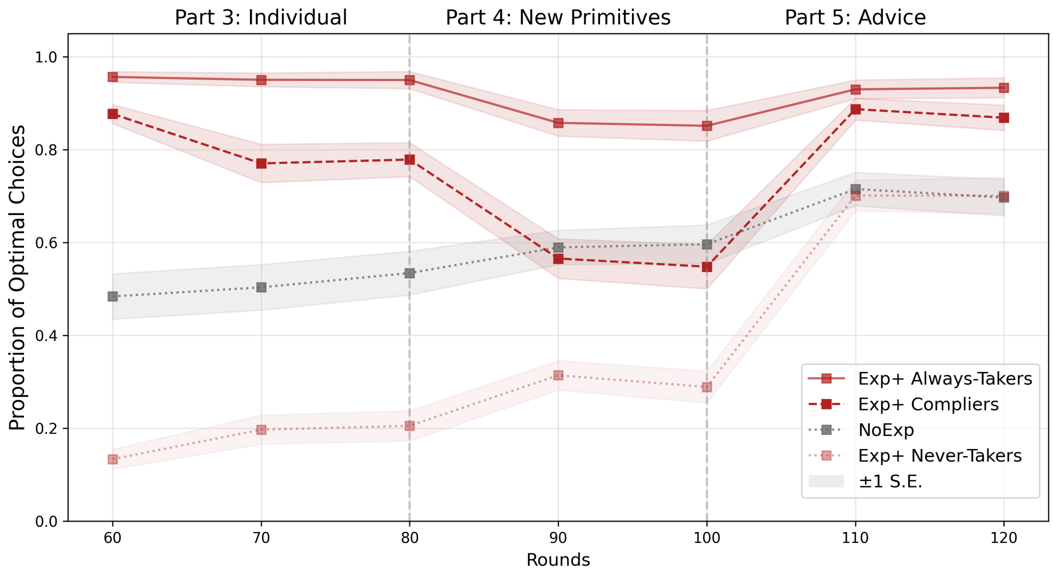
plots average optimal-choice shares, the stricter selection if anything shifts the trajectory upward relative to the baseline compliers. Never-takers under the 1.0 threshold (which now includes some subjects who would have been classified as compliers or always-takers under the 0.8 threshold) show a somewhat higher average performance than under the baseline, reflecting the reclassification of partially-optimal subjects into this group.

D.3. Alternative Classification Criteria

The threshold-based classification counts any subject who crosses the optimality threshold in Part 2 but not Part 1 as a complier. A potential concern is that some subjects may cross the threshold by chance rather than in response to treatment, especially near the boundary. To address this, I consider two alternative classifications that impose a minimum improvement (Δ) requirement: a subject is classified as a complier only if she both exceeds a level requirement in Part 2 and improves by at least Δ choices relative to Part 1.

Figure 29 uses a classification requiring at least 6 optimal choices in the final ten key rounds of Part 2 with an improvement of at least 4 choices relative to Part 1. Under this criterion, the complier group comprises 61 subjects, with 75 always-takers and 87 never-takers. The improvement requirement ensures that every complier exhibited a substantial behavioral shift between the baseline and exposure phases, ruling out subjects who were close to the boundary in both parts. The resulting dynamics are qualitatively indistinguishable from the baseline: compliers show declining

Figure 29: Learning Dynamics: Alternative Classification (≥ 6 in Part 2, $\Delta \geq 4$)

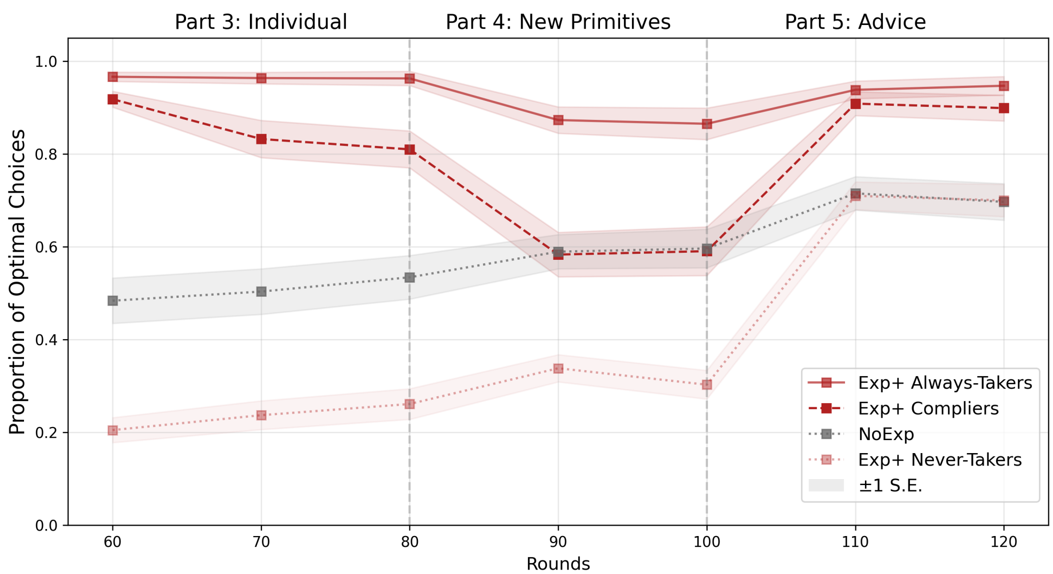


Notes. A subject in Exposure is classified as a complier ($N = 61$) if they make at least 6 optimal choices in the final ten key rounds of Part 2 and their improvement relative to Part 1 is at least 4 additional optimal choices ($\Delta \geq 4$). Always-takers ($N = 75$) meet the level requirement in both parts. Each line tracks the average proportion of optimal choices across Parts 3–5. Shaded regions indicate ± 1 standard error bands.

performance in Part 4 and recovery in Part 5, while always-takers remain stable and never-takers improve markedly only under advice.

Figure 30 uses a more stringent criterion: at least 7 optimal choices in Part 2 with an improvement of at least 5. This further restricts the complier group to 52 subjects, with 71 always-takers and 100 never-takers. Despite the more demanding definition, the learning dynamics remain qualitatively unchanged. The complier trajectory again shows the characteristic decline in Part 4 and recovery in Part 5, confirming that the pattern is not driven by marginal cases near the classification boundary. Across all four alternative specifications examined in this appendix, the core finding is robust: the decline-and-recovery trajectory of compliers — retention in the familiar environment, deterioration when primitives change, recovery under guidance — is not an artifact of any particular classification boundary. As in the main text, this trajectory characterizes behavior; on its own it does not distinguish fragile understanding from retention of a task-specific action rule.

Figure 30: Learning Dynamics: Alternative Classification (≥ 7 in Part 2, $\Delta \geq 5$)



Notes. A subject in EXPOSURE is classified as a complier ($N = 52$) if they make at least 7 optimal choices in the final ten key rounds of Part 2 and their improvement relative to Part 1 is at least 5 additional optimal choices ($\Delta \geq 5$). Always-takers ($N = 71$) meet the level requirement in both parts. Each line tracks the average proportion of optimal choices across Parts 3–5. Shaded regions indicate ± 1 standard error bands.

Table 8: Learning-Type Composition Across Classification Thresholds

Criterion	Compliers	Always-takers	Never-takers	Defiers	C1	C2	C3
$\geq$ 6 of 10	66 (30%)	75 (34%)	75 (34%)	7	32	17	17
$\geq$ 7 of 10	61 (27%)	71 (32%)	86 (39%)	5	27	17	17
$\geq$ 8 of 10 (baseline)	56 (25%)	67 (30%)	94 (42%)	6	22	17	17
$\geq$ 9 of 10	45 (20%)	59 (26%)	112 (50%)	7	21	15	9
10 of 10	42 (19%)	52 (23%)	125 (56%)	4	20	15	7

Notes: Each row classifies the 223 Exposure subjects using the stated criterion applied to the last ten key rounds of Parts 1 and 2 (y_1, y_2), and splits compliers into subtypes by applying the same criterion to Parts 3 and 4. Because y_t takes values on a ten-round grid, any threshold in $(0.7, 0.8]$ is equivalent to the baseline criterion of at least eight optimal choices.

D.4. Threshold Ladder and a Sampling-Noise Benchmark

Two further checks address the role of sampling noise in the classification. Because y_t is measured over ten key rounds, the classification threshold moves on a discrete grid: any cutoff in $(0.7, 0.8]$ implies the same criterion of at least eight optimal choices, so the relevant perturbations are one round in either direction. Table 8 reports the full ladder. The composition of the sample is stable: the complier share moves gradually from 30% at the most lenient cutoff to 19% at the strictest, and the ordering of the three subtypes is preserved throughout, with the C2 and C3 groups essentially unchanged between cutoffs of six and eight.

The second check quantifies how much of the classification could arise from chance crossings of the threshold in short windows, absent any true behavioral change. I simulate a null model in which each Exposure subject's choices are i.i.d. draws from a constant probability, set to that subject's empirical share of optimal choices over all fifteen key rounds of Part 1, and apply the classification to the simulated y_1 and y_2 windows (20,000 replications). Under this null, an average of 7.4 of the 223 subjects (3.3%) are labeled compliers, with 95% of replications producing between 3 and 13 — far below the 56 compliers observed in the data. Under this null, chance threshold-crossing accounts for about 13% of observed compliers on average, and for about 23% at the upper end of the simulation interval; these figures are benchmarks conditional on the fitted constant-propensity model rather than unconditional bounds, since serial dependence or nonstationarity in choices would alter both crossing and reversion rates. A variant that preserves within-subject serial dependence — resampling each subject's observed Part-1 key-round sequence in circular blocks of five — delivers nearly identical benchmarks (6.5 expected compliers on average, with 51% of them reverting), suggesting the constant-propensity simplification is not driving the conclusion. The same simulation also benchmarks the contribution of regression to the mean to the measured reversion: simulated null compliers "revert" in the Part 3 window 48% of the time, implying roughly 3.6 artifactual reverters — about one fifth of the 17 observed shallow imitators. Under this benchmark, the classification reflects genuine behavioral change for the large majority of compliers, though the size of the C3 group includes a modest noise component and should be interpreted accordingly.

E. Theoretical Framework

This appendix develops a conceptual framework for the central inference task in the experiment and for learning through exposure to another agent's optimal decisions. The first subsection formalizes the decision environment, including the signal structure and the Bayesian benchmark. The second subsection adapts an imitation learning model to the experimental setting. The third subsection distinguishes imitation from genuine belief updating.

E.1. Decision Setting

Consider a decision-maker who must choose between hiring a worker from group $g \in \{\text{Green}, \text{Orange}\}$. Each group consists of four workers with ability $a \in \{l, m, h\}$ drawn from group-specific distributions. Let k_a^g denote the proportion of workers with ability a in group g. Before making a hiring decision, the decision-maker observes a binary education signal for each worker, generated by the conditional probabilities $P(e \mid h) = 1$, $P(e \mid m) = p_m \in (0, 1)$, and $P(e \mid l) = 0$. This signal structure ensures that education is perfectly informative about the absence of high ability among the uneducated (since $P(e \mid l) = 0$ and $P(e \mid h) = 1$, an uneducated worker cannot be of high ability) and partially informative among the educated (since $P(e \mid m) = p_m > 0$, an educated worker may be of either high or medium ability).

Payoff structure. The decision-maker's payoff depends on the hired worker's education status and ability. In the experiment, hiring an educated worker yields \$18 if the worker is of high ability and \$13 if the worker is of medium ability; hiring an uneducated worker yields \$10 if the worker is of medium ability and \$5 if the worker is of low ability. The maximum possible payoff difference is therefore $\bar{\pi} = 18 - 5 = 13$.

Bayesian benchmark. When both workers are educated, low-ability workers are screened out ($P(e \mid l) = 0$), so the relevant comparison is the posterior probability of high ability conditional on being educated. By Bayes' rule:

$$P(h \mid e, g) = \frac{P(e \mid h) \cdot k_h^g}{P(e \mid h) \cdot k_h^g + P(e \mid m) \cdot k_m^g} = \frac{k_h^g}{k_h^g + p_m \cdot k_m^g}. \tag{1}$$

The optimal choice is to hire the worker from the group with the higher posterior $P(h \mid e, g)$, which reduces to comparing the ratio k_h^g / k_m^g across groups, since p_m is common.

Under the main primitives (Parts 1–3), the Green group has proportions $(k_h^G, k_m^G, k_l^G) = (0.25, 0.75, 0)$

and the Orange group has $(k_h^O, k_m^O, k_l^O) = (0.25, 0.50, 0.25)$, with $p_m = 0.9$. Applying equation (1):[50]

$$P(h \mid e, O) \approx 0.357 > 0.270 \approx P(h \mid e, G).$$

The posterior probability that the educated Orange worker is of high ability exceeds that of the educated Green worker, making hiring Orange optimal despite the Green group having higher average ability. The key insight is that education screening operates differently across groups: because the Orange group has a higher share of low-ability workers who are screened out, the surviving educated pool is of higher quality. A decision-maker who neglects the education signal and relies instead on unconditional group averages will suboptimally prefer the Green worker. The full derivation is in Section 2.

E.2. Learning Through Imitation

Setup. Consider an agent i who faces the binary choice problem over T rounds. In each round t, the agent selects an action $a_i^t \in \{G, O\}$ and receives a payoff π_i^t. Between rounds, agent i observes the action a_j^t chosen by another agent j who faces the same environment and uses a learning rule to determine her action in round $t+1$.

A learning rule is *improving* if, for any payoff distributions with $\mu_O > \mu_G$, the expected share of agents choosing the optimal action O weakly increases. Among all improving rules, the learning rule that maximizes the rate of convergence toward the optimal action is one in which the probability of switching to the observed action is proportional to the payoff advantage.[51] Under such a rule,

[50] For the main primitives: $P(h \mid e, G) = \frac{0.25}{0.25+0.9\times0.75} = \frac{0.25}{0.925} \approx 0.270$ and $P(h \mid e, O) = \frac{0.25}{0.25+0.9\times0.50} = \frac{0.25}{0.70} \approx 0.357$. For the new primitives (Parts 4–5), where $(k_h^G, k_m^G, k_l^G) = (0.25, 0.25, 0.50)$, $(k_h^O, k_m^O, k_l^O) = (0, 0.50, 0.50)$, and $p_m = 0.1$, the relevant case is both workers being uneducated. Then, since a medium-ability worker is uneducated with probability $1 - p_m = 0.9$, $P(m \mid \neg e, G) = \frac{(1-p_m)\cdot k_m^G}{(1-p_m)\cdot k_m^G + k_l^G} = \frac{0.9\times0.25}{0.9\times0.25+0.50} \approx 0.31$ and $P(m \mid \neg e, O) = \frac{0.9\times0.50}{0.9\times0.50+0.50} \approx 0.47$. In both cases, the Orange worker has a higher posterior probability of being the better type, making hiring Orange optimal.

[51] This is the Proportional Imitation Rule (PIR) introduced by Schlag (1998): $\Pr(a_i^{t+1} = a_j^t \mid a_j^t \neq a_i^t) = \max\{\pi_j^t - \pi_i^t, 0\}/\bar{\pi}$, where $\bar{\pi}$ is the maximum payoff difference. The PIR is optimal in a general class of environments where the agent does not know the underlying payoff distributions. In our experimental setting, the payoff structure is simpler: when both workers are educated, payoffs take one of two values ($18 for high ability, $13 for medium ability), and the payoff difference between a high-ability and a medium-ability hire is fixed at $5. The PIR reduces to a rule in which switching occurs with a fixed probability whenever the observed action yields a higher realized payoff. Since subjects observe participant X's action but not X's payoff, the rule cannot operate on round-level payoff comparisons; an implementable adaptation must substitute experience-based action values accumulated in Part 1, when virtually all subjects tried both actions repeatedly. This adaptation is a distinct learning rule rather than the literal PIR, and it does not by itself guarantee that subjects rank the observed action correctly; the PIR serves here as a theoretical benchmark for payoff-proportional imitation, not as a description of the implemented information structure.

the expected change in the share p^t choosing the optimal action follows:

$$\mathbb{E}[p^{t+1} - p^t] = \frac{\Delta\mu}{\bar{\pi}} \cdot p^t(1 - p^t), \tag{2}$$

where $\Delta\mu = \mu_O - \mu_G > 0$ is the expected payoff advantage and $\bar{\pi}$ is the maximum payoff difference. This is the discrete-time analogue of the replicator dynamic from evolutionary game theory. The term $p^t(1 - p^t)$ reflects the fact that switching can only occur when the two agents choose different actions.

Adaptation to the experimental setting. In the experiment, agent j is participant X, who always chooses optimally ($a_j^t = O$ for all t). This simplifies the dynamics: agent i can only switch *toward* the optimal action, never away from it. The per-round switching probability for a suboptimal agent is:

$$\Pr(a_i^{t+1} = O \mid a_i^t = G) \;\equiv\; \lambda > 0, \tag{3}$$

and the probability that agent i remains suboptimal after T rounds of exposure decays geometrically:

$$q^T = q^0 \cdot (1 - \lambda)^T. \tag{4}$$

With $T = 40$ rounds in Part 2, even a modest per-round switching probability generates substantial cumulative learning. Because switching is probabilistic rather than deterministic, the model implies that subjects do not switch simultaneously; instead, the cumulative share of switchers rises gradually, approaching its limit at a geometrically decaying rate, consistent with the gradual divergence between Exposure and NoExposure documented in Section 3.2.

This framework also clarifies why exposure to suboptimal choices has no effect. Under the learning rule, an agent never switches to an action that yielded a weakly lower payoff. When agent j always chooses suboptimally (the ExposureNeg treatment), a subject currently choosing Green observes $a_j^t = G = a_i^t$ and no switching occurs; a subject currently choosing Orange observes $a_j^t = G \neq a_i^t$ but since Orange yields a higher expected payoff, the switching probability toward Green is zero. The asymmetric response to optimal versus suboptimal exposure documented in Section 3.3 is therefore an implication of the model under the adaptation above. Similarly, because switching operates only when $a_j^t \neq a_i^t$, subjects who already choose optimally in Part 1 (always-takers) are unaffected by the treatment, while the largest impact falls on subjects who start suboptimally and are exposed to a superior alternative (Section 3.4).

The role of own-payoff feedback. A key feature of the experiment is that subjects observe only participant X's action, not participant X's payoff. However, subjects receive feedback on their own payoff after each round. This allows a modified learning rule in which the agent compares her realized payoff π_i^t with an inferred counterfactual payoff $\hat{\pi}_j^t$ for the action she did not take:

$$\Pr(a_i^{t+1} = O \mid a_i^t = G) = f\big(\pi_i^t,\ \hat{\pi}_j^t\big), \tag{5}$$

where f is increasing in $\hat{\pi}_j^t - \pi_i^t$. This implies that switching should be more likely in rounds where the subject's own payoff is low (e.g., hiring a medium-ability Green worker at \$13 rather than a high-ability one at \$18), a pattern consistent with the payoff-driven selective adoption documented in Section 3.3.

E.3. Imitation vs. Learning

The framework above describes *action imitation*: agents copy the observed action without necessarily understanding why it is optimal. Under pure imitation, a subject who switches to hiring the Orange worker in Part 2 would revert to suboptimal behavior once the environment changes (Part 4) or exposure is removed (Part 3), because the copied action may no longer be correct or the copying stimulus is absent.

A richer model allows for *belief updating*: through repeated exposure to optimal actions and the associated payoff feedback, agents may come to understand the underlying principle that education signals carry useful information about worker quality. Formally, let $\beta_i^t \in [0, 1]$ denote agent i's belief that the orange action is optimal. Under belief updating, each round of exposure shifts β_i^t upward:

$$\beta_i^{t+1} = \beta_i^t + \alpha \cdot \big(\mathbf{1}\{\text{evidence favoring } O \text{ in round } t\} - \beta_i^t\big), \tag{6}$$

where $\alpha \in (0, 1)$ is a learning rate. Under this specification, the probability of choosing optimally depends on the accumulated belief rather than on the presence of the external signal, so behavior persists even when exposure is removed.

Parts 3–5 of the experiment are designed to distinguish these channels. For most subjects, the evidence favors the action-based end of the spectrum: only about 40% of compliers continue to behave optimally under the new environment primitives in Part 4 (Section 4.2), and the group-level optimal choice rate becomes statistically indistinguishable from the control group's. For this minority of persisting subjects, behavior is consistent with belief updating — they have internalized the principle that the education signal carries useful information and apply it in a novel setting where the observed choice stream is no longer displayed and cannot simply be copied. For the majority, however, gains acquired under exposure do not survive the change in primitives, consistent with adoption of the observed action rather than revision of the underlying model.

F. Q-Learning Simulation

To illustrate how the observed dynamics can arise from a simple learning process, I simulate a Q-learning model that mirrors the main components of the NoExposure and Exposure treatments. In the simulation, agents update the value of actions based on realized payoffs and select actions probabilistically as a function of their current value estimates. The state space differs across treatments: in NoExposure, the state depends only on the agent's previous choice, while in Exposure, it additionally incorporates whether the agent's own realized payoff in the previous round was high or low. Both groups observe their own payoffs, so the difference across treatments is one of representation rather than information: the constant display of participant X's contrasting choice supplies a concrete alternative against which the agent's own payoff can be evaluated, which makes the payoff realization a behaviorally relevant component of the state. This modeling choice mirrors the selective adoption evidence in Section 3.3, where the observed choice identifies *what* to try while the agent's own payoff experience determines *when* it is tried and whether it is kept. In NoExposure, with no alternative on display, behavior conditions on the previous action alone. The simulation uses a learning rate of $\alpha = 0.05$, a discount factor of $\delta = 0.9$, and a softmax temperature of $\lambda = 0.1$; these values are chosen for illustration rather than estimated from the data, and the exercise is accordingly a demonstration that the qualitative patterns are reproducible by a minimal payoff-gated learning process, not a fitted model. Estimating the parameters from round-level choice sequences and testing the specification against parsimonious alternatives, such as a win-stay/lose-shift rule, is a natural next step.

Formally, each agent maintains a Q-value $Q(a, s)$ for each action $a \in \{\text{Green}, \text{Orange}\}$ and state s, representing the expected value of taking action a in state s. After each round, Q-values are updated according to the standard reinforcement learning rule

$$Q(a, s) \leftarrow (1-\alpha)Q(a, s) + \alpha \left[r + \delta \max_{a' \in \mathcal{A}} Q(a', s') \right],$$

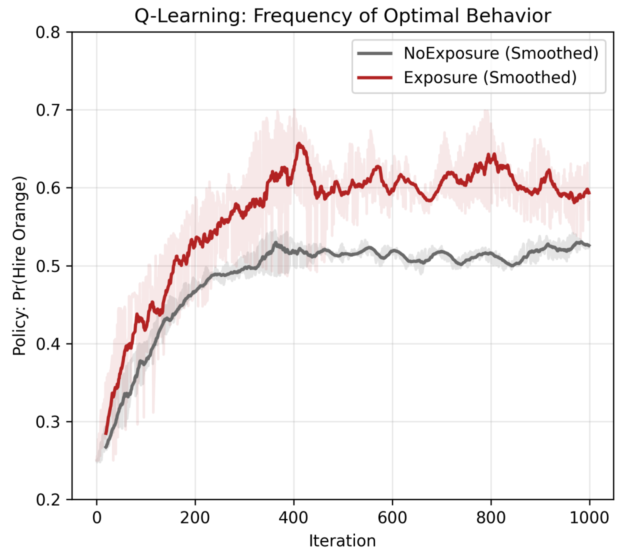
where r denotes the payoff from choosing action a in state s, s' is the subsequent state, $\alpha \in (0, 1)$ is the learning rate, and $\delta \in (0, 1)$ is the discount factor. Given these value estimates, actions are selected using a softmax rule,

$$\mathbb{P}(a \mid s) = \frac{e^{\lambda Q(a,s)}}{\sum_{a' \in \mathcal{A}} e^{\lambda Q(a',s)}},$$

where $\lambda > 0$ governs the sensitivity of choice probabilities to differences in Q-values.

The simulated paths in Figure 31 plot the evolution of the probability of choosing the optimal action over time and closely mirror the empirical patterns observed in the experiment. In particular,

Figure 31: Q-Learning Simulation for NoExposure and Exposure



Notes: This figure plots simulated behavior from a Q-learning model. In the NoExposure condition, the state depends only on the agent's previous choice. In the Exposure condition, the state additionally incorporates whether the agent's own realized payoff in the previous round was high or low. The simulated paths show the evolution of the probability of choosing the optimal action over time.

exposure to optimal behavior leads to faster increases in optimal choices early on, while agents learning only from their own feedback improve more gradually. The Exposure path reaches high optimality rates earlier because the richer state representation allows agents to learn faster from payoff-relevant signals; the simulation holds each treatment's representation fixed throughout, so it speaks to the speed of learning under exposure rather than to behavior after exposure is withdrawn.

The Q-learning framework also suggests a conjecture — not simulated here — about the fragility of learning gains documented in Section 4. Under reinforcement learning, action values are updated incrementally based on recent experience, and the weight placed on any given piece of information decays as new evidence accumulates. If exposure accelerates learning primarily by providing additional states that facilitate faster discrimination between actions, removing exposure collapses the state space back to its baseline size. Agents who have not yet accumulated sufficient experience to sustain their new behavior through the reduced state representation may gradually drift back toward suboptimal choices. This mechanism is consistent with the empirical observation that a subset of compliers revert once exposure is removed, while those who have had more rounds of reinforcement tend to retain their gains.

Overall, the transition evidence and the Q-learning simulation point to the same mechanism: learning from exposure operates through short-run, payoff-driven selective adoption, in which disappointment with one's own choice triggers experimentation with the observed alternative and favorable payoff realizations sustain the new behavior once adopted. This helps explain why exposure to optimal behavior can generate rapid improvements in performance while also producing learning gains that are sensitive to recent outcomes and may fail to persist or transfer once the environment changes.

Additional Result (Q-Learning): *Simulating a learning model in which observing another subject's choices expands the learner's state representation generates dynamic patterns consistent with the empirical ones documented in the experiment.*

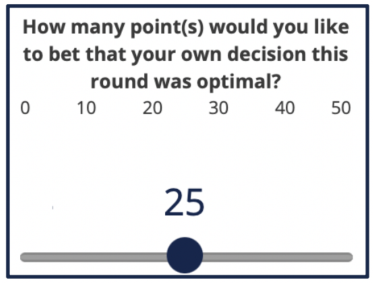
G. The Role of Confidence

Section 3.3 established that exposure to suboptimal choices has no detectable average effect on behavior. A natural question is whether this null result masks heterogeneity across subjects with different levels of confidence in their own strategy. Intuitively, subjects who are less certain about the optimality of their approach may be more susceptible to the influence of observed choices, regardless of their quality, whereas more confident subjects may be better able to evaluate and filter the information they receive. This appendix examines how baseline confidence moderates the treatment effects documented above.

Confidence elicitation. Confidence is elicited at six points during the experiment (rounds 10, 30, 50, 70, 90, and 110). In each elicitation round, after making a hiring choice but before observing the payoff, subjects are asked how many points out of 50 they wish to bet on the event that their current choice is optimal.[52] This betting task is incentive-compatible: if the subject's choice turns out not to be optimal, all bet points are lost and the final number of points equals the points not bet. If the choice is optimal, each bet point is multiplied by 1.5, and the final number of points equals the points not bet plus 1.5 times the bet points. The final point total is then converted into a probability of winning a \$2 bonus at a rate of one point per percentage point. Under risk neutrality, this rule makes the optimal bet a threshold report — zero points when the believed probability of an optimal choice is below 2/3 and all fifty points when it is above — so interior bets are not cardinally interpretable as beliefs; for risk-averse subjects the optimal bet instead rises smoothly with confidence. Accordingly, the analysis treats the 0–50 wager as an ordinal behavioral proxy for confidence rather than as an exact belief measure, which is how it enters the specifications below (as a standardized moderator).

[52] Subjects are informed that *optimal* means that the worker they hired would generate a higher payoff than the worker they did not hire. The bet is resolved using the realized abilities of both drawn workers: the choice counts as optimal if the hired worker's payoff exceeds the payoff the other worker would have delivered. Because the benchmark action maximizes the probability of this event conditional on the observed signals, the outcome-based criterion used to resolve the bets and the ex-ante benchmark rule used throughout the paper coincide in expectation.

Figure 32: Confidence Elicitation Interface



Notes: Screenshot of the confidence elicitation interface. Subjects use a slider to indicate how many points out of 50 they wish to bet on the event that their current hiring choice is optimal. The bet is incentive-compatible as described in the text.

The baseline confidence measure, elicited in round 10 of Part 1 before any treatment assignment, has a mean of 25.1 and a standard deviation of 15.7 across the 356 subjects in the main sample. Regressing baseline confidence on treatment dummies confirms that randomization successfully balanced this measure across groups ($p = 0.622$ for Exposure, $p = 0.214$ for ExposureNeg). Throughout this appendix, confidence is standardized to have mean zero and unit variance in the estimation sample.

Confidence interactions. To study whether confidence moderates the learning effects from exposure, I estimate specifications that add the standardized confidence measure and its interactions with the treatment dummies. Table 9 reports the results across three outcome measures: the share of optimal choices in the last ten rounds (columns 1–2), the optimal choice indicator in the last round (columns 3–4), and a binary indicator for whether a subject achieves at least 8 optimal choices out of 10 in the last ten key rounds (columns 5–6). Odd-numbered columns include confidence as a control; even-numbered columns add the interaction terms. Confidence is a small and insignificant predictor of optimal choices on its own (-0.007, $p = 0.782$ in column 1; -0.022, $p = 0.392$ in column 3; 0.005, $p = 0.853$ in column 5), and the treatment effect estimates are virtually unaffected by its inclusion.

The key result emerges in the even-numbered columns, which add the interaction terms. The interaction between Exposure and confidence is essentially zero across all three outcome measures (-0.002, $p = 0.970$ in column 2; 0.000, $p = 0.997$ in column 4; 0.024, $p = 0.720$ in column 6), indicating that the beneficial effect of observing optimal choices operates with similar strength regardless of the subject's baseline confidence level. In contrast, the interaction between ExposureNeg and confidence is large and statistically significant in every specification (0.172, SE = 0.076, $p = 0.024$ in column 2; 0.170, SE = 0.083, $p = 0.040$ in column 4; 0.188, SE = 0.082, $p = 0.023$ in column 6).

Table 9: The Role of Confidence in Part 2

	Optimal Choices in Last 10 Rounds		Optimal Choice in Last Round		Binary Optimality (≥ 8 of 10)	
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure	0.129** (0.056)	0.131** (0.056)	0.212*** (0.064)	0.213*** (0.064)	0.126* (0.065)	0.129** (0.065)
ExposureNeg	−0.005 (0.079)	−0.025 (0.075)	0.014 (0.090)	−0.005 (0.087)	−0.011 (0.088)	−0.029 (0.085)
Confidence (std)	−0.007 (0.024)	−0.033 (0.052)	−0.022 (0.026)	−0.050 (0.058)	0.005 (0.026)	−0.041 (0.057)
Exp × Confidence		−0.002 (0.060)		0.000 (0.066)		0.024 (0.066)
ExpNeg × Confidence		0.172** (0.076)		0.170** (0.083)		0.188** (0.082)
NoExposure Mean	0.491	0.491	0.463	0.463	0.425	0.425
Observations	3,560	3,560	356	356	356	356
R^2	0.017	0.033	0.043	0.060	0.016	0.032

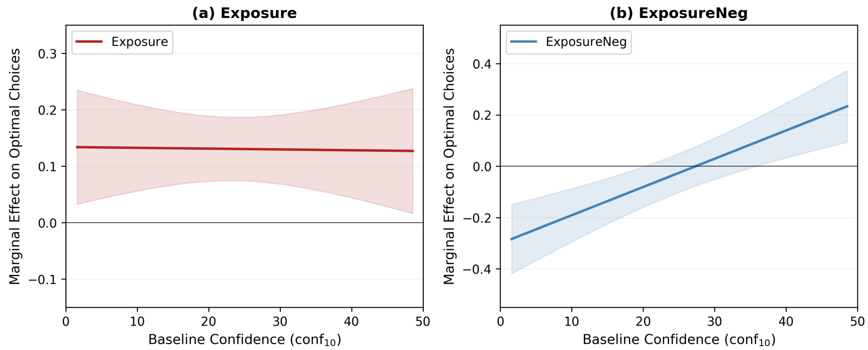
Notes: Each column reports a separate OLS regression of an optimality measure on treatment dummies, standardized baseline confidence, and their interactions. The sample includes NoExposure ($N = 80$), Exposure ($N = 223$), and ExposureNeg ($N = 53$). Columns (1)–(2) use a round-level optimal choice indicator in the last ten rounds of Part 2 with standard errors clustered at the subject level. Columns (3)–(4) use the optimal choice indicator in the last round of Part 2 with HC1 robust standard errors. Columns (5)–(6) use a binary indicator equal to one if a subject makes at least 8 optimal choices in the last 10 key rounds of Part 2, with HC1 robust standard errors. Odd-numbered columns control for confidence; even-numbered columns add the interaction terms. Confidence is the standardized baseline elicitation from round 10 of Part 1 (raw mean = 25.1, SD = 15.7, range 0–50). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

This implies that the effect of exposure to suboptimal choices varies substantially with baseline confidence: point estimates indicate harm concentrated among less confident subjects, while more confident subjects appear insulated. The consistency of this pattern across the intensive margin (columns 1–2), the last-round indicator (columns 3–4), and the extensive margin of the 80% optimality threshold (columns 5–6) confirms that the heterogeneity is not an artifact of any particular outcome measure.

Marginal treatment effects. To visualize how the treatment effects vary across the confidence distribution, Figure 33 plots the marginal effects of Exposure and ExposureNeg from column (2) of Table 9 as a function of baseline confidence. Because the interaction model is linear, each marginal effect is a linear function of confidence: for Exposure, the marginal effect at confidence level z equals $\hat{\beta}_{\text{Exp}} + \hat{\beta}_{\text{Exp}\times\text{Conf}} \cdot z$, with an analogous expression for ExposureNeg.

Panel (a) confirms that the Exposure effect is remarkably stable: the marginal effect hovers

Figure 33: Marginal Treatment Effects by Baseline Confidence



Notes: Each panel plots the estimated marginal treatment effect (relative to NoExposure) on the share of optimal choices in the last ten rounds of Part 2 as a function of baseline confidence. Panel (a) shows the Exposure effect; panel (b) shows the ExposureNeg effect. The marginal effects are computed from the interaction model in column (2) of Table 9: for each treatment T, the marginal effect at standardized confidence z equals $\hat{\beta}_T + \hat{\beta}_{T\times\text{Conf}} \cdot z$. The horizontal axis is expressed in raw confidence units (0–50). Shaded regions represent ± 1 standard error bands, computed from the variance-covariance matrix of the estimated coefficients.

around 0.13 throughout the confidence distribution, and the standard error band excludes zero across a wide range. Exposure to optimal choices benefits subjects of all confidence levels. Panel (b) reveals the stark heterogeneity in the ExposureNeg effect. At the bottom quartile of confidence (corresponding to a raw score of roughly 8 out of 50), the marginal effect of ExposureNeg is -0.214 ($p = 0.053$), a marginally significant estimate suggesting that low-confidence subjects exposed to suboptimal choices perform worse than those who learn independently. At the sample mean, the effect is indistinguishable from zero (-0.025, $p = 0.742$). At the top quartile of confidence (a raw score of roughly 40), the point estimate reverses sign to $+0.147$ ($p = 0.177$), suggesting that high-confidence subjects may even benefit from observing suboptimal behavior, though this positive estimate is imprecisely measured.

Splitting the sample into confidence quartiles and estimating separate regressions within each group provides a complementary, nonparametric view of this heterogeneity. In the bottom quartile ($N = 95$, confidence scores 0–10), the ExposureNeg coefficient is -0.146 ($p = 0.322$), while in the top quartile ($N = 70$, confidence scores 41–50), it is $+0.216$ ($p = 0.244$). Although neither estimate is individually significant given the reduced sample sizes, the sign reversal from the bottom to the top of the confidence distribution is consistent with the pattern identified by the interaction specification, though it should be read as descriptive corroboration rather than independent confirmation.[53] This pattern is consistent with a model in which less confident subjects are more

[53] Baseline confidence does not predict whether a subject will respond to exposure in general: within the Exposure group, mean confidence is virtually identical across compliers (24.95), always-takers (24.00), and never-takers (25.14), and a joint test of equality fails to reject the null ($p = 0.911$). Thus, the heterogeneity by confidence documented here, which concerns susceptibility to the *quality* of observed choices, operates through a different channel than the

willing to adopt observed strategies regardless of their quality, making them vulnerable when the observed choices are suboptimal, while more confident subjects evaluate the quality of what they observe and potentially recognize the inferiority of the suboptimal strategy, reinforcing their own approach.

Finally, while treatment assignment substantially affects behavior, it leaves no detectable imprint on subjects' stated confidence in their own strategy. Regressing later confidence elicitations (in rounds 50, 70, 90, and 110) on treatment dummies, controlling for baseline confidence, yields no significant treatment effects in any part of the experiment (all $p > 0.12$). These null results do not, absent formal equivalence bounds, rule out moderate effects; what can be said is that subjects who shift to optimal behavior under exposure do not report detectably greater certainty that their choices are correct, and subjects exposed to suboptimal behavior report no detectably lower confidence.

Additional Result (Confidence): *The beneficial effect of exposure to optimal choices does not vary with baseline confidence, whereas the response to suboptimal choices declines sharply with confidence, with point estimates indicating harm concentrated among the least confident subjects; treatment assignment leaves no detectable imprint on subsequent confidence elicitations.*

learning type classification in Section 3.4, which concerns whether subjects respond to exposure at all.

H. Experimental Instructions and Procedures

This section reproduces the complete experimental instructions and procedures as presented to participants in the Exposure treatment. It covers the overview of the task, the hiring decision interface, the payoff structure, the confidence elicitation mechanism, and the detailed instructions for each of the five parts of the experiment. Participants in the NoExposure treatment faced the exact same interface and instructions, except they did not observe the hiring choices of another participant in Part 2 and were not provided with that participant's written advice in Part 5.

Experiment Instructions

Welcome to the study!

This is an experiment in decision-making. Your final payment will be based on your performance and there will be an element of random chance. This will be paid to you privately at the end of the experiment. From now on, we ask that you do not communicate with each other. If you have a question, please raise your hand.

The experiment will consist of short, independent rounds. In each round, you will be asked to make one choice.

Next, we will review the main task description, followed by a series of comprehension questions. You need to answer these questions correctly to continue with the experiment. After that, you will proceed to the main part of the study, which will determine your final payoff.

On each page, please read all the information carefully, as you will **NOT** be able to return to any of the previous pages. During the task description part, please do not move to the next page until we have reviewed all the instructions on a given page and you are instructed to proceed. Do not exit the browser window at any point.

Before we proceed, please write down your desk number below:

Description

Overview

This study will consist of **120 independent short rounds**.

In each round, the interface will display two groups of workers: one with 4 Green workers and one with 4 Orange workers.

Each worker can have a **HIGH, MEDIUM or LOW ABILITY** level

- You will be able to see the number of workers at each ability level in both groups
- Rounds 1–80 will each randomly use one of the <u>same two distributions of abilities</u>
- Rounds 81–120 will use a <u>different distribution of abilities</u>

In each round, one worker from each group will be **RANDOMLY** selected — we will call these two workers **SELECTED WORKERS**

- You will NOT be able to see the ability level of these selected workers. You will instead see the **EDUCATION** status (Educated or Not educated) of the Green selected worker and the Orange selected worker

Regardless of the group that the worker belongs to, the following information will always be true:

- A high ability worker is educated with 100% chance (for sure)
- A medium ability worker is educated with 90% chance for rounds 1–80 and 10% chance for rounds 81–120
- A low ability worker is educated with 0% chance (for sure not)

After each round, the selected workers will be replaced back into their respective pools, so that any individual worker can be selected more than once across multiple rounds.

In rounds 21–60, you will also be provided with the **HIRING CHOICE OF ANOTHER STUDY PARTICIPANT** who took part in the exact same study in one of the past sessions, we will call this participant **PARTICIPANT X**

- In each round, participant X **saw the exact same information** as you will (distribution of abilities and education statuses of the selected workers)
- For example, if in round 30 you see that the abilities draw is Draw #1, the selected Green worker is educated, and the selected Orange worker is educated, it means that participant X in round 30 saw that the abilities draw is Draw #1, the selected Green worker is educated, and the selected Orange worker is educated
- All of the hiring choices you will see were made by **the same participant**

• In rounds 21–60, you will be provided with this hiring choice and the interface will, for instance, say "**Participant X chose to hire the Green worker**" or "**Participant X chose to hire the Orange worker**"

Below is an example of the interface that you will see in each round:

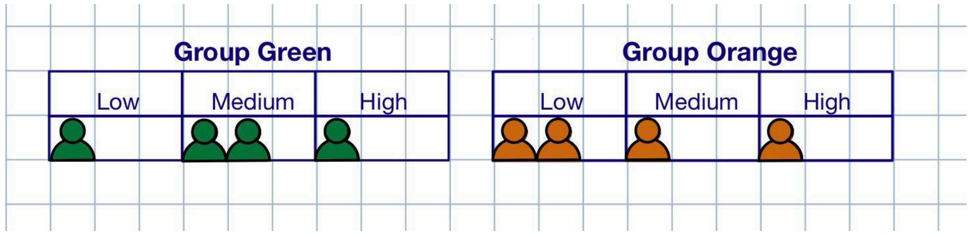


Figure: Example interface showing ability distributions for the two worker groups

In this example, among the workers in Green group:

• 1 has low ability, 2 have medium ability and 1 has high ability

In this example, among the workers in Orange group:

• 2 have low ability, 1 has medium ability and 1 has high ability

The interface will then **randomly** select one worker from the Green group and one worker from the Orange group — you will not see the ability levels of these selected workers

• You will instead see their education statuses

• For example, it could be that the Green worker is **Educated** and the Orange worker is **Not educated**

Your Task

In each round, your task is to hire one of the two selected workers. You will choose **one** of the following:

• Hire **GREEN** worker

• Hire **ORANGE** worker

After each round, you will learn the **ability of the worker that you hired.** You will then move on to the next round.

Let us know if you have any questions at this moment.

Your Payoff

Your final payoff will be determined based on your payoff from a **randomly selected round.**

In other words, after you complete all 120 rounds, the computer will randomly select one round and your final earnings in this experiment will be equal to your earnings in that randomly selected round.

Your earnings will be equal to:

Worker Type	Educated	Not Educated
High ability	$18	$11
Medium ability	$13	$10
Low ability	$12	$5

Therefore, you will get a higher payoff if you hire an Educated worker compared to a Not educated worker.

Also, within the same education status you will get a higher payoff if you hire a worker of Higher ability when compared to a worker of Lower ability.

You will be reminded of the payoff information on each decision page.

In the previous example, suppose that a randomly chosen worker from the Green group is Educated and of Medium ability, while a randomly chosen worker from the Orange group is Not educated and of Medium ability.

Then, if you hire a Green worker you will get $13 and if you hire an Orange worker you will get $10.

Bonus

Six times during the experiment, in rounds 10, 30, 50, 70, 90 and 110, you will make choices which can lead to a $2 bonus in addition to the payoff described earlier.

To determine whether you will win the bonus or not, we will randomly choose one out of your six bonus choices and the outcome of only that one choice will count.

Here is the description of the additional task you will face:

- You will be given a budget of 50 points
- You will use the slider to decide **how many points out of 50 to bet that your decision in a given round is optimal**
- "Optimal" here means that the worker you hired will bring you a higher or the same payoff as the worker you did not hire

- Every point you don't bet you get to keep.

Based on whether your decision in that round was actually optimal or not, we will determine your bonus as follows:

- If your decision was not optimal, i.e. if the worker you **did not hire** is associated with a higher payoff, every point you bet will be lost and the final amount of points will be:

Final amount of points = Points you did not bet

- If your decision was optimal, i.e. if the worker you **hired** is associated with a higher payoff, every point you bet will be multiplied by a factor of 1.5 and the final amount of points will be:

Final amount of points = Points you did not bet + 1.5 × Points you bet

All points will then be transformed to a percent chance to win a $2 bonus at the rate 1 point = 1 percent.

Example 1:

Suppose you chose to bet all 50 points. Then:

- If your decision this round was not optimal, you will have 0 points as a final amount, and 0% chance of winning a bonus
- If your decision this round was optimal, you will have 0 + 1.5×50 = 75 points as a final amount, and 75% chance of winning a bonus

Example 2:

Suppose you chose to bet 20 points. Then:

- If your decision this round was not optimal, you will have 30 points as a final amount, and 30% chance of winning a bonus
- If your decision this round was optimal, you will have 30 + 1.5×20 = 60 points as a final amount, and 60% chance of winning a bonus

Example 3:

Suppose you chose to bet 0 points. Then:

- Regardless of whether your decision this round was optimal or not, you will have 50 points as a final amount, and 50% chance of winning a bonus

Therefore, if you are certain that your decision in a given round is optimal, then betting more points is associated with a higher chance of winning a bonus.

And if you are not certain that your decision in a given round is optimal, then betting less points is associated with a higher chance of winning a bonus.

Summary

The study consists of 120 rounds.

In each round, the interface will show one group of 4 Green workers and one group of 4 Orange workers.

- Each of the 8 workers can be of **high, medium, or low ability**, and can be **educated or Not educated**
- Rounds 1–80 will each randomly use one of the <u>same two distributions of abilities</u> • Rounds 81–120 will use a <u>different distribution of abilities</u>
- In rounds 21–60, you will also be provided with the **hiring choice of another study participant** (participant X) from one of the past study sessions

One worker from each group will be **randomly** selected, and their education status will be shown:

- A high ability worker has a 100% probability of being educated
- A medium ability worker has a 90% probability of being educated for rounds 1–80 and a 10% probability of being educated for rounds 81–120
- A low ability worker has a 0% probability of being educated

You will **choose** to either hire the selected worker from the Green group or hire the selected worker from the Orange group.

Your **payoff** will depend on the Ability level and Education status of the worker you hired.

At the end of each round, you will learn the **ability of the worker you hired.**

After each round, the selected workers will be replaced back into their respective pools, so that any individual worker can be selected more than once across multiple rounds.

In six different rounds you will also bet on whether your choice in a given round is a choice that maximizes your payoff.

We will pay you based on the hiring outcome in **one randomly selected round**.

After you finish, please <u>do not leave</u> as we will wait for everyone to complete the experiment.

Comprehension

Comprehension Questions

Next, you will see 6 comprehension questions, which you need to answer **correctly** in order to proceed to the main part. Similar to the main part of the study, when answering these questions you will be provided with the ability distributions for the Green and Orange groups.

Q1. In the example above, how many workers in the Green group are of high ability?

A. 2
B. 0
C. 4
D. 1

Q2. In the example above, how many workers in the Orange group are of low ability?

A. 3
B. 2
C. 4
D. 0

Q3. From which group(s) will the two randomly selected workers come from?

A. **Both workers from the Green group**
B. **Both workers from the Orange group**
C. **One worker from each group**
D. **All of the above are possible**

Q4. What information will you be provided with about the randomly selected worker from each group?

A. **Both their education status and ability level**
B. **Only their ability level**
C. **Only their education status**
D. **Neither their education status nor ability level**

Q5. Which choice out of four presented below will result in the highest payoff?

A. **Hiring an educated worker of medium ability**
B. **Hiring a not educated worker of high ability**
C. **Hiring a not educated worker of medium ability**
D. **Hiring an educated worker of low ability**

Q6. Suppose you are certain that you made the optimal hiring decision in round 50. Which of the following options will result in the highest chance of receiving the $2 bonus?

A. **Betting 35 out of 50 points**
B. **Betting 25 out of 50 points**
C. **Betting 45 out of 50 points**
D. **Betting 5 out of 50 points**

Pause

Now, you will begin the main part of the study, which includes 120 independent rounds of the task with which you just became familiar.

In each round, please make your hiring choice when instructed to do so and click ‘Next’ to go to the next page. DO NOT hit ‘Back’ or exit the full screen interface. If you have a question at any point, raise your hand.

By the end of 120 rounds, you will answer a few questions before learning about your earnings.

Remember that once you are done, you should wait for others to finish before receiving your payment and leaving.

Go to the next page and begin whenever you are ready.

Tasks — Part 1

```
Click next to begin Round 1
```

In Rounds 21–60 you will additionally see which worker Participant X chose to hire.

In each round, participant X saw the exact same information as you will (distribution of abilities and education statuses of the selected workers)

- For example, if in round 30 you see that the selected Green worker is educated and the selected Orange worker is educated, it means that participant X in round 30 saw that the selected Green worker is educated and the selected Orange worker is educated

Example: Round 1 (Individual)

Distribution of abilities based on Draw 1.

The randomly drawn worker from the Green group is **Educated**
The randomly drawn worker from the Orange group is **Educated**

Which worker do you want to hire?

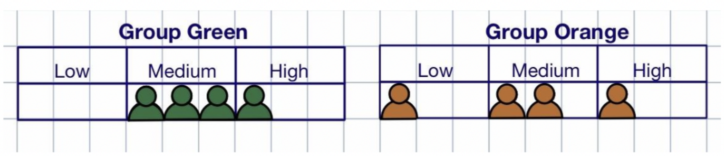
Hire GREEN worker

Hire ORANGE worker

Figure: Decision screen without Participant X (Rounds 1–20)

Example: Round 21 (Social Learning)

Round 22.



Distribution of abilities based on Draw 1.

The randomly drawn worker from the Green group is **Educated**
The randomly drawn worker from the Orange group is **Educated**

In this round, Participant X chose to hire the **Orange** worker.

Which worker do you want to hire?

Hire GREEN worker

Hire ORANGE worker

Figure: Decision screen with Participant X's choice (Rounds 21–60)

Reminder

Reminder:

Regardless of the group that the worker belongs to:

- A high ability worker is educated with 100% chance (for sure)
- **A medium ability worker is educated with 90% chance**
- A low ability worker is educated with 0% chance (for sure not)

Your earnings will be equal to:

- $18 if you hire an Educated and High-ability worker
- $13 if you hire an Educated and Medium-ability worker
- $12 if you hire an Educated and Low-ability worker
- $11 if you hire a Not educated and High-ability worker
- $10 if you hire a Not educated and Medium-ability worker
- $5 if you hire a Not educated and Low-ability worker

Figure: Reminder panel shown on each decision page

Feedback

In this round, you chose to hire a Green Educated worker.
The worker you hired is of **MEDIUM** ability.

If this round is chosen for payment, your earnings would be $13.

Figure: Example feedback after a hiring decision

Betting Interface

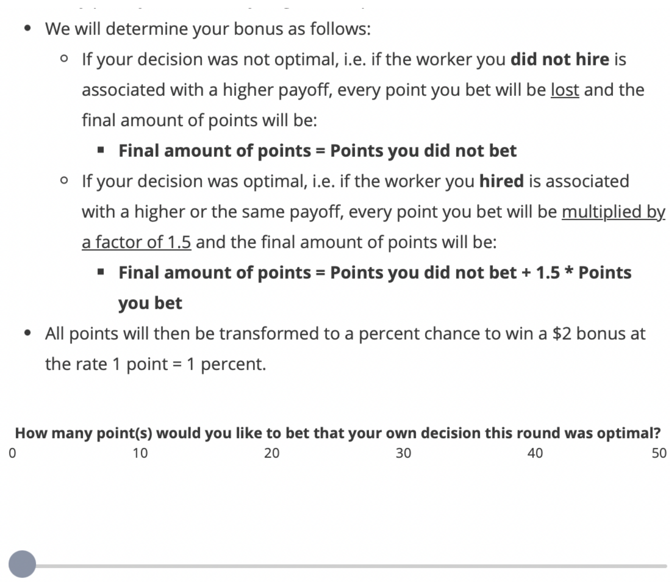


Figure: Confidence elicitation (betting) interface

Pause

You have completed rounds 1–80. The final 40 rounds will use a different distribution of abilities than rounds 1–80.

In addition, medium ability workers will now be educated with 10% chance instead of 90% chance.

Go to the next page and begin whenever you are ready.

Tasks — Part 2

For the remaining rounds of the study, you will also see the hiring recommendations with explanations that participant X wrote for you.

In each round, participant X saw the exact same information as you will (distribution of abilities and education statuses of the selected workers)

- For example, if in round 110 you see that the abilities draw is Draw #3, the selected Green worker is educated, and the selected Orange worker is educated, it means that participant X in round 110 saw that the abilities draw is Draw #3, the selected Green worker is educated, and the selected Orange worker is educated

```
Click next to begin Round 101
```

Participant X's Advice

Scenario 1. Green is Educated and Orange is Educated

Recommendation: Hire **GREEN** worker

Explanation: Both guarantee at least a medium skill worker, but green has a high chance (10/11) of being high skilled.

Scenario 2. Green is Educated and Orange is Not educated

Recommendation: Hire **GREEN** worker

Explanation: Educated has higher payoff than non-educated and guarantees at least medium skill worker with a high chance (10/11) of being high skilled.

Scenario 3. Green is Not educated and Orange is Educated

Recommendation: Hire **ORANGE** worker

Explanation: Educated has higher payoff than non-educated and guarantees medium skill worker. Since green is not educated, it cannot be the high skill worker, so this choice is strictly better.

Scenario 4. Green is Not educated and Orange is Not educated

Recommendation: Hire **ORANGE** worker

Explanation: Both are uneducated, so the worker cannot be high skilled. There are 2 medium workers in orange as opposed to only 1 in green, so orange has a 50% chance of being medium while green only has a 33% chance.

Pause

You have completed the main part of the study. Click next to continue.

Reflection

Reflection on Your Experience

As we conclude this experiment, we would greatly appreciate your insights into how you approached the task at hand. Your responses will be invaluable in enhancing our understanding of decision-making processes.

1. Strategy Reflection:

- Could you please describe the strategy you employed in making your hiring decision in the study?
- Did your approach evolve or change over the course of the 120 rounds? If so, how?

2. Use of Participant X:

- How did you use participant X's hiring choices in Rounds 21–60 when making your own hiring choices?
- How did you use participant X's recommendations and explanations in Rounds 101–120 when making your own hiring choices?

Your candid and thoughtful reflections on these aspects of the experiment are crucial for our study. We appreciate the time and effort you have invested in participating in this research. Note that your response in this question does not affect your final payment.

Final Payoff

Before learning your final payoff, please answer the demographic questions below.

What is your age?

What is your gender identity?

■ Female ■ Male ■ Non-binary ■ Prefer not to say

What is your major? If you have multiple majors, please choose one.

■ Economics ■ Business Economics ■ Joint Mathematics-Economics ■ Other (specify below)

Click next to view your final payoff

Final Payoff

The round that was randomly chosen for the main payment is Round 47

- You chose to hire the Green worker in Round 47, who turned out to be of Medium ability
- Your payoff in Round 47 was $13, so your total earnings in the main part of the study are $13

The round that was randomly chosen for bonus payment is Round 50

- Your bet on your own choice was 34 points
- You chose to hire the Orange worker, who turned out to be of High ability
- Your earnings in Round 50 were $18
- You did not choose to hire the Green worker, who turned out to be of Low ability
- Your earnings in Round 50 would have been $5 were you to choose the Green worker
- Therefore, the final amount of points you have earned from betting is 67, so you have 67 per cent chance of winning a bonus
- As an outcome of the lottery with 67 per cent chance of win, your final bonus in the study is $2

This is the end of the experiment, please remain seated and wait for other participants in the session to finish.

In the meantime, please record the amount you earned on the paper form on your desk — take that form with you as you exit the lab, and then return it back to the experimenter once you have received the payment.

Your earnings from the main part: $13

Your earnings from the bonus: $2

Response record code (for experimenter use only, do not input anything here):

If you have any questions, feel free to email abatmanov@ucsd.edu.

This is the end of the experiment, please remain seated and do NOT leave your desk!

Thank you for participating!