Do Teams Alleviate or Exacerbate Overreaction in Beliefs?*

Ricardo Barahona† Stefano Cassella‡ Kristy A. E. Jansen§ Vincenzo Pezone¶

January 28, 2025

Abstract

We investigate whether teams exhibit increased or reduced overreaction in expectation formation relative to individuals in a financial context. A field analysis of money managers operating both individually and as part of a team shows that teams attenuate their members' overreaction to the recent performance of an investment. In preregistered randomized experiments where expectations about future returns are elicited directly, we find consistent evidence of lower overreaction in teams. A quantitative decomposition of this team effect shows that the reduced overreaction stems from a "self-selection" mechanism, whereby the most biased team members choose to influence the team decision less.

Keywords: Behavioral Finance, Expectation Formation, Extrapolation, Lab Experiment, Overreaction, Performance, Teams
JEL classifications: G41, D91

*We thank Bronson Argyle, Nicholas Barberis, Tabea Bucher-Koenen, Annamaria Conti, Zhi Da, Stephen Dimmock, Joost Driessen, Richard Evans, Anastassia Fedyk, Miguel Ferreira, Rik Frehen, Cary Frydman, Andreas Fuster, Mariassunta Giannetti, Jasmin Gider, Huseyin Gulen, David Hirshleifer, Xing Huang, Cosmin Ilut, Alex Imas, Alexander Kempf, Bart Lambrecht, Yan Liu, Alexandra Niessen-Ruenzi, Cameron Peng, Jan Potters, Melissa Prado, Alessandro Previtero, Lenka Fiala, Pedro Rey-Biel, Emanuele Rizzo, Paolo Sodini, Andrei Shleifer, Oliver Spalt, Per Strömberg, Francisco Urzua, Marno Verbeek, Bas Werker, Jun Yang, and conference and seminar participants at the NBER Behavioral Finance Meeting, AFA, SFS Cavalcade North America, USC Social and Behavioral Finance Conference, Research in Behavioral Finance Conference, ECBE Harvard Business School, FMA, 10th Research Workshop Banco de España-CEMFI, Hong Kong University, City University of Hong Kong, National University of Singapore, Stockholm School of Economics, CFR Research Seminar, and Tilburg University for very useful comments. Views expressed are those of the authors and do not necessarily reflect official positions of Banco de España, De Nederlandsche Bank (DNB), or the Eurosystem.

†Banco de España; ricardo.barahona@bde.es.
‡Tilburg University; s.cassella@tilburguniversity.edu.
§Marshall School of Business, University of Southern California and DNB; kjansen@marshall.usc.edu.
¶Tilburg University; v.pezone@tilburguniversity.edu

1. Introduction

Using surveys and laboratory experiments, a growing body of research shows that overreaction to incoming information is an important feature of how individuals form expectations about the future (Greenwood and Shleifer, 2014; Frydman and Nave, 2017; Bordalo et al., 2020; Hartzmark, Hirshman, and Imas, 2021; Afrouzi et al., 2023). While the evidence of overreaction in individual belief formation is pervasive, in many settings decisions are not made by individuals acting in isolation, but by teams. Crucially, a large experimental literature in economics indicates that teams' behavior can differ systematically from the behavior displayed by individuals.[1] As teams are entrusted with a large variety of tasks, ranging from writing and enforcing the laws of a country to setting its monetary and fiscal policies, or making financial decisions on behalf of their clients, it is important to ask whether the evidence of overreaction that permeates the literature on individuals' expectations also extends to teams.

Studies have pointed out that team members' ability to identify each others' judgment mistakes can be a key determinant of teams' decisions and performance (e.g., Sah and Stiglitz, 1986). However, much work in cognitive psychology shows that in key areas individuals share common intuitive heuristics, which lead to systematic and widespread errors in expectations (e.g., Tversky and Kahneman, 1974; Kahneman, 2003). The shared use of common heuristics can generate similar overreaction for teams and individuals, or even lead to the exacerbation of individual-level overreaction in teams due to group-think (Janis, 1972; Bénabou, 2013). Overall, whether teams attenuate or exacerbate the pervasive overreaction in individuals' expectation formation is the open question that we address in this paper.

We tackle this question in a context of financial decision making. In such a context, past work shows that investors often form beliefs about the future performance of a financial

[1] See for instance Hill (1982), Kerr, MacCoun, and Kramer (1996) and Charness and Sutter (2012) for a survey of the literature. In these studies, teams are compared to individuals in a variety of tasks. In some tasks, teams fare better than individuals (learning and concept-attainment tasks), while individuals achieve better outcomes in other situations (e.g., when dealing with problems that do not have a clear answer).

investment by overreacting to its recent performance (e.g., Greenwood and Shleifer, 2014; Barberis, 2018; Da, Huang, and Jin, 2021).[2] Thus, we study whether teams help alleviate or instead exacerbate the overreaction to past investment returns. We study overreaction in teams both in the field and in the laboratory. In the field, we study the investment behavior of US mutual fund managers, an important category of institutional investors who collectively hold more than $25 trillion in assets under management as of 2023. We document that a large fraction of fund managers makes investment decisions that are consistent with overreaction to past returns of US equities. Importantly, we show that teams of fund managers display a significantly lower overreaction to past investment performance relative to individuals, in a "within-subject" design that isolates the effect of teams on judgment biases by comparing how *the same fund managers* behave when operating individually or in a team.

The field result offers prima facie evidence of reduced overreaction in teams. However, the field results also have limitations. First, these results use trading behavior to infer overreaction, and such behavior could reflect both belief biases and risk preferences. Second, the fact that teams observed in the field are not formed at random leaves open an important question of whether our results can be generalized beyond fund managers' teams.

Therefore, we perform a series of preregistered online experiments. We adopt the state-of-the-art approach recently proposed by Afrouzi et al. (2023) to measure belief overreaction. The key innovation of our design is that some participants are randomly assigned to a two-person "team" treatment, where they are required to agree on a joint prediction via online chat interaction, while other participants are randomly assigned to a standard individual-level task. By comparing the extent of overreaction observed in the individual treatment with the overreaction observed in the team treatment, we can measure how effective a team is in reducing the influence of individual biases on judgment. Our experimental result

[2]While the subjects we study, both in the field and in the experiment, exhibit belief overreaction on average in the context of financial decision making, past work also highlights the importance of belief underreaction. See for instance theoretical work by Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), and Hirshleifer (2001) and more recent empirical evidence by Wang (2021) for financial markets, as well as work on asymmetric updating and the good-news/bad-news effect (Eil and Rao, 2011; Möbius et al., 2014).

aligns with the field result: while on average participants display substantial overreaction to past investment performance, teams display significantly less overreaction than individuals. Our experimental design enables us to uncover the exact mechanism driving the results by introducing additional treatments, allowing for a novel quantitative decomposition of the "team effect" into distinct alternative channels.

We now move to a description of the empirical design of our field evidence. To measure overreaction in the field, we rely on a recent literature linking investor overreaction to extrapolative beliefs (e.g., López-Salido, Stein, and Zakrajšek, 2017; Barberis, 2018; Bordalo, Gennaioli, and Shleifer, 2018; Greenwood, Hanson, and Jin, 2025). This literature shows that investors' beliefs about future asset returns become more optimistic after good asset performance, leading to increased asset demand. This sensitivity of beliefs and demand to past investment returns is indicative evidence of overreaction, because past investment performance is not a strong predictor of future performance. Given the micro-level evidence linking belief overreaction to a recency effect (e.g., Greenwood and Shleifer, 2014; Barberis et al., 2015, 2018; Da et al., 2021; Roussanov, Ruan, and Wei, 2020; Afrouzi et al., 2023), we measure a fund manager's overreaction as the sensitivity of their stock trades to a weighted sum of past stock returns. We assign the higher weight to more recent returns calibrated to the direct estimates of extrapolative beliefs from surveys, similar to, for instance, Jin and Sui (2022).[3]

We first validate our approach by showing that our fund-manager overreaction metric predicts lower future investment performance (raw, risk-adjusted, or benchmark-adjusted), lower future ranking in the cross-section of funds, and a lower future growth of assets under management compared to managers who do not overreact. Given the evidence that investment professionals' compensation and career are influenced by fund performance, ranking, and size (e.g., Guercio and Tkac, 2008; Ibert et al., 2018; Ma, Tang, and Gómez,

[3]When estimating a fund's overreaction metric, we control for key determinants of fund trading behavior, including stock characteristics and flow-induced trading. Our results also remain robust across alternative overreaction metrics. Additionally, as we show, other return-based investor behaviors, such as momentum trading (Jegadeesh and Titman, 1993) and the disposition effect (Odean, 1998), have little correlation with our overreaction metric.

2019; Han, Roussanov, and Ruan, 2021), these results indicate that fund managers who overreact to past stock returns are systematically worse off compared to fund managers who do not overreact. Further tests show that risk aversion, hedging motives (e.g., Breeden, 1979; Campbell and Cochrane, 1999), or preference for skewness (e.g., Barberis and Huang, 2008) are unlikely to explain the negative relationship between our overreaction measure and fund performance.

In the second and central step of the field data analysis, we compare overreaction in solo-managed funds versus team managed funds. Since fund managers are not randomly assigned to a team or an individually managed fund, a sheer between-comparison of overreaction in team funds and solo-managed funds may lack internal validity (Angrist and Pischke, 2009). That is, those managers operating in solo-managed funds and those working as part of a team can exhibit differences in individual-level overreaction, so that the comparison of overreaction between these two groups of managers would not recover the impact of teams for overreaction.

To address this issue, we take advantage of one key feature of our empirical setting; namely, that there are team-managed funds whose members have all at some point in their career managed one or more funds by themselves. Using this sample of managers and funds, we test whether team-based asset management alleviates or exacerbates the overreaction exhibited at the individual level, while keeping the characteristics of managers in team-managed and solo-managed funds constant.

We show that teams attenuate overreaction by up to 50%. This result (i) is robust to various ways of estimating managers' and teams' overreaction; (ii) cannot be explained by differences in compensation structures or bounded rationality constraints faced by fund managers when operating in a team or individually; (iii) is not due to learning, i.e., fund managers joining teams when they have acquired skills, because we also document it in a smaller sample of fund managers operating both in teams and individually *at the same point in time*; (iv) is not due to measurement error, as demonstrated using an instrumental-variable procedure in the spirit of Jegadeesh et al. (2019).

We complement the field evidence with a series of preregistered laboratory experiments. The experimental setup presents a series of attractive features relative to the analysis of the real-world data. The first is that we collect expectations directly, to avoid the confounding effects of preferences that we face in the field when analyzing trading behavior. Second, we have control over the data generating process, allowing us to pin down exactly the extent of belief overreaction among experimental subjects. This rules out the possibility that overreaction is affected by omitted variables that differ between teams and individuals and that may simultaneously affect trading behavior and stock returns. Third, when forming the expectation of next-period investment returns, subjects operating in a team or individually are presented with the same history of past investment returns, thus allowing for a comparison of overreaction in these two groups that holds constant the information set between the two groups. Fourth, by design, participants are randomly assigned to a "Group treatment" or to a "Individual treatment". Moreover, in the former case, every participant is randomly matched with a team member.

We design these experiments via Labvanced, a platform that allows us to create both individual-level and team-level studies, and we run these experiments using subjects recruited via Prolific. In our experiments, inspired by recent work on overreaction by Afrouzi et al. (2023), participants are presented with a time series of returns for a hypothetical US stock, and are then asked to report their forecast regarding the future return of the stock. This task is repeated for twenty rounds, and in each round participants are compensated based on the accuracy of their prediction. The experiment is performed both for individuals (Individual treatment) as well as teams (Group treatment), where participants communicate via a chat box and need to agree on a joint prediction in each round.

After replicating the finding in the literature of widespread overreaction at the individual level (e.g., Bordalo et al., 2020; Afrouzi et al., 2023), we show that overreaction is greatly reduced in the Group treatment relative to the Individual treatment. The economic magnitude of the reduction is also large, since across specifications we find that overreaction is reduced by 30% to 55% in teams relative to individuals, an effect whose magnitude is close

to the one documented in the field.

In the last part of the paper, we perform a quantitative decomposition of the effect of teams on overreaction. To the best of our knowledge, this is the first attempt in the literature to quantify the relative importance that different mechanisms play in shaping team-level predictions and choices in the domain of judgment under uncertainty. We consider three separate mechanisms, motivated by the existing literature. First, to the extent that overreaction in individual expectations is driven by intuitive heuristics (Stanovich and West, 2000; Kahneman and Frederick, 2002; Kahneman, 2003; Bordalo et al., 2020), it can be diminished by replacing intuition with reasoning (Frederick, 2005). Crucially, engaging the scarce resources available for reasoning is more likely when stakes are high (Ilut and Valchev, 2023). We connect this with social psychology theories that suggest that individuals strive to perform well in the presence of others (Hertel, Kerr, and Messé, 2000). Thus we conjecture that a team setting is perceived as a high-stake decision scenario where anticipating the need to justify one's beliefs to others leads to deeper reasoning and lower overreaction (Scott and Lyman, 1968; Lerner and Tetlock, 1999). This leads us to propose an "internal reflection hypothesis", where teams reduce overreaction through their members' increased individual-level reflection. We examine the role of this channel in an additional preregistered treatment that we call Internal Reflection, where we elicit team members' forecasts individually *before* the team-level interaction.

Second, we consider a "self-selection hypothesis" (Silver, Mellers, and Tetlock, 2021; Enke, Graeber, and Oprea, 2023). Simply put, to the extent that individuals who are more prone to overreaction are aware that their beliefs might be biased, they may choose to exert less influence on the decision process when operating in teams.[4] This, in turn, would lead the team-level overreaction to more closely reflect the beliefs of the less biased team members. We measure the impact of this mechanism in an additional preregistered treatment that we label Self-Selection. This treatment differs from the Internal Reflection treatment in that,

[4]The idea behind self selection is rooted in psychological theories of metacognition and calibration. See, for instance, Nelson (1984) and Yaniv, Yates, and Smith (1991).

after eliciting team members' individual-level beliefs, we ask each teammate to cast a vote for their own prediction, with the prediction featuring the most votes becoming the team prediction. Thus, the voting system is meant to allow a team member to express their willingness, or lack thereof, to influence the team.

Lastly, past work argues that the beneficial impact of teams for decision making stems from the process of joint deliberation, exchange of ideas, and mutual scrutiny that takes place through actual interaction between team members (Smith et al., 2009; Minson, Mueller, and Larrick, 2017). Even though there is evidence that belief overreaction can be widespread and thus shared among team members, team interaction can still help filter belief biases, as individuals are more likely to recognize behavioral biases in others than in themselves (Pronin, Lin, and Ross, 2002). Thus, we cast an "external screening hypothesis."

The result of our decomposition is striking. Between 70% and 90% of the gap in overreaction between teams and individuals can be attributed to self-selection, with only a residual role for the two other mechanisms. Both our baseline findings and this decomposition's conclusions are robust to the inclusion of a host of control variables, capturing demographic and financial factors, cognitive sophistication, as well as the use of an alternative overreaction metric that incorporates the notion of recency effects in belief formation that recent work has highlighted.

Our work contributes to a recent strand of literature documenting overreaction at the individual level both in the field (e.g., Bordalo et al., 2020; Da et al., 2021), in the laboratory (e.g., Frydman and Nave, 2017; Afrouzi et al., 2023), and in surveys (e.g., Greenwood and Shleifer, 2014). Our paper differs from previous work in that we study whether *teams* attenuate or exacerbate overreaction.

Our paper is also related to work analyzing consensus beliefs (e.g., Coibion and Gorodnichenko, 2015; Bordalo et al., 2018), as these beliefs represent the prevailing beliefs of a potentially large group of individuals. We differ from this work since these consensus beliefs reflect the average beliefs within a group of individuals that do not need to reach a joint prediction, unlike the economic agents we study in the field and in the lab. Additionally,

recent work points to the coexistence of over- and under-reaction as a function of the information environment, with overreaction arising in more complex environments such as financial markets (Ba, Bohren, and Imas, 2024), and in environments in which the information available to a decision maker, such as past investment performance, is only a weak signal of an unobservable state, e.g., good or bad future performance (e.g., Augenblick, Lazarus, and Thaler, 2024). Our results, broadly in line with this work, offer a new perspective on overreaction in the context of team decision making and provide insights into the mechanisms that can reduce overreaction in a team setting.

Our paper also contributes to the literature in economics that studies the impact of teams on decision making (e.g., Holmstrom, 1982; Sah and Stiglitz, 1986; Charness, Rigotti, and Rustichini, 2007; Sutter, 2009; Cooper and Kagel, 2023). Most of this work studies teams either in an experimental setting or in the field (Harvey et al., 2021; Evans et al., 2024; Fedyk, Patel, and Sarkissian, 2024). We bring to this literature a comprehensive set of new results on the role of teams for belief overreaction, which encompasses both field and experimental evidence. In the lab, a related paper is Enke et al. (2023), who study self-selection in teams and show that such self-selection leads to performance improvement for some cognitive tasks but not for others. We differ from this work in two important ways. First, our decomposition allows us to quantify the relative importance of self-selection for team performance vis-à-vis other potential mechanisms considered in the literature. Second, we study belief overreaction in the context of financial and economic forecasting, using the state-of-the-art cognitive task designed in Afrouzi et al. (2023).

2. Field Evidence

2.1. Data Sources

We focus on US active equity mutual funds. By combining data from several sources, we obtain for each fund and fund manager a panel of holdings of US stocks, which we update

at the end of every quarter based on mutual funds' mandatory filings with the SEC. We also collect information about funds' managerial structure (team versus individual), the identity of the fund managers working at each fund, and fund characteristics such as fund size, investment objectives, fund family, fund returns, and fee structure. Additional details on the procedure used to merge the involved data sources (Thomson Reuters, WRDS, CRSP, Morningstar) are provided in Internet Appendix IA1.

2.2. Fund Manager Sample

To identify the role of mutual fund teams for overreaction, the simplest approach would be to compare overreaction in funds managed by individuals and the overreaction displayed in mutual fund teams. However, this *between-subject* approach would allow to recover the causal role of teams for overreaction only if fund managers were randomly assigned to either individual management or team-based management, or if the managers operating individually or in teams were similar to each other. Unfortunately, fund managers in the mutual fund industry are not assigned at random to funds. What is more, theory work suggests that the managers working alone are likely to differ from those working in teams in terms of their skills, which can correlate with overreaction and are unobservable to the econometrician. For instance, Massa (1998) and Huang, Qiu, Tang, and Xu (2019) suggest that a fund family firm seeking to maximize its overall profits will offer its managers a menu of incentive contracts, so that better managers join an individually managed fund, while less skilled managers join a team.[5]

Due to these concerns, we argue that a between-subject design is inadequate to study the causal impact of teams in the field. Thus, we propose a *within-subject* design, whereby the trading behavior of a team-managed fund is compared with the behavior that *the members of that same team* display when they manage a fund individually. This setup does not require

[5]Other reasons for differences between managers operating in teams and individually also exist, for instance related to managers' social preferences (e.g., Kocher, Strauß, and Sutter, 2006) or skilled managers' preference to work alone rather than in a team with the aim of being more visible to their customers and prospective employers (e.g., Choi, Kahraman, and Mukherjee, 2016).

that individuals are randomly assigned to the "treatment" (i.e., team-based management) or to the "control" (i.e., individual management) and naturally reduces concerns of fundamental differences, observable and unobservable, between managers operating alone and managers operating in a team (Charness, Gneezy, and Kuhn, 2012). To perform this analysis, we identify a sample of mutual fund teams whose managers have all at some point in their career managed a fund by themselves. To ensure that we identify actual team-managed funds, we require that management teams operate for at least four consecutive quarters to be included in the sample. This procedure renders a total sample of 467 unique managers, 847 unique funds, and 308 unique teams observed between 1980 and 2018, which makes up 33% of the full sample of mutual funds in this same period.

Table 1 shows the summary statistics for our sample. Panel A summarizes information on funds managed individually and Panel B presents statistics on team-managed funds. The main message is that there are no sizable differences between individually managed and team-based funds. For instance, the experience that the managers in our sample have while operating a fund individually is 32 quarters, or 8 years, while this equals 7 years for teams. The median number of stocks held by solo and team-managed funds equals 52 and 54, respectively. Moreover, the individually managed and team-managed funds in our sample are also similar with respect to two return-based factors—momentum trading (Jegadeesh and Titman, 1993) and the disposition effect (Odean, 1998)—so that it is unlikely that differences in overreaction between the team-managed and the individually managed funds are due to differences in such factors.

[Place Table 1 about here]

2.3. Measuring Overreaction

To directly measure belief overreaction among mutual fund managers, we would need extensive data on these managers' stock-level expectations. This data, to the best of our knowledge, is not readily available for research. To circumvent this limitation, we rely on

a key insight from the theory and empirical work in behavioral finance (e.g. Greenwood and Shleifer, 2014; Barberis et al., 2015). This work suggests that belief overreaction leads investors to increase their demand for a risky financial asset excessively on the back of good recent asset performance.

Therefore, to obtain a measure of overreaction for the managers of fund j, we estimate a panel-level regression of the fund's trades at the end of quarter $t+1$ on stock return observed at the end of quarter t. More formally, we estimate the following regression:

$$trade_{s,j,t+1} = \alpha_j + \beta_j^X r_{s,t-4\to t} + \gamma_j' C_{s,t} + \theta_{j,t} + e_{s,j,t+1} \quad \text{for} \quad j = 1, \ldots, J. \tag{1}$$

The dependent variable $trade_{s,j,t+1}$ measures changes in a fund's demand for stock s in the next quarter. As such, it is defined as the split-adjusted change in the holdings of stock s held by fund j at time $t+1$, where the trade value is in dollars based on the price of the share at time $t+1$, $P_{s,t+1}$, and normalized by the fund total net assets (TNA) at that time (as in, e.g., Gantchev, Giannetti, and Li, 2024):[6]

$$trade_{s,j,t+1} \equiv \frac{(shares_{s,j,t+1} - shares_{s,j,t}^{split-adj})P_{s,t+1}}{TNA_{j,t+1}}. \tag{2}$$

The main parameter of interest in Equation (1) is β_j^X, which measures the sensitivity of fund managers' demand to past stock performance $r_{s,t-4\to t}$.[7] Past theoretical work (Barberis et al., 2015) refers to those investors for whom $\beta_j^X > 0$ as *extrapolators*. This work shows that in a market with extrapolators, rational traders understand that asset prices can grow too high following a period of good recent asset returns, and thus for these rational investors $\beta_j^X < 0$, i.e., they are more likely to sell an investment after its good performance. We borrow from, e.g., Conrad and Kaul (1998) and Barberis and Shleifer (2003), and refer to a

[6] The output of the regressions is nearly identical if we instead use the lagged price $P_{s,t}$ and lagged total net assets $TNA_{j,t}$ in Equation (2).

[7] In Appendix IA2 we explore alternative assumptions when estimating β_j^X and Appendix IA3.1 shows several tests to underscore the conceptual and empirical differences between our measure of overreaction β^X from Equation (1) and momentum or the disposition effect, most notably the low correlation between them (0.15 and -0.13, respectively).

team or a manager whose $\beta_j^X < 0$ as a *contrarian*. In Section 2.7.1, we show that, consistent with these theoretical insights, extrapolators achieve lower performance, while contrarians obtain higher investment performance. Because of this dichotomy, we equate overreaction to extrapolation in the remainder of this section. Furthermore, to shed light on whether teams deal differently with individual-level behavior that is performance enhancing versus performance destroying, we separately study teams whose member extrapolate past returns versus teams whose members exhibit contrarian behavior.

Equation (2) is estimated using the full history of the holdings of a fund manager or a team,[8] and it includes time fixed effects ($\theta_{j,t}$) and a host of controls ($C_{s,t}$) that influence trading such as standard stock characteristics and controls for flow-induced trading at the stock level (Lou, 2012). We provide details about these controls as well as the definition of the investment universe used to estimate Equation (1) in Internet Appendix IA2.

To better micro-found our measure of overreaction to past investment returns, we measure fund managers' tendency to extrapolate past stock returns based on insights from Greenwood and Shleifer (2014). They use direct data on beliefs from surveys to show that overreaction features a "recency effect", with more recent returns playing a much bigger role in shaping belief biases than distant ones. Thus, following Greenwood and Shleifer (2014), the measure of past returns in Equation (1) is a weighted sum of past quarterly stock returns with exponentially declining weights:

$$r_{s,t-4\to t} \equiv \sum_{j=0}^{k} w_j r_{t-j}, \tag{3}$$

where $w_j \equiv \frac{\lambda^j}{\sum_{i=0}^{k} \lambda^i}$ and λ equals extrapolators' memory parameter. Following the estimates of the parameter λ from Greenwood and Shleifer (2014) (Table 4), we set $\lambda = 0.56$, which corresponds to the average estimate across six distinct surveys of investor expectations. This memory parameter implies a steep decline in weights in Equation (3), whereby the most

[8]The use of the full sample implicitly treats overreaction as time invariant. In Section 2.7.2, we examine whether our results on the role of teams for overreaction are due to changes in overreaction over time. We find this not to be the case.

recent four quarterly stock returns explain 90% of the variation in extrapolators' beliefs about future asset returns.[9] We therefore set $k = 3$, that is, we use four quarters of past stock returns.

2.4. Empirical Approach: Conceptual Framework

In our econometric approach, we estimate the degree to which team members transmit overreaction to a team. To this end, for each team in the restricted sample, we construct a *statistical team counterfactual*. The use of a statistical counterfactual in tests of teams' decision making is common in the experimental literature on teams (e.g., Charness and Sutter, 2012). The idea behind it is to observe how each team member deals with a task alone, and then compare team members' average individual behavior (i.e., the statistical team counterfactual) with the behavior observed when the same individuals complete the same task as part of a team. The comparison between teams and statistical counterfactuals is informative about the value of teams to decisions in that the human capital deployed in both the actual and the counterfactual team is the same, but the synergistic benefits of team members' interactions are absent in the counterfactual. We adopt this approach in our setting.

To this end, we define $\hat{\beta}_j^{TM}$ as the sensitivity of trades to past investment returns observed in a team and $\hat{\beta}_j^{CF}$ as the average sensitivity of trades to returns that is measured among the individual members of a team when they manage alone. We provide summary statistics on $\hat{\beta}_j^{CF}$ and $\hat{\beta}_j^{TM}$, along with difference-in-means tests, in Table 2. The table presents full-sample statistics (Panel A), as well as separate statistics for teams whose individuals do not exhibit overreaction on average (i.e., contrarians, Panel B) and teams that do (extrapolators, Panel C). In case of the former (i.e., $\hat{\beta}_j^{CF} < 0$), we see that individual-level behavior remains virtually unchanged when moving from individual management to teams. In stark contrast, in the case of the latter, the table shows that overreaction is reduced by about 50% (0.18 to

[9] Formally, based on the value of $\lambda = 0.56$ estimated on average across surveys, the relative weight of the first four quarters in the sum is $\sum_{j=0}^{3} \frac{\lambda^j}{\sum_{l=0}^{\infty} \lambda^l} \approx 1 - \lambda^4 = 0.9016$.

0.09), and this reduction is significant at the 5% level.[10]

This difference between how teams retain extrapolative versus contrarian behavior is important, because in the field, as we show in Section 2.7.1, extrapolative behavior is performance-destroying, while contrarian behavior is performance-improving. Thus, these results suggest that teams substantially mitigate belief overreaction, while at the same time teams retain individual-level behavior that is performance-enhancing.

[Place Table 2 about here]

We then formally test how teams inherit the trading behavior of their members with the following regression:

$$\hat{\beta}_j^{TM} = \alpha + \delta_0 \hat{\beta}_j^{CF} + \delta_1 \hat{\beta}_j^{CF} \times D_j^E + \delta_2 D_j^E + \delta_3 C_j + \epsilon_j, \tag{4}$$

where D_j^E is a dummy variable that is equal to one if the counterfactual team is extrapolative (i.e., $\hat{\beta}_j^{CF} > 0$), and C_j indicates a set of team controls. Thanks to this regression framework, we can answer two key questions. First, the regression indicates whether team members' overreaction survives the scrutiny and the aggregation of beliefs that occur in a team. In particular, the sum $\delta_0 + \delta_1$ represents the extent to which the overreaction at the individual level is transmitted to a team. Under the null hypothesis of no effect of teams on decision making, that is, $\delta_0 + \delta_1 = 1$, there is a full transmission of overreaction from team members to the team. The alternative hypothesis is that teams either exacerbate ($\delta_0 + \delta_1 > 1$) or attenuate ($\delta_0 + \delta_1 < 1$) overreaction. In the remainder of the section, we refer to the "team effect" as the evidence of differences between overreaction at the individual level versus at the team level, i.e., evidence that $\delta_0 + \delta_1 \neq 1$.

Second, the regression framework contrasts the transmission of overreaction with the way in which teams inherit contrarian behavior. In this respect, the coefficient δ_0 shows the extent to which teams inherit contrarian behavior present at the individual level. Relatedly,

[10] We report the results of difference-in-means tests for a host of fund characteristics of teams and counterfactuals in Table IA2 of the Appendix. The table shows little evidence of statistically significant differences between characteristics of team funds and individual funds.

the coefficient δ_1 reveals whether teams can distinguish between contrarian behavior that is performance improving and mistakes in investment decisions associated with overreaction.

2.5. Empirical Approach: Instrumental-Variable Methodology

Estimating Equation (4) via OLS likely results in biased estimates, because our right-hand variable is a generated regressor and thus likely subject to measurement error. In the presence of measurement error in one of the regressors (uncorrelated with the error term), the coefficient estimates for that regressor are downward-biased (Theil, 1971). As a result, measurement error could lead to an over-rejection of the null hypothesis of no team effects in favor of the alternative hypothesis of a positive team effect.

To address the issue of measurement error in our regressions, we rely on an instrumental-variable (IV) approach in the spirit of Jegadeesh et al. (2019).[11] Their IV approach leverages the richness of the data to mitigate measurement error in asset pricing tests. Specifically, factor exposures are estimated in two disjoint subsamples, and the resulting estimates are used as independent and instrumental variables in the second-stage regression. The approach satisfies the relevance and exclusion restriction criteria: the variables are highly correlated, but their measurement errors are uncorrelated due to the disjoint samples.

Our setting shares similarities with Jegadeesh et al. (2019) in that our main regressor is estimated from a rich dataset of fund holdings that spans many stocks over many quarters. As a result, we propose a similar approach by estimating $\hat{\beta}_j^{CF}$ on disjointed samples. Specifically, we randomly partition a fund's stock holdings in every quarter into two subsamples. We then separately estimate Equation (1) in both subsamples. Thus, we get two separate estimates of $\hat{\beta}_j^{CF}$ for each team, $\hat{\beta}_j^{CF,1}$ and $\hat{\beta}_j^{CF,2}$. We then use $\hat{\beta}_j^{CF,2}$ as an instrument for $\hat{\beta}_j^{CF,1}$ in the

[11] Gillen, Snowberg, and Yariv (2019) follow a similar approach to address measurement error in experimental studies.

following two-stage least squares regression:

$$\text{1st stage: } \begin{cases} \hat{\beta}_j^{CF,1} & = c_1 + \lambda_{1,0}\hat{\beta}_j^{CF,2} + \lambda_{1,1}\hat{\beta}_j^{CF,2} \times D_j^{E,2} + \lambda_{1,2}D_j^{E,2} + \lambda_{1,3}C_j + u_{1,j} \\ \hat{\beta}_j^{CF,1} \times D_j^{E,1} & = c_2 + \lambda_{2,0}\hat{\beta}_j^{CF,2} + \lambda_{2,1}\hat{\beta}_j^{CF,2} \times D_j^{E,2} + \lambda_{2,2}D_j^{E,2} + \lambda_{2,3}C_j + u_{2,j} \\ D_j^{E,1} & = c_3 + \lambda_{3,0}\hat{\beta}_j^{CF,2} + \lambda_{3,1}\hat{\beta}_j^{CF,2} \times D_j^{E,2} + \lambda_{3,2}D_j^{E,2} + \lambda_{3,3}C_j + u_{3,j} \end{cases} \tag{5}$$

$$\text{2nd stage: } \hat{\beta}_j^{TM} = \alpha + \delta_0\hat{\beta}_j^{CF,1,pred} + \delta_1\hat{\beta}_j^{CF,1,pred} \times D_j^{E,1,pred} + \delta_2 D_j^{E,1,pred} + \delta_3 C_j + \epsilon_j,$$

where *pred* indicates the predicted values from the first-stage regressions. In Internet Appendix IA4, following Jegadeesh et al. (2019), we provide the results of simulations that strongly support the use of the IV in our setting. Therefore, we provide both OLS estimates and IV estimates in our main tables to draw more robust conclusions.

2.6. Results

Table 3 presents the regression results described earlier in this section. We start by estimating the following "simplified" version of Equation (4) in Columns 1, 2, 5, and 6:

$$\hat{\beta}_j^{TM} = \alpha + \delta_0\hat{\beta}_j^{CF} + \delta_1 C_j + \epsilon_j. \tag{6}$$

The regression does not differentiate extrapolative teams from teams of contrarians, but this simpler specification helps illustrate how measurement error can lead to quantitatively inaccurate conclusions on the role of teams for decision making. In Columns 1 and 2, the OLS estimates indicate that the transmission of individual behavior is imperfect. The coefficient δ_0 for $\hat{\beta}_j^{CF}$ is always lower than one, and the null hypothesis of perfect transmission (δ_0=1) is always rejected at the 1% level. However, without further analysis, we cannot rule out that this evidence is due to measurement error in the independent variable. Therefore, we adopt the IV approach described above to this simpler specification and estimate δ_0 again. As expected, the IV estimator of the coefficient of interest δ_0 is larger in magnitude than the

OLS estimate.[12] This increase corroborates the reasoning that the OLS coefficient estimates are shrunk toward zero and that an IV procedure can correct this issue. Based on the IV, we cannot reject the null hypothesis of a full transmission of individual behavior to the team.[13]

Columns 3, 4, 7, and 8 show the results for our main regression of interest. For brevity, we concentrate on the IV results, but the conclusions we draw are similar to those from the OLS results. The coefficient δ_0 captures the transmission of contrarian behavior to the team. The estimate of δ_0 is close to one, and the null hypothesis that contrarian behavior is fully transmitted to the team (i.e., $\delta_0 = 1$) cannot be rejected. In stark contrast, the transmission of overreaction from individuals to teams indicate a large attenuation of individual biases. In particular, the sum $\delta_0 + \delta_1$, which captures the effect of teams on overreaction, is equal to about 0.45. This result is economically important, because it implies that overreaction at the individual level is attenuated in teams by close to 55%. This large positive team effect is also statistically significant. In particular, in all specifications we reject the hypothesis that overreaction is fully transmitted to teams i.e., the hypothesis that $\delta_0 + \delta_1 = 1$, at standard levels of significance.

[Place Table 3 about here]

These results are robust to the inclusion of a large set of controls in the regressions, such as the team's average experience, disposition effect, and investment styles.[14] Notice that, in constructing the team counterfactuals, we assign all the members of a team equal weights. This implicitly assumes that all managers carry the same weight in a team's decision. In reality, some team members may have more influence on the team decision than others. To address this concern, we compute alternative team counterfactuals which assign higher

[12]The critical value for the weak instrument test based on correlations proposed by Nelson and Startz (1990) and applied in Jegadeesh et al. (2019) is 0.057 and is based on the number of teams. We find a correlation between the two sets of disjoint coefficients of 0.53.

[13]In Internet Appendix IA4.5, we show that the results for the IV methodology are not sensitive to the choice of the random sample that is used in the IV procedure. We repeat the analysis for 2,000 randomly disjointed, drawn samples. We find that, as long as the IV passes a standard weak-instrument diagnostic e.g., Nelson and Startz (1990), the results are similar.

[14]We classify investment styles based on Morningstar's categories, which include blend, value, and growth, across three size segments: small-cap, mid-cap, and large-cap stocks.

weights to those team members that have more experience and, as such, may have a larger influence on the team decision. Formally, we compute three alternative team counterfactuals as weighted averages of team members' sensitivity to past returns, where managers' weights are based on: (i) quarters of industry experience; (ii) the number of funds managed; (iii) and the aggregate size of the funds managed prior to team formation.[15] Table IA11 in the Internet Appendix shows that our conclusions remain unchanged for all three metrics. So, taking all findings together, we show evidence of the attenuation of overreaction in teams.

2.7. Robustness

We perform several robustness checks for our field evidence, which have two main goals. Our first goal is to show that our measure of overreaction is consistent with the leading interpretation of overreaction or extrapolation in the literature, namely that it is a bias in belief formation that leads to investment mistakes. We do this by showing that funds that overreact perform worse, grow less, and are less likely to become top performers compared to funds who do not overreact (Section 2.7.1). The second goal is to rule out that confounding effects such as learning and differences in compensation drive the observed reduction of overreaction in teams (Section 2.7.2).

2.7.1. Is Overreaction Costly?

Previous work on overreaction (e.g., Barberis, 2018) suggests that, insofar as overreaction reflects a bias, it should lead fund managers who overreact to past returns to achieve worse outcomes compared to managers who do not overreact. To test whether overreaction makes mutual fund managers indeed worse off, we rely on insights from Guercio and Tkac (2008), Ibert et al. (2018), and Ma et al. (2019), who show that managers' compensation is tied to fund performance as well as the growth of the funds' assets. Based on these insights, we

[15] All three metrics are measured as of the first quarter in which a fund manager appears in the team. For example, if weights are based on quarters of experience, and one manager has 1 quarter of experience and the other one 19 at the time of team formation, then $\hat{\beta}_j^{CF}$ is constructed with weights that are equal to 5% and 95%, respectively.

test the implications of overreaction for fund performance, fund flows, and the likelihood of achieving top-fund status for the full sample of active US mutual funds, which are summarized in Table IA1 in the Internet Appendix.

Figure 1 presents graphical evidence on the relation between overreaction and fund performance. In Panel A, we estimate the sensitivity of funds' trades to past returns over the full sample, and then sort funds into either two groups (extrapolators and contrarians, left panel), or five groups based on quintile breakpoints (right panel). The figure reports style-adjusted average yearly gross fund performance in each of the aforementioned groups. Performance is measured in a variety of ways: (i) raw returns; (ii) fund returns in excess of the benchmark (henceforth, benchmark-adjusted returns); (iii) CAPM alpha; (iv) Fama-French 3-factor model (FF3, Fama and French, 1993); and (iv) Fama-French 5-factor model (FF5, Fama and French, 2015).[16] The graphical evidence in Panel A provides strong support for the negative relation between overreaction and performance. For instance, in the top left panel, we observe that funds that exhibit overreaction feature underperformance relative to their style peers, their benchmarks, and funds with similar risk exposures. On the contrary, funds that display contrarian behavior outperform across a variety of metrics. This positive relation between contrarian behavior and fund performance is consistent with the view that in a market with funds that exhibit overreaction, rational investors trade as contrarians (e.g., Barberis et al., 2015). The right side of Panel A lends further support to our conjecture, because it shows that funds' investment outcomes worsen monotonically when going from a low β_j^X (Q1, contrarian funds, whose average beta is -0.48) to a high β_j^X (Q5, extrapolative funds, whose average beta is 0.78). Panel B shows that results are very similar if we repeat the analysis in a predictive setting, where we ask whether a recursively estimated overreaction metric can predict future fund returns.[17]

[Place Figure 1 about here]

[16] The construction of these performance metrics is standard. Thus, further details are offered in Internet Appendix IA3.2.

[17] This second approach allows to study the relation between overreaction and performance without the look ahead bias that full-sample estimates of overreaction introduce.

To investigate our hypothesis more formally, we perform multivariate predictive regressions of future gross yearly fund returns on the lagged sensitivity of funds' trades to past returns, recursively estimated over the prior 8 quarters. In the regression, we enact the within-style analysis that reflects peer benchmarking in managerial compensation by means of style × time fixed effects. Moreover, we control for fund characteristics that prior literature has linked to fund performance (details are in the Internet Appendix IA3.2).[18]

The results of the analysis are in Table 4. In Columns 1 to 5, we analyze the relationship between fund overreaction and returns, using either simple fund returns, benchmark adjusted returns, CAPM alphas, FF3 alphas, or FF5 alphas. Across all specifications, the relationship between overreaction and performance is negative and statistically significant, which indicates that funds with stronger overreaction underperform their peers.

[Place Table 4 about here]

The results are robust to the inclusion of controls that account for alternative ways in which past stock returns can affect investor trading, namely, fund managers' disposition effect and momentum trading. With regards to the latter, we also perform an additional check in Column 6 of Table 4. In particular, to reduce the concern that our overreaction metric captures fund managers' propensity to follow a momentum strategy, and that the negative impact of overreaction on performance is due to momentum crashes (Daniel and Moskowitz, 2016), we exclude momentum crashes from our estimation.[19] Excluding momentum crashes strengthens the relation between fund's momentum trading and future performance, as expected. More importantly, if overreaction captured momentum trading, we would expect a smaller negative coefficient β_j^X. Instead, the coefficient remains negative, significant, and even slightly larger in magnitude.

Further tests in Internet Appendix IA3.2 show that the negative impact of overreaction

[18] To obtain inference that is robust to unmodeled dependencies in fund returns over time within a fund or across funds at a given point in time, we follow Petersen (2009) and cluster standard errors both at the fund level and by time. Other modeling choices, such as the use of Newey-West standard errors, render the same results.

[19] Following Daniel and Moskowitz (2016), we exclude the years 2001, 2002, and the second and third quarters of 2009.

on managerial financial outcomes extends to other determinants of managers' pay such as fund flows and the achievement of top-fund status. We also show in Internet Appendix IA2.3 that our results are robust to alternative overreaction metrics. These metrics, among others, include a no-momentum variant of our measure, whereby we estimate Equation (1) after excluding winners and losers in the cross-section, and a variant that uses portfolio weight changes as opposed to dollar trades. Finally, in Internet Appendix IA5, we perform a back-of-the-envelope calculation and show that overreaction has an economically meaningful impact on the managers' income.

To distinguish our overreaction measure from a preference-based story, we also investigate whether managers' risk attitudes and preferences help reconcile the negative relation between overreaction and fund performance in Internet Appendix IA3.3. Managers who overreact might for instance do so in an attempt to reduce the volatility of their compensation, or overreaction might lead to lower fund returns because it offers hedging properties against bad states. We find no evidence that this is the case. In addition, to further strengthen the case that our evidence captures the reduction of biases in beliefs rather than preferences, we show that our results remain unchanged when repeating our team regressions by focusing only on first-time purchases to measure fund managers' overreaction (Da et al., 2021; Liao, Peng, and Zhu, 2022).

Overall, these results strengthen the interpretation that our overreaction metric captures a form of bias in investor beliefs that is consistent with the leading interpretation of overreaction in the literature.

2.7.2. Alternative Explanations

Learning and Experience

Suppose that fund managers learn from experience that overreaction hurts their performance, and hence they progressively overreact less. If fund managers manage a fund first individually, and only later join a team, then learning outside of the team could explain why overreaction is reduced in teams.

The learning explanation hinges on the existence of differences in the timing of individual management and team-based management for a given manager. Therefore, in Table 5, we repeat our main tests to examine individual and team-based behavior at comparable points in time. For instance, in the first two columns, we restrict the measurement of teams' overreaction to a period when at least one of the team managers is also managing a fund individually. The number of teams that survive when looking at this case equals 280, which means that teams and individual management often occur at a similar point in time. We then re-estimate the regression of Table 3, and find that the result of bias attenuation in teams remains the same. An alternative approach in Columns 3–6 that restricts the measurement of teams' overreaction to a time frame that starts (ends) when the counterfactual team starts (ends) gives similar results.[20]

[Place Table 5 about here]

A second test relies on the intuition that, for learning to explain team-level reduction in biases, individual management should *precede* team-based management. If so, the reduction of overreaction in teams should only occur in cases in which managers first manage individually and later manage as part of a team. However, as we detail in Internet Appendix IA3.4.1, we do not find that to be the case.

In our third and final test, we probe the learning story by focusing on the sample of *non-learners*, i.e., managers whose overreaction remains largely unchanged over the years in which they operate individually. As we show in Internet Appendix IA3.4.2, overreaction is also significantly reduced within teams composed of non-learners, providing further support that our results are robust to a learning story.

Portfolio Managers' Compensation

[20]To be precise, the start date of the counterfactual team is determined by averaging the start dates of individual operation for each team member, and the end date is determined in a similar manner. To illustrate, suppose that Manager A starts individual fund operations in 2005Q1 and ends in 2015Q1, while Manager B starts operating individually in 2009Q1 and ends in 2019Q1. In this scenario, the start date of the counterfactual team would be 2007Q1, and the end date would be 2017Q1.

A second alternative interpretation of our findings is that the attenuation of overreaction is driven by systematic differences in portfolio managers' compensation between team-managed and solo-managed funds. For instance, if managers have a compensation structure that is more strongly linked to their performance when they manage in teams as opposed to when they manage alone, then these stronger compensation incentives induce managers to use more effortful deliberation in their investment decisions. More effortful decisions could reduce the reliance on heuristics observed in teams (Kahneman and Frederick, 2002), consistent with our main finding.

To address this alternative interpretation, we hand-collect data from each fund's statement of additional information (SAI) through the SEC's EDGAR as of 2006.[21] For each fund that is in our restricted sample as of the first quarter of 2006, we collect data on the compensation structure, based on the last available SEC report before the fund leaves our sample.[22] We closely follow Ma et al. (2019) to extract the compensation structures. Specifically, we generate four dummy variables that are equal to one if managers have (i) (only) a fixed compensation, (ii) compensation based on the performance of the fund, (iii) compensation based on the assets under management of the fund, and (iv) share ownership in their own funds, and zero otherwise. We convert the data at the fund level to the team or manager level by averaging the compensation structures across all the funds they manage. We then compare the compensation structure of the managers working in a team with the compensation structure of the counterfactual, that is, the average compensation structure of the managers that compose the team when they manage alone.

We have a total of 203 teams that appear in our restricted sample as of 2006q1. Of these teams, we were able to collect data on 128 teams, out of which 65 are extrapolative teams. Table 6 compares the compensation structures for all teams and for the contrarians and extrapolators separately. First, taking all teams together, we find that the compensation

[21] We choose 2006, because as of that year mutual funds are required to disclose information on how they compensate portfolio managers.

[22] Ma et al. (2019) show that the compensation structure of a fund is stable over time, so we assume that the last available report is representative for the history of the fund's compensation structure.

structures do not differ between the actual and counterfactual team, except for fund ownership. The counterfactual team is more likely to own shares of their own funds, which implies that, if anything, managers in solo-managed funds have steeper compensation incentives compared to managers who operate in a team. This pattern is similar if we take contrarian and extrapolative teams separately. Because teams face compensation structures that are similar or less tied to performance compared to their solo-managed funds, we take this finding as evidence that differences in compensation incentives between team-managed and solo-managed funds are unlikely to rationalize our result.

[Place Table 6 about here]

Other Channels

Aside from these two important checks, we also consider additional confounders: (i) fund managers may experience a systematic decrease in workload when managing in a team, allowing them to have better judgement at a time of lower cognitive overload; (ii) differences in style between solo-managed and team funds could reconcile the result; (iii) managers' employment at different investment firms when operating individually and as part of a team could explain our result. As we show in Internet Appendix IA3.5, we do not find empirical evidence consistent with these alternative explanations.

3. Experimental Evidence

3.1. Overview

The experiment we describe in this section is a key part of our empirical analysis, because it addresses several shortcomings and limitations of the field analysis. First, while the within-subject analysis in the field helps boost internal validity, it does not address the fact that fund managers are not randomly assigned to teams. If teams observed in the field are only those whose members work well together, our results could overstate the beneficial impact of teams

relative to a situation in which teams are randomly formed. Second, our way of measuring belief overreaction in the field is based on an analysis of trades. Since trading behavior is driven by both beliefs and risk preferences, this raises concerns that our measure of belief biases may be contaminated by investor preferences. Third, in the field analysis where the exact data generating process underlying stock returns is unknown to the econometrician, it is difficult to clearly tell apart a rational reaction of beliefs to recent stock performance from the overreaction of investor beliefs to such performance.

The experiment has another important goal: in our experimental setting, we can design several treatments that not only shed light on the causal role of teams in beliefs and decision making but also study the mechanisms through which teams achieve lower overreaction.

To test the role of teams in alleviating individual level overreaction, we design two experimental treatments, which we label "Individual treatment" and "Group treatment." In both treatments, subjects are shown a figure with the history of the past returns of a hypothetical stock, and are then asked to make predictions on the future return of the stock. Subjects are then compensated based on their accuracy. The cognitive task used in the experiment is close to the one used in Afrouzi et al. (2023) to measure belief overreaction among individuals. However, we introduce a key difference in the decision making process between the two treatments. In the Individual treatment participants complete the prediction task in a setting that does not involve any team interaction. In the Group treatment, instead, participants are required to make joint predictions in each round as part of a randomly formed two-person team, coordinating their choices via a live chat. Comparing the predictions made in these two treatments allows us to recover how teams alter the forecasting process.

In a preview of the results, presented in Section 3.6.2, we find robust evidence that teams exhibit lower overreaction relative to individuals. In the Group treatment, our baseline overreaction metric drops by about 30%, relative to the Individual treatment. This result is robust to the inclusion of several controls in our econometric specification, as well as the use of a different overreaction measure in the spirit of Greenwood and Shleifer (2014).

We then investigate the mechanism through which teams can influence belief overreaction

as part of a quantitative decomposition that we detail in Section 3.7. The decomposition concerns three possible mechanisms, namely (a) internal cognitive reflection; (b) self-selection; (c) external screening. To investigate such mechanisms, we perform two additional preregistered treatments. In the "Internal Reflection" treatment, we gauge the extent to which individuals engage in deeper cognitive reflection when in a team decision making context, by asking them to make an individual prediction *prior* to the team interaction taking place. In the "Self-Selection" treatment, inspired by Enke et al. (2023), we ask participants to first make an individual prediction, and then "submit a vote" in support of their prediction. The prediction with the highest vote is selected as the team prediction. This treatment differs from the Internal Reflection treatment, in that participants are, de facto, allowed to drop out of the team decision process by casting only few votes in support of their prediction. We find the self-selection mechanism to be the most important determinant of teams' ability to reduce overreaction, with the decomposition attributing about 70% percent of the documented impact of teams' overreaction to this mechanism.

3.2. Timeline

The timeline of the experiment is as follows: subjects (i) are hired from Prolific; (ii) read a set of instructions, which include a detailed explanation of the steps outlined below, as well as an explanation of how they will be compensated; (iii) are randomly assigned to one of the four treatments above (and, in the treatments involving a team interaction, are randomly assigned to a teammate); (iv) complete a prediction task in which they forecast the one-period ahead return of a stock based on the history of the past returns of the stock; (v) observe the ex-post return realization, and are awarded a score based on how accurate their ex-ante prediction was; (vi) repeat steps iv and v for twenty times; (vii) take a short exit survey that collects socio-demographics information, financial literacy measures, cognitive reflection scores, and information on investment experience; (viii) they are informed about their final payoff.

3.3. Cognitive Task

In each treatment of our experiment, we elicit subjects' belief overreaction, based on the experimental task in Afrouzi et al. (2023), which we reframe in terms of predicting future stock returns. Participants in the experiment are presented with the history of past returns x_t of a hypothetical US stock. The law of motion of stock returns is an AR(1) process $x_t = \rho x_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. To reflect the fact that past stock returns only have a weak predictive power for future returns, we set $\rho = 0$. As Afrouzi et al. (2023) we set $\sigma = 20$, and compensate participants based on the accuracy of their predictions.[23] In the first round, subjects are presented with a time series of forty past stock returns, and are asked to predict the next-period return. Participants select their prediction using a vertical scroll bar. Subsequently, the actual return that is realized in the next period is announced, and participants receive a score based on the accuracy of their prediction in that round. A total score is also displayed at the end of every round, and it cumulates current and previous scores. The time series of stock returns is then updated to reflect the most recent stock return realization, and the next round begins. For illustration, Figure 2 presents snapshots from the prediction tasks of the Individual and Group treatments in two generic rounds of the experiment.[24] Importantly, participants in our treatments are presented with exactly the same visual cues and information (i.e., the same history of past returns).

[Place Figure 2 about here]

3.4. Incentives

In the experiment, participants receive a completion fee that is proportional to the estimated duration of the experiment, plus a bonus based on the total score that subjects earn in the experiment. This total score is the sum of the score received in each individual round.

[23] One difference between our experiments and Afrouzi et al. (2023) is that, in our experiments, participants repeat the prediction task for 20 rounds, instead of 40. As the treatments involving team interaction are more time consuming, we were concerned about disengagement or fatigue among participants.

[24] Additional figures for the other two treatments, as well as more details on the experimental design, are in Internet Appendix IA6.

Following Dwyer Jr, Williams, Battalio, and Mason (1993); Hey (1994); Afrouzi et al. (2023), the score S_t earned in round t by a subject or a group is a (weakly) increasing function of the accuracy of a prediction, i.e., $S_t = 100 \times \max(0, 1 - |FE_t|/\sigma)$. In this expression, FE_t is the forecast error in round t, and σ is the standard deviation of the returns. We convert this score to a dollar amount by dividing it by 600 (and rounding it to the nearest cent). Participants are made aware of how scores are converted to dollars in the last part of their instruction session, which takes place immediately before the treatment starts.

We adopt the same payment structure in all treatments, in that all participants are offered a completion fee at the same hourly rate of \$9 per hour, and all participants are presented with the same prospect of bonus earned from accuracy. See Internet Appendix IA6 for additional details.

3.5. Logistics

All experiments were conducted using the Prolific US-based sample. The Individual treatment, the Group treatment, and the Internal Reflection treatment were preregistered in May 2024 and fielded in June 2024 (RCT Id: AEARCTR-0013710). In November 2024, we preregistered and fielded the Self-Selection treatment (RCT Id: AEARCTR-0014914).[25]

To ensure high data quality, we took several steps. First, we chose to conduct our experiment on Prolific, which has been shown to provide higher-quality data compared to alternatives such as Amazon Mturk or SONA (Douglas, Ewell, and Brauer, 2023). Second, the participants were pre-screened so that any panelist with less than 98% approval rating on Prolific would not be allowed to participate. Third, we conducted several pilot studies during which we collected participants' overall experience with the experiment and with respect to the clarity of the experimental task. Fourth, when recruiting participants for a new experimental treatment, we excluded all Prolific panelists that had participated in the experiment previously, be that in the same treatment or in another one, and irrespective

[25] We designed the Self-Selection treatment following feedback by seminar participants, which is why it is described in a separate preregistration.

of whether prior participation concerned the main experimental treatment or a previous pilot. Fifth, after the completion of the experiment, we checked for abnormalities in the rate of perfectly correct answers. Given the nature of the experimental task featuring a quasi-continuum of possible predictions, a high such rate should be very rare, and thus we treat it as an indication of unethical behavior. Therefore, we discard the corresponding participants, which are straightforward to identify, from the analysis. Section IA6.6 in the Internet Appendix describes this filtering in detail.

Lastly, to incentivize accurate forecasts, we compensate participants in the experiment both with a completion fee and a bonus for accuracy. Both aspects of the compensation process are the same across four treatments, in that we set the hourly wage participants expect to receive to \$9 per hour, and we set an identical reward for accuracy in all the treatments. After accounting for both components of the compensation, the average earnings across all the treatments are \$6.09, and the average per hour earnings are \$12.17. Thus, the compensation is close to the hourly wage of \$12 that is recommended by Prolific.

We preregistered the two key aspects of the experiment. First, we preregistered the targeted sample size of each treatment: 250 for the Individual treatment and 500 for the three team treatments (that is, 250 teams). Due to the quality checks described above, as well as cases of multiple attempted submissions, the final sample comprises 248 observations for the Individual treatment, 456 observations for the Group treatment, 405 observations for the Internal Reflection treatment, and 403 observations for the Self-Selection treatment, resulting in a final sample of 1,512 observations. In Table 7 we report descriptive statistics on some key demographic characteristics for our two main treatments, with more details available in Internet Appendix IA7. We also perform a balance test of these covariates, and find that in almost all instances we cannot reject the null hypothesis that the means are equal across treatments. Second, we preregistered the treatments we intended to conduct, as well as the key object of interest of our analysis, namely differences in overreaction across the treatments.

[Place Table 7 about here]

3.6. Baseline Tests

3.6.1. Econometric Design

Although the experimental task follows from Afrouzi et al. (2023), the key innovation of our design is that we randomly assign participants to an "Individual" or to a "Group" treatment, henceforth I and G. Participants in the G treatment are randomly assigned to a two-person group; instead, in the I arm, participants work independently of one another. In the G arm, two teammates communicate via a chatbox and need to reach a joint prediction to move to the next round. Hence, this arm of the experiment simulates in a stylized fashion the problem faced by money managers in a team, who ultimately need to converge on a joint investment choice.

To measure belief overreaction, for every participant i in our sample we run the following time-series regression:

$$\widehat{E}_t x_{i,t+1} = \alpha_1 + \beta_i x_t + \varepsilon_{i,t}, \tag{7}$$

where $\widehat{E}_t x_{i,t+1}$ is the participant's prediction regarding the next-period return, and x_t is the most recent return. The coefficient of interest is β_i. Given the law of motion of x_t, the objective benchmark, which corresponds to rational expectations, is $\beta_i = 0$. Conversely, the larger β_i, the larger the participant's overreaction. Based on the evidence in Afrouzi et al. (2023), as well as our field-level analysis, we expect $\beta_i > 0$ for the majority of participants.

We can then test whether group interaction leads to changes in the degree of overreaction by running the following regression:

$$\tilde{\beta}_i = \alpha_2 + \gamma G_i + \delta' X_i + \eta_i, \tag{8}$$

where G_i is a dummy equal to one if the participant is in the Group treatment and zero

otherwise. Our hypothesis, based on the field evidence, is that that $\gamma < 0$, i.e., *group decision making reduces overreaction*. In the regression, we also control for X_i, a vector of participant characteristics based on information collected in a survey conducted after the prediction tasks.

Whereas the regression framework in Equation (7) closely follows the measurement of overreaction in Afrouzi et al. (2023), we also consider an alternative model, whereby beliefs about future asset returns overreact not just to the most recent return realization, but to a weighted sum of past returns (Greenwood and Shleifer, 2014), along the lines of the model used in Section 2.3. We present further details on this alternative metric and the results in Internet Appendix IA7.

3.6.2. Key Result

Table 8 presents our main results. Column 1, in which we estimate Equation (8), shows that the reduction in the overreaction coefficient $\tilde{\beta}$ is 0.092, significant at the 1% level ($t-$statistic$= -2.67$). As discussed in the previous section, this is a large effect, relative to the mean overreaction of the participants in the Individual treatment (0.311, reported at the bottom of the table), as it corresponds to a 30% reduction. In Column 2 we include an extensive set of demographics controls collected through the survey part of the experiment, including, for example, gender, age, and ethnicity dummies.[26] We describe these variables in detail in Section IA6 in the Internet Appendix. The coefficient on the Group treatment dummy is now -0.101 and retains its statistical significance. In Column 3 we also include financial sophistication controls, as more experienced or skilled investors may be less prone to overreaction. Again, the coefficient is unchanged. In Column 4 we further include 13 income

[26]The full list of demographics controls include ethnicity dummies; 1-to-7 measures of trust in others, patience, and risk tolerance; age; a college graduate dummy; an employed dummy; a 1-to-6 measure of political opinions, ranging from 1 for very conservative to 6 for very liberal; dummies for the favorite presidential candidate; a single dummy; the number of people living in the household. Financial sophistication controls include dummies equal to one if the participant holds any stocks, has traded in stocks in the previous year, or holds shares in a mutual fund; a 0-to-4 score assigned on the basis of a cognitive reflection test; a 0-to-3 score assigned on the basis of a financial literacy test.

category fixed effects and US region fixed effects.[27] The estimate from this specification is -0.099 ($t-$statistic$= -2.83$). The stability of the coefficient across specifications suggests that the randomization has been successful, and that the controls are virtually orthogonal to the treatment.

[Place Table 8 about here]

The results of these tests point toward a robust, economically large, and precisely estimated reduction of overreaction when individuals engage in team decision-making relative to when they do not interact with other participants. The next section discusses how additional experimental treatments can be used to interpret and decompose this team effect.

3.7. Mechanism of the Team Effect

3.7.1. Conceptual Framework

To interpret and decompose the team effect, we turn to additional experimental treatments. The effect of teams on overreaction is defined as:

$$\Delta\beta_G = \beta_G - \beta_I$$

where β_G is the average overreaction observed among groups, and β_I is the average overreaction observed among individuals. Going back to our econometric framework, $\Delta\beta_G$ corresponds to γ in Equation (8).[28] We identify three possible channels that map onto a straightforward quantitative decomposition of the team effect $\Delta\beta_G$. These are: (i) Internal Reflection ($\Delta\beta_{IR}$); (ii) Self-Selection ($\Delta\beta_{SS}$); and (iii) External Screening ($\Delta\beta_{ES}$):

$$\Delta\beta_G = \Delta\beta_{IR} + \Delta\beta_{SS} + \Delta\beta_{ES} = (\beta_{IR} - \beta_I) + (\beta_{SS} - \beta_{IR}) + (\beta_G - \beta_{SS}) \tag{9}$$

[27] There is a single participant who claims not to be in the US at the time of the experiment (possibly by mistake, given the Prolific filters), which is why we lose one observation due to collinearity with the region fixed effects.

[28] This correspondence is exact only in the absence of control variables in (8). On the other hand, introducing controls in our regression framework has little to no effect on the estimates of overreaction. Thus, we ignore this aspect here for simplicity of exposition.

Internal Reflection

A large body of work argues that biases in beliefs at the individual level arise in a dual system of cognition, when intuition rather than reasoning is used to make predictions about the future (Stanovich and West, 2000; Kahneman and Frederick, 2002; Kahneman, 2003).[29] Reasoning that is deliberative and reflective can override intuition (Frederick, 2005; Ilut and Valchev, 2023), but its cognitive resources are scarce. As a result, reasoning will be engaged selectively based on the perceived stakes associated with the judgment being considered. We bridge this literature with work in social psychology arguing that individuals wish to perform well in the presence of others (Hertel et al., 2000), and one's anticipation that she may be called on to justify her beliefs, feelings, and actions to others can alter her behavior (Scott and Lyman, 1968; Lerner and Tetlock, 1999). The nexus of these two strands of literature leads us to cast the hypothesis that teams filter belief overreaction because, when preparing for a team-level prediction about future stock returns, individuals are more likely to engage in cognitive reflection.

The internal cognitive reflection that we aim to measure is meant to capture the ex-ante intervention of reasoning in individual judgment that occurs when individuals prepare to participate in a team, rather than the result of the actual social interaction that happens in a team. Thus, in our Internal Reflection treatment (IR), our goal is to collect team members' individual-level beliefs *before* any team interaction actually takes place. From a practical standpoint, this is accomplished by modifying the experimental task used in the Group treatment in three ways: after making our experiment takers aware that they are part of a team, (i) we add a stage of individual-level forecasting before team members are allowed into the chat room where they make a joint forecast; (ii) to avoid cheap talk in the individual-level forecasting task, we incentivize prediction accuracy in this task in the

[29]For instance, in Barberis et al. (1998), Greenwood and Shleifer (2014), and Barberis (2018), overreaction is considered an outcome of the representativeness heuristic. In Malmendier and Nagel (2011), belief distortions arise due to availability. Recently, other explanations for evidence of belief overreaction have been offered at the individual level, e.g., cued memory recall (Enke, Schwerter, and Zimmermann, 2024). Moreover, overreaction can arise more naturally when constraints of bounded rationality are binding (Hong and Stein, 1999; Glaeser and Nathanson, 2017).

same way that we incentivize accuracy in the Individual treatment; (iii) to further prime the accountability of individual-level judgment in the team setting, after both team members have made their prediction individually, and before a chat interaction begins, team members' see each others' predictions.[30]

Self-Selection

Belief overreaction in the Internal Reflection treatment is by construction the average overreaction that is observed at the individual level among the members of a team. Thus, one can interpret it as the overreaction that would be observed if team members' individual beliefs had equal influence on the team choice. Of course, it need not be the case that team members have equal influence on the team decisions, and past work has shown that some members of a team often exert disproportionate control over the team-level choice, while others de-facto self-select out of the team decision by delegating such decision to their teammates.[31] For this reason, in the next treatment, we aim at measuring the influence of self-selection (SS) on the team choice. This is done in the spirit of Enke et al. (2023) by introducing a voting process in each round of the experiment. Specifically, after being teamed up with the respective partners, participants are first asked to provide individual-level predictions completely independently of each other, and are then endowed with 100 votes, which they can use to vote for their prediction. The prediction getting the most votes in a round is selected as the team prediction. Thus, the voting mechanism allows a participant in the team to self-select out the team decision by casting few or no votes for their prediction. Vice versa, the participant can choose to "lead" the team by casting a high number of votes, or even 100.

It is important to stress that the effect of self-selection on overreaction is ex-ante unclear.

[30] Whereas, just like the Group treatment, the Internal Reflection treatment also entails a stage of team-level decision making, these decisions are not directly comparable to the outcome of the Group treatment. This is because in the Group treatment participants do not get to see each others' individual forecasts prior to the team interaction, while team members see this information in the Internal Reflection treatment.

[31] Self-selection can have several underlying explanations, such as individual or collective confidence calibration (Silver et al., 2021; Enke et al., 2023), willingness to avoid social conflict (Moscovici, Lage, and Naffrechoux, 1969), herding or false consensus effects (Ross, Greene, and House, 1977).

To the extent that self-selection and overreaction are negatively related, i.e., those team members who overreact more are the ones who self-select out of the team decision process more aggressively, self-selection will result in a lower overreaction. This would occur if, for example, less biased individuals are more confident in the accuracy of their predictions (Enke et al., 2023). The opposite, however, is true if overreaction and self-selection are positively correlated. In such a case, self-selection might lead to an exacerbation of overreaction.

It is equally important to stress that, while our Group treatment does not feature a formal self-selection stage as part of the experimental task, self-selection can nonetheless occur endogenously in the Group treatment. This is because some participants can still choose to influence more aggressively the decision process via the chat interaction, while some team members let their teammates take the initiative. From this perspective, the chat interaction can serve as a tool for selecting the leading decision maker in the team.

External Screening

The last channel that we consider concerns external screening (ES). The idea, which has a long tradition in psychology, is that a key part of the team decision process revolves around the actual interaction of the team members (e.g., Smith et al., 2009; Minson et al., 2017; Silver et al., 2021). Even though belief overreaction can be widespread and thus shared among team members, team interaction can still help filter belief biases, in that individuals are known to recognize behavioral biases in others more than in themselves, i.e., they exhibit "bias blind spot" (Pronin et al., 2002). Thus, we cast an "external screening hypothesis."

We think of external screening as a mechanism that entails further improving a team's forecast *after* internal cognitive reflection and self-selection. This is natural, since our Internal Reflection and Self-Selection treatments measure the other two mechanisms in the decomposition before or in the absence of any team interaction.[32]

[32]Of course, in our Group treatment, team interactions could further prompt internal reflection beyond what we measure in the Internal Reflection treatment. Similarly, team interaction could aid team members' decision to select out of the decision making process. These more nuanced interactions between distinct mechanisms represent an interesting avenue for future work.

3.7.2. Decomposing the Team Effect

To identify the channel driving the team effect, we augment specification (8) as follows:

$$\tilde{\beta}_i = \alpha_3 + \gamma_{IR} IR_i + \gamma_{SS} SS_i + \gamma_G G_i + \delta' X_i + \epsilon_i, \tag{10}$$

where IR_i, SS_i, and G_i correspond to the Internal Reflection, Self-Selection, and Group treatment, respectively. It is straightforward to map the estimated coefficients to the decomposition outlined in Equation (9). Specifically, we have that γ_{IR} corresponds to $\Delta\beta_{IR} = (\beta_{IR} - \beta_I)$, that is, the effect of internal reflection. The difference between γ_{SS} and γ_{IR} corresponds to $\Delta\beta_{SS} = (\beta_{SS} - \beta_{IR})$, which identifies the self-selection channel. Finally, the difference between γ_G and γ_{SS} measures the effect of external screening, and corresponds to $\Delta\beta_G = (\beta_G - \beta_{SS})$.

We report in Table 9 the results obtained after estimating Equation (10). As before, the omitted group includes participants in the Individual treatment, which correspond to the reference group with respect to which the effects are estimated. At the bottom of the table we also report the differences between the three associated coefficients, which are the key objects of interest to perform decomposition (9).

[Place Table 9 about here]

Column 1 reports the specification without controls, and Columns 2 through 4 are increasingly more conservative. As before, the point estimates change little across specification, again suggesting that demographic characteristics are orthogonal to the treatment dummies.

We do not detect a meaningful role for internal reflection. The coefficient on the IR dummy, which measures this component of the team effect, is small and insignificant. As participants' choices in this treatment are made individually, this result implies that reducing overreaction does require some form of aggregation of individuals' beliefs by the team.

Conversely, the Self-Selection treatment explains *the majority* of the reduction in overreaction due to the Group treatment. In Column 1, the coefficient on the SS dummy is -0.087,

significant at the 5% level. The inclusion of the controls slightly reduces the magnitude of this effect. However, in the most conservative specification of Column 4, the coefficient is still -0.068, and remains significant at the 10% level. The self-selection component, given by $\gamma_{SS} - \gamma_{IR}$, is large in magnitude and significant at the 5% level.

Finally, the effects of the Self-Selection and Group treatments are similar in magnitude. The external screening component $\gamma_G - \gamma_{SS}$ ranges between -0.002 in Column 1 to -0.031 in Column 4, and is never statistically different from zero. Thus, external screening appears to play a limited additional role in reducing overreaction.

We are now ready to quantify decomposition (9). Focusing on the most conservative specification reported in Column 4, we have:

$$\underset{\underbrace{(0.035)}_{\Delta\beta_G}}{-\ 0.098} = \underset{\underbrace{(0.034)}_{\Delta\beta_{IR}}}{-\ 0.001} \underset{\underbrace{(0.034)}_{\Delta\beta_{SS}}}{-\ 0.067} \underset{\underbrace{(0.035)}_{\Delta\beta_{ES}}}{-\ 0.031} \tag{11}$$

Hence, the self-selection channel explains about 70% of the reduction in overreaction. The message of this experimental decomposition is quite stark. As far as the objective is reducing overreaction, there is very little gain in having participants explicitly communicating with each other, as teams do equally well by simply letting the most confident team member make the decision for the group.

In Internet Appendix Table IA17 we perform the same analysis using the alternative approach described in Section 3.6.1, where beliefs are allowed to overreact to a weighted average of past returns. Results are qualitatively and quantitatively similar: we find a significant reduction of overreaction in the Group treatment relative to the Individual treatment. In fact, with this alternative measure, we find that teams reduce overreaction by 55% ($-0.228/0.411$). Moreover, the self-selection channel explains the majority of this effect.

3.8. Additional Outcomes

We report the results of additional tests in Table 10, showing for brevity just the coefficients from the specification with the full slate of controls. Notice that these tests were not preregistered, but we consider them informative nevertheless. The first test is motivated by the observation that, in our experimental setup, both positive and negative sensitivity to past returns lead to belief biases. Thus, in Column 1 we re-estimate Equation (8) using the *absolute value* of $\tilde{\beta}$ as the dependent variable. We find again a negative and statistically significant effect of the Group treatment (coefficient equal to -0.101, significant at the 1% level), suggesting that the lower overreaction documented earlier in this section indeed corresponds to forecasts closer to the rational benchmark.

[Place Table 10 about here]

Although performance per se is not the key object of interest of this analysis, we also briefly discuss whether a reduction in overreaction does indeed correspond to improvements in prediction accuracy. Suppose that joining a team makes participants more biased in dimensions other than overreaction. For example, they could become worse at estimating the mean of the time series (zero in this case). In this case, participants joining a team would be less likely to overreact to recent realizations, but would not necessarily perform better than individuals and, as a result, would not necessarily be better off.

In Columns 2 and 3, we use two different measures of prediction accuracy as dependent variables, namely the mean squared error (MSE) and the mean absolute error (MAE). We find again large and statistically significant effects, with a reduction in MSE and MAE among participants in the Group treatment in the order of 40% of the cross-sectional standard deviation, and 10% to 20% of the average size of the forecast error.[33]

Finally, in Column 4, we find that the reduction in belief biases documented in previous columns indeed corresponds to a higher compensation in teams. Participants in the Group

[33] As before, these statistics, reported at the bottom of the table, refer to the "control" group of participants in the Individual treatment.

treatment earn, on average, about 6 cents more than participants in the Individual treatment. This corresponds to about a quarter of the sample standard deviation for participants in the Individual treatment, and to about 6.5% of the sample mean.[34]

Interestingly, we find that participants in the Group treatment earn the highest dollar amount and exhibit the lowest forecast errors of *all* the participants, including those in the Self-Selection treatment (Internet Appendix Table IA16). Moreover, the distribution of their predictions is characterized by the smallest dispersion. One possible interpretation is that external reflection may be beneficial in dimensions other than mitigating overreaction, beyond the role played by self-selection. In Appendix IA7 we make use of the content of the chat box to show that in the Group treatment participants are more likely to challenge predictions made by teammates when these predictions are more extreme and further from rational expectations. Studying when different forms of belief aggregations may add value is beyond the scope of this paper, as the focus of our analysis is overreaction (which is the outcome of interest that we explicitly preregistered). However, it represents an interesting avenue for future research.

4. Conclusion

Evidence of belief overreaction is pervasive in economics, in settings ranging from macroeconomics expectations formation to beliefs about the returns of financial investments. However, most of the extant evidence concerns belief formation at the individual level, and it is an open question whether the belief overreaction is attenuated or amplified when teams rather than individuals cast predictions about the future.

To address this question, in this paper we adopt a two-pronged approach, as we study the role of teams for overreaction both in the field and in the lab. In the field, we show that teams of professional fund managers in the US heavily attenuate the adverse impact

[34]Recall that the per-round score is $100 \times \max(0, 1 - |\Delta|/\sigma)$, where $\sigma = 20$. Hence, the effects on compensation are somewhat attenuated, relative to those on forecast accuracy, by the fact that the score earned by participants in each round is censored at zero for predictions that are more than 20 units away from the correct answer.

of extrapolation-induced overreaction. We establish this result in a novel within-subject approach, whereby we compare overreaction in teams with the overreaction exhibited by the members of the team when managing a fund individually.

In the lab, we show in an experimental setting how teams mitigate overreaction in a different environment. In this experiment, we compare overreaction of individuals versus teams in a between-subject design involving a stock forecasting task. We find remarkably similar results to the ones documented in the field, with overreaction decreasing by about a third when participants are asked to collaborate in teams. Aside from eliminating any residual concerns regarding the lack of random assignment of our field evidence, and from allowing a more direct measure of beliefs, the experiment also allows us to gather a deeper understanding of the channels driving the results. We are the first to propose a formal quantitative decomposition of the team effect on overreaction. We perform such a decomposition by running additional experimental treatments, and establish that self-selection plays a dominant role in explaining our results relative to alternative mechanisms such as internal reflection or external screening.

We consider this paper as a first step towards a deeper understanding of how organizational structures contribute to the attenuation or exacerbation of behavioral biases. Studying the implications of different forms of belief aggregation for outcomes beyond overreaction represents an interesting avenue for future research.

References

Afrouzi, H., S. Y. Kwon, A. Landier, Y. Ma, and D. Thesmar (2023). Overreaction in expectations: Evidence and theory. *Quarterly Journal of Economics 138*(3), 1713–1764.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Augenblick, N., E. Lazarus, and M. Thaler (2024). Overinference from weak signals and underinference from strong signals. *Quarterly Journal of Economics*. Forthcoming.

Ba, C., J. A. Bohren, and A. Imas (2024). Over-and underreaction to information. *Available at SSRN 4274617*.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of Behavioral Economics: Applications and Foundations 1*, pp. 79–175. Elsevier.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics 115*(1), 1–24.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2018). Extrapolation and bubbles. *Journal of Financial Economics 129*(2), 203–227.

Barberis, N. and M. Huang (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review 98*(5), 2066–2100.

Barberis, N. and A. Shleifer (2003). Style investing. *Journal of Financial Economics 68*(2), 161–199.

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*(3), 307–343.

Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets. *Review of Economic Studies 80*(2), 429–462.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020). Overreaction in macroeconomic expectations. *American Economic Review 110*(9), 2748–82.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *Journal of Finance 73*(1), 199–227.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics 7*(3), 265–296.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), 205–251.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

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

Charness, G., L. Rigotti, and A. Rustichini (2007). Individual behavior and group membership. *American Economic Review 97*(4), 1340–1352.

Charness, G. and M. Sutter (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives 26*(3), 157–76.

Choi, D., B. Kahraman, and A. Mukherjee (2016). Learning about mutual fund managers. *Journal of Finance 71*(6), 2809–2860.

Cici, G. (2012). The prevalence of the disposition effect in mutual funds' trades. *Journal of Financial and Quantitative Analysis 47*(4), 795–820.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–2678.

Conrad, J. and G. Kaul (1998). An anatomy of trading strategies. *Review of Financial Studies 11*(3), 489–519.

Cooper, D. J. and J. H. Kagel (2023). Using team discussions to understand behavior in indefinitely repeated prisoner's dilemma games. *American Economic Journal: Microeconomics 15*(4), 114–145.

Da, Z., X. Huang, and L. J. Jin (2021). Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics 140*(1), 175–196.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under-and overreactions. *Journal of Finance 53*(6), 1839–1885.

Daniel, K. and T. J. Moskowitz (2016). Momentum crashes. *Journal of Financial Economics 122*(2), 221–247.

Douglas, B. D., P. J. Ewell, and M. Brauer (2023). Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. *PLOS ONE 18*(3), e0279720.

Dwyer Jr, G. P., A. W. Williams, R. C. Battalio, and T. I. Mason (1993). Tests of rational expectations in a stark setting. *The Economic Journal 103*(418), 586–601.

Eil, D. and J. M. Rao (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics 3*(2), 114–138.

Enke, B., T. Graeber, and R. Oprea (2023, July). Confidence, self-selection, and bias in the aggregate. *American Economic Review 113*(7), 1933–1966.

Enke, B., F. Schwerter, and F. Zimmermann (2024). Associative memory, beliefs and market interactions. *Journal of Financial Economics 157*, 103853.

Evans, R. B., M. Prado, A. E. Rizzo, and R. Zambrana (2024). Identity, diversity, and team performance: Evidence from US mutual funds. *Management Science*. Forthcoming.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fedyk, A., S. Patel, and S. Sarkissian (2024). Managerial structure and performance-induced trading. *Working Paper*.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 25–42.

Frydman, C. and G. Nave (2017). Extrapolative beliefs in perceptual and economic decisions: Evidence of a common mechanism. *Management Science 63*(7), 2340–2352.

Gantchev, N., M. Giannetti, and R. Li (2024). Sustainability or performance? Ratings and fund managers' incentives. *Journal of Financial Economics 155*, 103831.

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

Glaeser, E. L. and C. G. Nathanson (2017). An extrapolative model of house price dynamics. *Journal of Financial Economics 126*(1), 147–170.

Greenwood, R., S. G. Hanson, and L. Jin (2025). Reflexivity in credit markets. *Journal of Finance*. Forthcoming.

Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *Review of Financial Studies 27*(3), 714–746.

Guercio, D. D. and P. A. Tkac (2008). The effect of morningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 907–936.

Han, X., N. L. Roussanov, and H. Ruan (2021). Mutual fund risk shifting and risk anomalies. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Hartzmark, S. M., S. D. Hirshman, and A. Imas (2021). Ownership, learning, and beliefs. *Quarterly Journal of Economics 136*(3), 1665–1717.

Harvey, C. R., Y. Liu, E. K. Tan, and M. Zhu (2021). Crowding: Evidence from fund managerial structure. *Working Paper*.

Hertel, G., N. L. Kerr, and L. A. Messé (2000). Motivation gains in performance groups: paradigmatic and theoretical developments on the köhler effect. *Journal of Personality and Social Psychology 79*(4), 580.

Hey, J. D. (1994). Expectations formation: Rational or adaptive or…? *Journal of Economic Behavior & Organization 25*(3), 329–349.

Hill, G. W. (1982). Group versus individual performance: Are n+ 1 heads better than one? *Psychological Bulletin 91*(3), 517.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *Journal of Finance 56*(4), 1533–1597.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324–340.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance 54*(6), 2143–2184.

Huang, J., Z. Qiu, Y. Tang, and X. Xu (2019). Why do we have so many funds? The organizational structure of mutual fund families. *Working Paper*.

Ibert, M., R. Kaniel, S. Van Nieuwerburgh, and R. Vestman (2018). Are mutual fund managers paid for investment skill? *Review of Financial Studies 31*(2), 715–772.

Ilut, C. and R. Valchev (2023). Economic agents as imperfect problem solvers. *Quarterly Journal of Economics 138*(1), 313–362.

Janis, I. L. (1972). *Victims of Groupthink: A Psychological Study of Foreign Policy Decisions and Fiascoes*. New York: Houghton Mifflin.

Jegadeesh, N., J. Noh, K. Pukthuanthong, R. Roll, and J. Wang (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics 133*(2), 273–298.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jin, L. J. and P. Sui (2022). Asset pricing with return extrapolation. *Journal of Financial Economics 145*(2), 273–295.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review 93*(5), 1449–1475.

Kahneman, D. and S. Frederick (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment*, 49–81.

Kerr, N. L., R. J. MacCoun, and G. P. Kramer (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review 103*(4), 687.

Kocher, M., S. Strauß, and M. Sutter (2006). Individual or team decision-making—causes and consequences of self-selection. *Games and Economic Behavior 56*(2), 259–270.

Lerner, J. S. and P. E. Tetlock (1999). Accounting for the effects of accountability. *Psychological Bulletin 125*(2), 255.

Liao, J., C. Peng, and N. Zhu (2022). Extrapolative bubbles and trading volume. *Review of Financial Studies 35*(4), 1682–1722.

López-Salido, D., J. C. Stein, and E. Zakrajšek (2017). Credit-market sentiment and the business cycle. *Quarterly Journal of Economics 132*(3), 1373–1426.

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies 25*(12), 3457–3489.

Ma, L., Y. Tang, and J.-P. Gómez (2019). Portfolio manager compensation in the US mutual fund industry. *Journal of Finance 74*(2), 587–638.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics 126*(1), 373–416.

Massa, M. (1998). Why so many mutual funds? Mutual fund families, market segmentation and financial performance. *Working paper*.

Minson, J. A., J. S. Mueller, and R. P. Larrick (2017). The contingent wisdom of dyads: When discussion enhances vs. undermines the accuracy of collaborative judgments. *Management Science 64*, 4177–4192.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2014). Managing self-confidence. *NBER Working paper 17014*.

Moscovici, S., E. Lage, and M. Naffrechoux (1969). Influence of a consistent minority on the responses of a majority in a color perception task. *Sociometry*, 365–380.

Nelson, C. and R. Startz (1990). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *The Journal of Business 63*(1), S125–S140.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin 95*(1), 109–133.

Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance 53*(5), 1775–1798.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435–480.

Pronin, E., D. Y. Lin, and L. Ross (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin 28*(3), 369–381.

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

Roussanov, N. L., H. Ruan, and Y. Wei (2020). Mutual fund flows and performance in (imperfectly) rational markets? *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Sah, R. K. and J. E. Stiglitz (1986). The architecture of economic systems: Hierarchies and polyarchies. *American Economic Review 76*(4), 716–727.

Scott, M. B. and S. M. Lyman (1968). Accounts. *American Sociological Review*, 46–62.

Silver, I., B. A. Mellers, and P. E. Tetlock (2021). Wise teamwork: Collective confidence calibration predicts the effectiveness of group discussion. *Journal of Experimental Social Psychology 96*, 104157.

Smith, M. K., W. B. Wood, W. K. Adams, C. Wieman, J. K. Knight, N. Guild, and T. T. Su (2009). Why peer discussion improves student performance on in-class concept questions. *Science 323*, 122–124.

Stanovich, K. E. and R. F. West (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences 23*(5), 645–665.

Sutter, M. (2009). Individual behavior and group membership: Comment. *American Economic Review 99*(5), 2247–57.

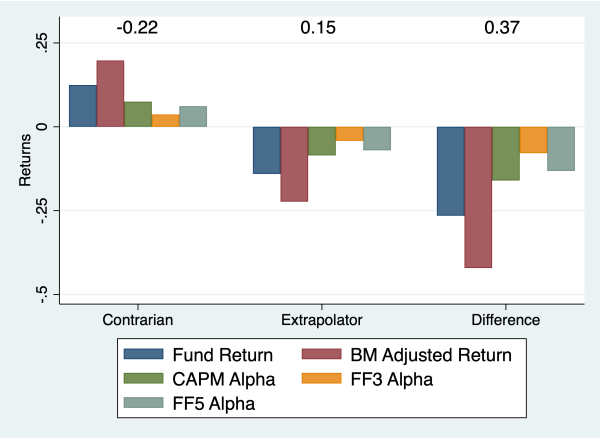
Theil, H. (1971). *Principles of Econometrics*. New York: Wiley.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131.

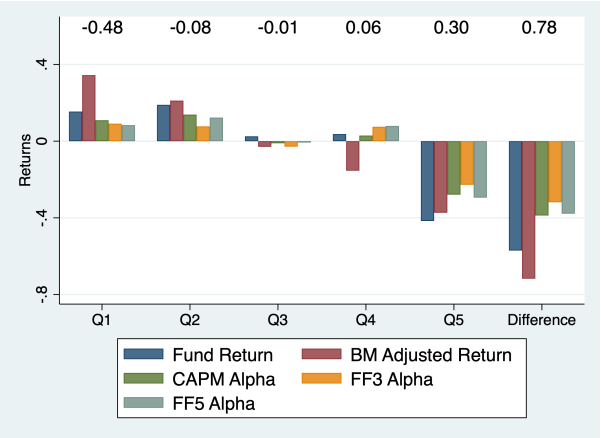
Wang, C. (2021). Under-and overreaction in yield curve expectations. *Available at SSRN 3487602*.

Yaniv, I., J. F. Yates, and J. K. Smith (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin 110*(3), 611.

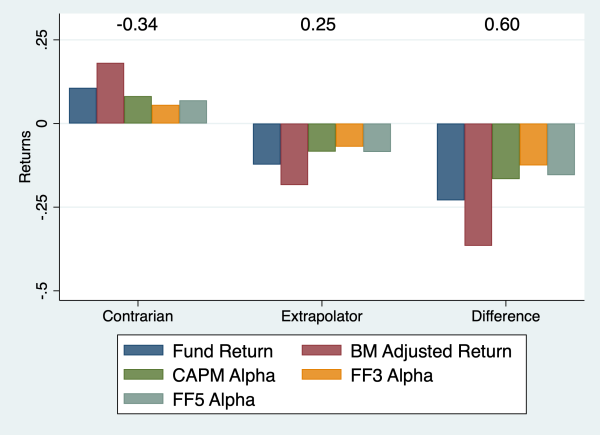
Figure 1. **Field Evidence – Overreaction and Performance**: In Panel A we estimate funds' overreaction over the full sample, and then sort funds into either two groups (A1, extrapolators and contrarians, left panel), or five groups based on quintile breakpoints (A2, right panel). Each subpanel reports style-adjusted average yearly gross fund performance in each of the aforementioned groups, as well as the difference between the top and the bottom group. Performance is measured in a variety of ways (from left to right in each group): (i) raw returns (blue); (ii) benchmark-adjusted returns (red); (iii) CAPM alpha (green); (iv) Fama-French 3-factor alpha (FF3, Fama and French, 1993) (orange); (v) Fama-French 5-factor alpha (FF5, Fama and French, 2015) (grey). Above each group, we report the respective average β_j^X. Panel B repeats the analysis in a predictive setting to examine whether a recursively estimated overreaction metric predicts future fund returns. Further details on the construction of the performance metrics are in Internet Appendix IA2.



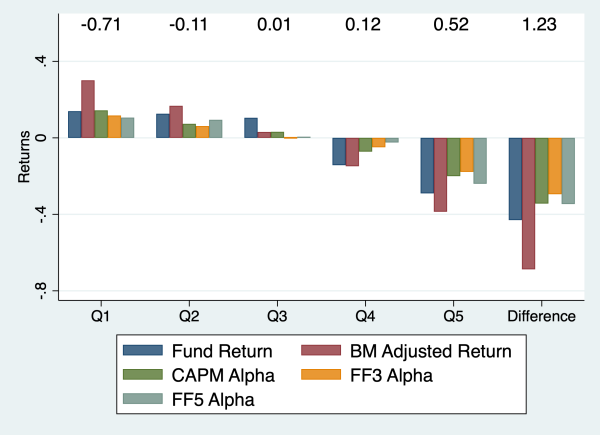
(a) Panel A1: Extrapolators and contrarians (full sample)



(b) Panel A2: Quintile sorts (full sample)

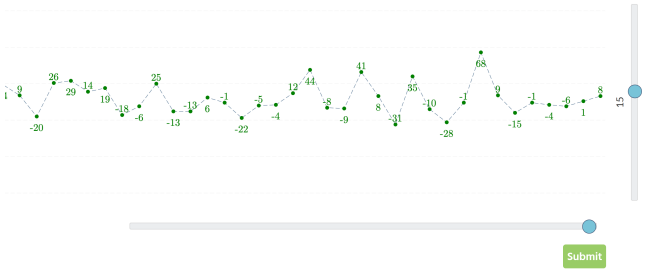


(c) Panel B1: Extrapolators and contrarians (recursive)

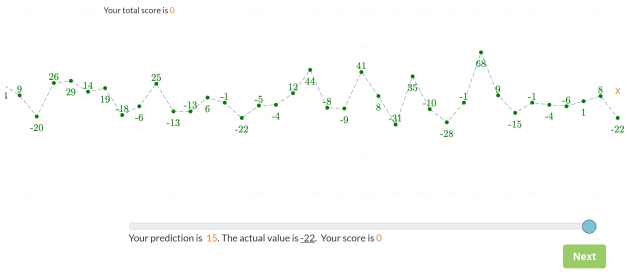


(d) Panel B2: Quintile sorts (recursive)

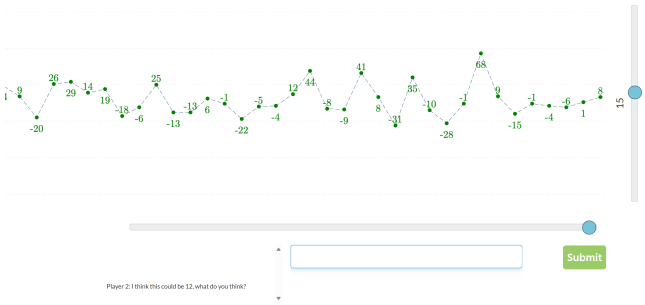
Figure 2. **Experiment – Task**: This figure shows an example of the return time series of a hypothetical US stock presented to participants as part of our online experiments. A horizontal slider allows participants, be it in a team or in an individual task, to scroll back to the beginning of the series, which initially includes forty return realizations. A vertical slider enables participants to make their prediction for the next period's return by moving the marker. The task in the individual treatment and the team treatment is identical, with the exception that in the team treatment participants are also allowed to chat with each other in real time, and they can proceed in the experiment only after reaching a joint forecast of the stock's future return.



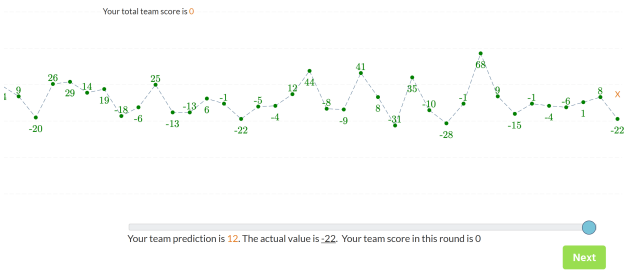
(a) Panel A1: Individual treatment - Prediction Task



(b) Panel A2: Individual treatment - Feedback and Score



(c) Panel B1: Group treatment - Prediction Task



(d) Panel B2: Group treatment - Feedback and Score

Table 1. **Field Evidence – Summary Statistics**: Panel A reports statistics for the solo-managed funds and Panel B for the team-managed funds. *Fund TNA* is the fund's total net asset value (millions); *Manager Experience* is the experience of managers (quarters); *Fund Age* is the age of the fund since inception (years); *Expense Ratio* is the fund's total expense ratio (percentage points); β_{MOM} is the fund's exposure to momentum; *Disposition* is the measured disposition effect. Details on the measurement of β_{MOM} and the disposition effect are in Internet Appendix IA3.1.

Panel A: Solo-Managed Funds

Mutual Fund Characteristics	Mean	St. Dev.	5th Pct.	Median	95th Pct.
Fund TNA	1017.82	3327.68	12.03	232.15	3893.80
Number of Managers	1.00	0.00	1	1	1
Manager Experience	38.44	25.55	7.00	31.95	88.00
Number of Stocks	89.88	193.98	18	52	215
Fund Age	13.92	13.38	2.27	9.74	43.40
Expense Ratio	1.25	0.46	0.57	1.21	2.09
β_{MOM}	0.03	0.13	-0.15	0.02	0.24
Disposition	-0.03	0.08	-0.18	-0.02	0.09

Panel B: Team-Managed Funds

Mutual Fund Characteristics	Mean	St. Dev.	5th Pct.	Median	95th Pct.
Fund TNA	1200.26	2914.56	20.28	300.62	4890.17
Number of Managers	2.09	0.26	2	2	3
Manager Experience	31.36	18.18	8.00	28.33	70.17
Number of Stocks	103.07	214.42	18	54	267
Fund Age	14.93	13.34	2.29	9.83	43.92
Expense Ratio	1.26	0.44	0.56	1.24	2.03
β_{MOM}	0.04	0.15	-0.18	0.02	0.29
Disposition	-0.03	0.09	-0.20	-0.01	0.10

Table 2. **Field Evidence – Overreaction in Solo-Managed Funds versus Team-Managed Funds**: This table compares the team's sensitivity to past returns, $\hat{\beta}_j^{TM}$, with its statistical counterfactual, $\hat{\beta}_j^{CF}$, i.e., the average level of sensitivity to past returns among team members when they manage a fund alone. We report the results for all teams combined (Panel A), for contrarian teams (Panel B), and for extrapolative teams (Panel C). A team consists mainly of contrarians if $\hat{\beta}_j^{CF} \leq 0$ and of extrapolators if $\hat{\beta}_j^{CF} > 0$. We report the mean, standard error of the mean (s.e.), the corresponding t-stat, the total number of teams in each group (N), and the median number of observations that are used to estimate the respective sensitivities (N^X) for each group.

Panel A: All Teams

	Mean	s.e.	t-stat	N	N^X
$\hat{\beta}_j^{CF}$	-0.011	0.022	-0.489	308	7773
$\hat{\beta}_j^{TM}$	-0.034	0.028	-1.200	308	3898
Difference $\hat{\beta}_j^{CF}$ - $\hat{\beta}_j^{TM}$	0.023	0.026	0.891		

Panel B: Contrarian Teams

	Mean	s.e.	t-stat	N	N^X
$\hat{\beta}_j^{CF}$	-0.224	0.029	-7.593	143	7506
$\hat{\beta}_j^{TM}$	-0.178	0.043	-4.127	143	3825
Difference $\hat{\beta}_j^{CF}$ - $\hat{\beta}_j^{TM}$	-0.046	0.034	-1.374		

Panel C: Extrapolative Teams

	Mean	s.e.	t-stat	N	N^X
$\hat{\beta}_j^{CF}$	0.184	0.024	7.815	165	7916
$\hat{\beta}_j^{TM}$	0.090	0.035	2.589	165	4008
Difference $\hat{\beta}_j^{CF}$ - $\hat{\beta}_j^{TM}$	0.094	0.039	2.414		

Table 3. **Field Evidence – Transmission of Overreaction from Individuals to Teams**: In this table we estimate the transmission of overreaction from solo managers to their respective teams. To this end, we estimate the following regression: $\hat{\beta}_j^{TM} = \alpha + \hat{\beta}_j^{CF}(\delta_0 + \delta_1 D_j^E) + \delta_2 D_j^E + \delta_3 C_j + \epsilon_j$. In the regression, $\hat{\beta}_j^{TM}$ measures the team's actual overreaction, while $\hat{\beta}_j^{CF}$ is the team's counterfactual overreaction, based on team members' trading behavior when managing alone. D_j^E is an indicator variable that is equal to 1 when the members of the team exhibit overreaction on average when managing a fund alone. Team controls (C_j) are the time-series average log TNA of the team-managed portfolio, the time-series average log experience of the team members, the time-series average log of fund age of the team-managed portfolio, the time-series average of the exposure to momentum of the team, the time-series average of the disposition effect of the team, and style fixed effects, and are included as reported. Columns (1) to (4) perform the analysis using OLS. Columns (1) and (2) estimate a simpler model without any interaction terms. Thus, in these columns, the coefficient δ_0 captures the transmission of all individual-level return-based trading behavior (be it extrapolative or contrarian) to the team. Columns (3) and (4) estimate the full model. In these columns, we analyze separately the transmission of individual contrarian behavior (δ_0), the transmission of extrapolative behavior ($\delta_0 + \delta_1$), and the difference between the two (δ_1). Columns (5) to (8) apply an IV methodology in the spirit of Jegadeesh et al. (2019), described in detail in Section 2.5. Standard errors are in brackets. Significance: ***99%, **95%, *90%.

	OLS				**IV**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\beta}_j^{CF}$	0.6219***	0.5593***	0.9126***	0.8481***	0.9338***	0.8534***	1.1703***	1.1703***
	[0.0654]	[0.0663]	[0.1012]	[0.1030]	[0.1160]	[0.1179]	[0.1658]	[0.1682]
$\hat{\beta}_j^{CF} \times D_j^E$			-0.6737***	-0.6093***			-0.7167***	-0.6706***
			[0.1493]	[0.1475]			[0.2172]	[0.2080]
D_j^E			0.0221	-0.0036			0.022	-0.1224
			[0.0570]	[0.0573]			[0.1820]	[0.1858]
Style fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Team controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	308	307	308	307	308	307	308	307
Adj. R-squared	0.2253	0.264	0.2697	0.3002	0.0371	0.1011	0.0746	0.128
Hypothesis testing:								
p-value $H_0 : \delta_0 = 1$	0.000	0.000	0.389	0.141	0.568	0.214	0.304	0.311
p-value $H_0 : \delta_0 + \delta_1 = 0$			0.030	0.027			0.043	0.019
p-value $H_0 : \delta_0 + \delta_1 = 1$			0.000	0.000			0.015	0.018

Table 4. **Field Evidence – Overreaction and Investment Performance**: This table reports coefficient estimates from a regression of future yearly fund performance on the lagged fund's overreaction. We analyze fund performance using the following metrics: raw returns in Column 1; returns in excess of the benchmark (BM Adj. Ret) in Column 2; CAPM Alpha in Column 3; Fama-French 3-factor model Alpha in Column 4; Fama-French 5-factor model Alpha in Column 5. In Column 6, we redo the analysis from Column 5, but we exclude from the sample years of momentum crashes (Daniel and Moskowitz, 2016). Controls in the multivariate regressions are: the fund's expense ratio; the log of the fund's total net assets (TNA) and its square; the log of fund age; the fund's CRSP turnover ratio; the log number of stocks in the fund's portfolio; lagged fund flows; average manager experience; β_{MOM}, the fund's exposure to the momentum risk factor; and the fund's measured disposition effect. All regressions also control for style-quarter fixed effects. Data frequency is quarterly. Standard errors are clustered by quarter and at the fund level, and reported in brackets. More details on the control variables can be found in Internet Appendix IA2. Significance: ***99%, **95%, *90%.

	Raw Return	**BM Adj. Return**	**CAPM Alpha**	**FF3 Alpha**	**FF5 Alpha**	**FF5 Alpha NC**
	(1)	(2)	(3)	(4)	(5)	(6)
β_j^X $(t-1)$	-0.237*	-0.355**	-0.298**	-0.331***	-0.372***	-0.405***
	[0.135]	[0.151]	[0.134]	[0.112]	[0.108]	[0.109]
Expense Ratio $(t-1)$	-0.024	0.17	-0.308	-0.051	0.332*	0.309
	[0.197]	[0.287]	[0.198]	[0.182]	[0.196]	[0.202]
Log Fund TNA $(t-1)$	-0.192	-0.670***	-0.339**	-0.235	-0.208	-0.214
	[0.174]	[0.220]	[0.161]	[0.145]	[0.151]	[0.147]
Log Fund TNA^2 $(t-1)$	0.015	0.044***	0.026**	0.018	0.023*	0.024**
	[0.013]	[0.017]	[0.012]	[0.011]	[0.012]	[0.011]
Log Fund Age $(t-1)$	-0.029	-0.036	-0.101	-0.073	-0.109	-0.128
	[0.087]	[0.118]	[0.087]	[0.074]	[0.076]	[0.079]
Fund Turnover $(t-1)$	0.106	-0.039	-0.278*	-0.340**	-0.075	-0.049
	[0.151]	[0.165]	[0.142]	[0.138]	[0.131]	[0.141]
Log N Stocks $(t-1)$	0.174**	0.092	-0.007	0.038	-0.027	-0.011
	[0.069]	[0.088]	[0.067]	[0.063]	[0.066]	[0.069]
Flow $(t-1)$	-0.005	-0.005	0.007*	0.013***	0.015***	0.016***
	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]	[0.005]
Avg. Manager Exp. $(t-1)$	0.013	0.017	0.018	-0.004	-0.005	-0.008
	[0.045]	[0.054]	[0.047]	[0.041]	[0.042]	[0.042]
β_{MOM} $(t-1)$	-3.492*	-3.508*	0.665	1.106	1.068	2.444
	[2.023]	[1.806]	[1.838]	[1.555]	[1.677]	[1.739]
Disposition $(t-1)$	0.162	0.392	0.324	0.156	0.217	0.215
	[0.446]	[0.473]	[0.417]	[0.349]	[0.357]	[0.378]
Time × Style FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67634	67634	67629	67629	67629	61258
Adj. R-squared	0.8788	0.1012	0.435	0.2235	0.2255	0.2375

Table 5. **Field Evidence – Timing Solo-Managed versus Team-Managed Funds**: This table shows robustness of the main results in Table 3 with respect to the timing of solo-managed versus team-managed funds. In particular, we repeat the OLS regressions of Table 3 on the transmission of overreaction from individual team members to teams, but we re-estimate team-level overreaction over a sample period that is constrained to be closer to the sample period over which we estimate individual behavior. Columns (1) and (2) confirm that overreaction is attenuated in teams when we restrict the measurement of teams' overreaction to the period when at least one manager is also managing a fund individually. Columns (3) and (4) show that our results hold when we restrict the measurement of teams' overreaction to the period starting one year before and ending one year after the period during which we observe the counterfactual team. Columns (5) and (6) show the results hold when we begin (end) the estimation of the team-level behavior exactly when the estimation of the counterfactual team begins (ends). The regression is estimated with the same specification and the same controls as Table 3. Standard errors are in brackets. Significance: ***99%, **95%, *90%.

	Overlap Team & Single Manager		**Overlap Team & CF -1/+1 year**		**Overlap Team & CF Exact**	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}_j^{CF}$	0.9464***	0.8937***	0.7010***	0.6345***	0.6810***	0.5936***
	[0.1253]	[0.1254]	[0.1553]	[0.1561]	[0.1529]	[0.1561]
$\hat{\beta}_j^{CF} \times D_j^E$	-0.7040***	-0.6275***	-0.4317**	-0.3513*	-0.4657**	-0.3659*
	[0.1850]	[0.1790]	[0.2036]	[0.2023]	[0.2003]	[0.2002]
D_j^E	0.0676	0.0181	0.1242*	0.0921	0.1283*	0.1031
	[0.0720]	[0.0714]	[0.0723]	[0.0734]	[0.0716]	[0.0729]
Style fixed effects	No	Yes	No	Yes	No	Yes
Team controls	No	Yes	No	Yes	No	Yes
Observations	280	279	263	263	249	249
Adj. R-squared	0.2322	0.2774	0.153	0.2027	0.1497	0.1954
Hypothesis testing:						
p-value $H_0 : \delta_0 = 1$	0.669	0.397	0.055	0.038	0.020	0.010
p-value $H_0 : \delta_0 + \delta_1 = 0$	0.076	0.043	0.042	0.097	0.030	0.077
p-value $H_0 : \delta_0 + \delta_1 = 1$	0.000	0.000	0.000	0.000	0.000	0.000

Table 6. **Field Evidence – Compensation Structures**: This table compares the compensation structure of the fund managers in our sample when they manage individually with the compensation structures these managers face when managing as part of a team. The contractual incentives are measured using hand-collected data from the statement of additional information (SAI) that mutual funds file with the SEC. We summarize these incentives using four dummy variables tracking whether managers have (i) a fixed compensation; (ii) compensation based on the performance of the fund; (iii) compensation based on the AUM of the fund; and (iv) share ownership in their own funds. For more details on the data, see Section 2.7.2. We report the average β_j^X and compensation of the team, their counterfactual, and the difference between the counterfactual and the team (CF - team). Furthermore, we report the results for all teams combined (Panel A), for contrarian teams (Panel B), and for extrapolative teams (Panel C). A team consists mainly of contrarians if $\hat{\beta}_j^{CF} \leq 0$ and of extrapolators if $\hat{\beta}_j^{CF} > 0$.

Panel A: All teams

	β_j^X		**Fixed pay**		**Performance pay**		**AUM pay**		**Ownership**	
	CF	Team	CF	Team	CF	Team	CF	Team	CF	Team
Mean	0.015	-0.040	0.036	0.035	0.830	0.860	0.189	0.179	0.682	0.576
s.e.	0.037	0.042	0.014	0.015	0.029	0.029	0.031	0.032	0.032	0.038
CF - Team		0.054		0.001		-0.031		0.010		0.105
t-stat		1.232		0.127		-1.640		0.433		3.046
Obs.		128		125		124		124		119

Panel B: Contrarian teams

	β_j^X		**Fixed pay**		**Performance pay**		**AUM pay**		**Ownership**	
	CF	Team	CF	Team	CF	Team	CF	Team	CF	Team
Mean	-0.216	-0.175	0.029	0.032	0.815	0.857	0.209	0.165	0.714	0.608
s.e.	0.027	0.063	0.017	0.019	0.040	0.039	0.047	0.043	0.044	0.053
CF - Team		-0.040		-0.003		-0.042		0.044		0.106
t-stat		-0.782		-0.251		-1.273		1.101		2.133
Obs.		63		63		63		63		61

Panel C: Extrapolative teams

	β_j^X		**Fixed pay**		**Performance pay**		**AUM pay**		**Ownership**	
	CF	Team	CF	Team	CF	Team	CF	Team	CF	Team
Mean	0.238	0.092	0.044	0.038	0.842	0.861	0.172	0.197	0.642	0.535
s.e.	0.055	0.051	0.024	0.024	0.043	0.043	0.042	0.049	0.046	0.056
CF - Team		0.146		0.005		-0.019		-0.025		0.107
t-stat		2.112		0.293		-1.137		-1.077		2.162
Obs.		65		61		60		60		57

Table 7. **Experiment – Balance Tests of Demographic Variables**: This table presents means of key variables for participants in the Individual treatment (first column) and Group treatment (second column). The third column presents the difference between the means. College is a dummy equal to one if the participant holds a college, bachelor, master, or doctoral degree. Income bracket is a 1-to-13 indicator measuring the income range of the participant's household (where 1 stands for "less than $10,000" and 13 stands for "more than $200,000"). Employed is a dummy equal to one if the participant has either a full-time or part-time job. People in household is the number of individuals living in the participant's household. Single is a dummy equal to one if the participant is single, separated, divorced, widowed, or does not answer the question. Financial sophistication is a 0-to-3 measure of financial sophistication based on answers to financial literacy questions. CRT score is a 0-to-4 measure of cognitive skills based on answers to four cognitive reflection questions. Financial experience is a 0-to-4 measure of financial experience based on answers to financial experience questions. Liberal is a 1-to-5 measure of political opinions, ranging from 1 for "very conservative" to 5 for "very democrat." Risk tolerance, trust, and patience are measured on a 1-to-7 scale. $t-$stats are reported in brackets.

Variable	**Individual**	**Group**	**Group – Individual**
Male	0.52	0.51	–0.01
	[16.36]	[22.49]	[–0.24]
Age	37.57	38.38	0.80
	[47.83]	[70.44]	[0.84]
College	0.76	0.70	–0.06
	[27.82]	[33.82]	[–1.64]
Income Bracket	7.95	7.61	–0.34
	[35.79]	[45.57]	[–1.24]
Employed	0.77	0.78	0.01
	[28.45]	[41.04]	[0.38]
People in Household	2.17	2.30	0.13
	[22.47]	[32.86]	[1.08]
Single	0.40	0.48	0.08
	[12.81]	[20.06]	[2.06]
Financial Sophistication	2.57	2.45	–0.12
	[54.86]	[64.19]	[–2.04]
CRT Score	2.50	2.48	–0.02
	[37.38]	[49.74]	[–0.21]
Financial Experience	2.29	2.21	–0.08
	[25.24]	[33.63]	[–0.68]
Liberal	3.38	3.30	–0.09
	[50.16]	[60.50]	[–1.00]
Risk Tolerance	4.31	4.38	0.07
	[42.97]	[62.38]	[0.57]
Trust	3.61	3.85	0.24
	[37.25]	[51.09]	[1.97]
Patience	4.98	5.03	0.05
	[54.42]	[72.26]	[0.40]

Table 8. **Experiment – Baseline Results**: In this table we present results from regressions of the overreaction coefficient $\tilde{\beta}$ of participants on a dummy variable equal to one for participants in the Group treatment and zero otherwise. The dependent variable is the overreaction coefficient estimated from Equation (7). In the two bottom rows, we report summary statistics for the dependent variable for participants in the Individual treatment. Column 2 includes demographic controls, namely ethnicity dummies; 1-to-7 measures of trust in others, patience, and risk tolerance; age; a college graduate dummy; an employed dummy; a 1-to-6 measure of political opinions, ranging from 1 for very conservative to 6 for very liberal; Biden or Trump dummies for the favorite presidential candidate; a single dummy; the number of people living in the household. In Column 3 we further include financial sophistication controls, namely dummies equal to one if the participant holds any stocks, has traded in stocks in the previous year, or holds shares in a mutual fund; a 0-to-4 score assigned on the basis of a cognitive reflection test; a 0-to-3 score assigned on the basis of a financial literacy test. In Column 4, we also include US region and income category fixed effects. Standard errors are in brackets and are clustered at the team level (which corresponds to the single participant in the case of the Individual treatment). Significance: ***99%, **95%, *90%.

	(1)	(2)	(3)	(4)
Group	-0.092***	-0.101***	-0.101***	-0.099***
	[0.035]	[0.035]	[0.035]	[0.035]
Controls:				
Demographics	No	Yes	Yes	Yes
Fin. Sophistication	No	No	Yes	Yes
Region and Income FE	No	No	No	Yes
Observations	704	704	704	703
R-squared	0.014	0.050	0.052	0.086
Avg. $\tilde{\beta}$ Individual	0.311	0.311	0.311	0.311
Std. Dev. $\tilde{\beta}$ Individual	0.424	0.424	0.424	0.424

Table 9. **Experiment – Decomposition**: In this table we regress the overreaction coefficient $\tilde{\beta}$ on dummy variables that equal one if participants are assigned to the Group (G), Internal Reflection (IR), or Self Selection (SS) treatments, and zero otherwise, with the overreaction coefficient $\tilde{\beta}$ estimated from Equation (7). We also report the differences $\widehat{\gamma_G} - \widehat{\gamma_{IR}}$, $\widehat{\gamma_{SS}} - \widehat{\gamma_{IR}}$, and $\widehat{\gamma_G} - \widehat{\gamma_{SS}}$, with associated standard errors, where $\widehat{\gamma_j}$ is the coefficient with treatment dummy $j = G, IR, SS$. In the two bottom rows, we report summary statistics for the dependent variable for participants in the Individual treatment. Column 2 includes demographics controls, namely ethnicity dummies; 1-to-7 measures of trust in others, patience, and risk tolerance; age; a college graduate dummy; an employed dummy; a 1-to-6 measure of political opinions, ranging from 1 for very conservative to 6 for very liberal; Biden or Trump dummies for the favorite presidential candidate; a single dummy; the number of people living in the household. In Column 3 we further include financial sophistication controls, namely dummies equal to one if the participant holds any stocks, has traded in stocks in the previous year, or holds shares in a mutual fund; a 0-to-4 score assigned on the basis of a cognitive reflection test; a 0-to-3 score assigned on the basis of a financial literacy test. In Column 4, we also include US region and income category fixed effects. Standard errors are in brackets and are clustered at the team level (which corresponds to the single participant in the case of the Individual treatment). Significance: ***99%, **95%, *90%.

	(1)	(2)	(3)	(4)
Group	-0.092***	-0.098***	-0.098***	-0.098***
	[0.035]	[0.034]	[0.035]	[0.035]
Internal Reflection	-0.006	-0.004	-0.005	-0.001
	[0.033]	[0.033]	[0.034]	[0.034]
Self-Selection	-0.090**	-0.071*	-0.073*	-0.068*
	[0.039]	[0.039]	[0.039]	[0.039]
Controls:				
Demographics	No	Yes	Yes	Yes
Fin. Sophistication	No	No	Yes	Yes
Region and Income FE	No	No	No	Yes
Observations	1,512	1,512	1,512	1,512
R-squared	0.013	0.039	0.041	0.057
Avg. $\tilde{\beta}$ Individual	0.311	0.311	0.311	0.311
Std. Dev. $\tilde{\beta}$ Individual	0.424	0.424	0.424	0.424
$\widehat{\gamma_G} - \widehat{\gamma_{IR}}$	-0.086***	-0.095***	-0.093***	-0.098***
$se(\widehat{\gamma_G} - \widehat{\gamma_{IR}})$	(0.029)	(0.029)	(0.029)	(0.029)
$\widehat{\gamma_{SS}} - \widehat{\gamma_{IR}}$	-0.084**	-0.067**	-0.069**	-0.067**
$se(\widehat{\gamma_{SS}} - \widehat{\gamma_{IR}})$	(0.035)	(0.034)	(0.034)	(0.034)
$\widehat{\gamma_G} - \widehat{\gamma_{SS}}$	-0.002	-0.027	-0.025	-0.031
$se(\widehat{\gamma_G} - \widehat{\gamma_{SS}})$	(0.036)	(0.035)	(0.035)	(0.035)

Table 10. **Experiment – Additional Outcomes**: In this table we present results from regressions of several outcomes on a dummy variable that equals one if participants are assigned to the Group treatment and zero otherwise. The dependent variables are: the absolute value of the overreaction coefficient estimated from Equation (7) (Column 1), the mean squared forecast error (Column 2), the mean absolute forecast error (Column 3), the bonus earned based on the forecasting accuracy (Column 4). In the two bottom rows, we report summary statistics for the dependent variable for participants in the Individual treatment. All the regressions include demographic controls (namely ethnicity dummies; 1-to-7 measures of trust in others, patience, and risk tolerance; age; a college graduate dummy; an employed dummy; a 1-to-6 measure of political opinions, ranging from 1 for very conservative to 6 for very liberal; Biden or Trump dummies for the favorite presidential candidate; a single dummy; the number of people living in the household), financial sophistication controls (dummies equal to one if the participant holds any stocks, has traded in stocks in the previous year, or holds shares in a mutual fund; a 0-to-4 score assigned on the basis of a cognitive reflection test; a 0-to-3 score assigned on the basis of a financial literacy test), and region and income category fixed effects. Standard errors are in brackets and are clustered at the team level (which corresponds to the single participant in the case of the Individual treatment). Significance: ***99%, **95%, *90%.

	$\lvert\tilde{\beta}\rvert$	**MSE**	**MAE**	**Bonus**
	(1)	(2)	(3)	(4)
Group	-0.101***	-157.722***	-2.015***	0.059**
	[0.027]	[30.394]	[0.408]	[0.023]
Controls:				
Demographics	Yes	Yes	Yes	Yes
Fin. Sophistication	Yes	Yes	Yes	Yes
Region and Income FE	Yes	Yes	Yes	Yes
Observations	703	703	703	703
R-squared	0.088	0.111	0.099	0.063
Avg. Dep. Var. Individual	0.413	829.703	22.757	0.910
Std. Dev. Dep. Var. Individual	0.326	396.046	4.972	0.245

Internet Appendix of "Do Teams Alleviate or Exacerbate Overreaction in Beliefs?"

Ricardo Barahona Stefano Cassella Kristy A. E. Jansen Vincenzo Pezone

Internet Appendix Table of Contents

Appendix IA1. Field Evidence: Dataset Construction IA.1
Appendix IA2. Field Evidence: Measuring Overreaction IA.5
Appendix IA3. Field Evidence: Robustness IA.8
Appendix IA4. Field Evidence: IV Simulations – Transmission Overreaction to Teams IA.20
Appendix IA5. Field Evidence: Cost of Overreaction – Back-of-the-Envelope Analysis IA.25
Appendix IA6. Experimental Evidence: Details on the Design IA.29
Appendix IA7. Experimental Evidence: Additional Results IA.41

IA1. Field Evidence: Dataset Construction

In this appendix, we explain how we merged the CRSP, Morningstar, and Thomson Reuters databases, based on work by (Berk and van Binsbergen, 2015; Pástor, Stambaugh, and Taylor, 2015; Kim, 2020). We start with the cleaning of the raw CRSP database and Morningstar database separately, followed by a detailed explanation of the merge between the two. We then explain how we match this merged database to mutual fund holdings data from s12 filings, obtained from CRSP and Thomson Reuters.

IA1.1. Cleaning Raw CRSP Database

We download the monthly returns (mret), size per share class (mtna), tickers (ticker), and cusip numbers (ncusip) from the raw CRSP database over the period 1979M1-2019M9. We first delete observations for which total net asset values or returns are missing within a given month. The number of observations then equals 6,986,661 and there are 66,453 unique fund share classes (*CRSP fundno*'s).

We forward- and backward-fill the tickers within each fund share class. We then perform the following four checks:

1. *We check if a CRSP fundno has multiple tickers in a given month.* There are no such cases.

2. *We check if a CRSP fundno has multiple tickers over the entire sample period.* There are 2,970 *CRSP fundno*'s with time-varying tickers. We use the latest ticker for each *CRSP fundno* available, following (Pástor et al., 2015).

3. *We check if a ticker has multiple CRSP fundno's in a given month.* There are 24,699 combinations of ticker and month that correspond to more than one *CRSP fundno*. As in (Pástor et al., 2015), we replace these cases with a missing value.

4. *We check if a ticker has multiple CRSP fundno's over the sample.* There are 3,834 tickers with multiple *CRSP fundno*'s. These cases are automatically taken care of in the merge.

We follow the exact same procedure for cusip numbers. We list here the number of cases to which 1-4 apply in the case of cusip numbers:

1. There are no such cases.

2. There are 12,131 *CRSP fundno*'s with time-varying cusips.

3. There are 8,185 combinations of cusip and month that correspond to more than one *CRSP fundno*.

4. There are 247 cusips with multiple *CRSP fundno*'s.

IA1.2. Cleaning Raw Morningstar Database

We select the domestic equity Morningstar funds, excluding index funds, and download the monthly returns, size per share class, tickers, and cusip numbers over the period 1980M1-2019M12. We first delete observations for which total net asset values or returns are missing within a given month. The number of observations then equals 1,838,776 and there are 15,947 unique fund share classes (*SecId*).

We again apply the four checks as we did for the CRSP database for both tickers and cusips. We summarize the number of cases to which 1-4 apply here:

1. There are no such cases for both ticker and cusip.

2. There is no *SecId* that has time-varying tickers or cusips. A *SecId* either never has a ticker (cusip) over the entire sample period or a *SecId* has the same ticker (cusip) over the entire sample period.

3. There are 547 combinations of ticker and month that correspond to more than one *SecId* for ticker and 650 for cusip and month.

4. There are 23 tickers with multiple *SecId*s for ticker and 3 for cusip.

IA1.3. Matching CRSP and Morningstar Databases

The CRSP database is our master file and we merge this database to Morningstar using first a match based on ticker and then a match based on cusip (the results of the merge are exactly the same if we first merge based on cusip and then on ticker). In order to make sure that missing values are never matched, we replace the ticker (cusip) with the *CRSP fundno* number in CRSP and with the *SecId* in Morningstar (Berk and van Binsbergen, 2015).

After we merge based on cusip and ticker, we also merge CRSP and Morningstar based on an exact match between year, month, monthly return, and monthly total net asset value. After that, we perform two near merges:

1. Exact match based on year, month, and total net asset value and a difference in monthly returns in the CRSP versus the Morningstar database that is at most two basis points.

2. Exact match based on year, month, and monthly return and a difference in total net asset value that is at most 20,000 USD.

We then correct for potential errors in the merge as some tickers and cusip numbers may be reused. We first check whether the same unique share class identifier from CRSP, *CRSP fundno*, consistently matches the Morningstar unique identifier for the history of that share class. Following (Berk and van Binsbergen, 2015), we drop all funds where the same share classes are matched less than 60% of the time. On the other hand, if a given share class is matched more than 60% of the time, we assume that this match is the correct match and change the observations that don't match accordingly.

We then use the Morningstar *FundId* to group funds that have multiple share classes and check whether we are able to match all of the share classes of a given fund. Following (Pástor et al., 2015), if we are not able to find a full match, we drop those observations.

Finally, we select CRSP fund share classes that are defined as domestic equity, but excluding the index funds. We only keep quarterly observations and the period we consider is 1980Q1-2018Q4. We end up with a final dataset of 436,984 observations. The merge matches 80% of the CRSP active US domestic equity universe.

At the end of this process we have a key that allows us to match any given unique CRSP share class number and portfolio number, *CRSP fundno* and *CRSP portno*, to a Morningstar portfolio level number, *FundId*. This key is then used to match funds to fund information from CRSP such as fund TNA, expense ratios, and returns, fund holdings from Thomson Reuters and CRSP, and fund managers from Morningstar.

IA1.4. Merging Mutual Fund Holdings Data

The next step in constructing our dataset is to match funds to its respective fund holdings. For this part we use two sources, the Thomson Reuters s12 Holdings file for mutual funds and the CRSP s12 Mutual Fund Holdings database. For the first part of the sample, from 1980 to 2008, we use Thomson Reuters s12 Holdings, and CRSP thereafter.

Merging is then straightforward. We are able to match the CRSP Mutual Fund Holdings to our master data file using their unique fund identifiers which are present in both files, and we use the MFLINKS dataset from Wharton Research Data Services to match Thomson Reuters data to our master file that links Thomson Reuters fund identifiers to CRSP identifiers. We then collapse stock holdings every quarter at the portfolio level by adding all shares of a given stock for the fund's Morningstar FundId.

After matching the holdings, using stock CUSIP numbers from CRSP, we link mutual funds' holdings to the stock-level information (prices, returns, book-to-market, profitability, investments) contained in the merged CRSP-COMPUSTAT database. We consider the universe of stocks with codes 10 and 11 that trade on the NYSE, NASDAQ and AMEX exchanges, and we exclude stocks trading below $5. Finally, we link each mutual fund to their respective managers. This results in a dataset that contains manager-fund-stock-quarter holdings data.

In Table IA1 we report summary statistics for all the funds collected and in Table IA2 we report additional summary statistics on the sample of managers and teams used for the within-subject design analysis.

IA2. Field Evidence: Measuring Overreaction

IA2.1. Control Variables for Measuring Overreaction

To measure overreaction both at the manager and at the fund level, we control for several stock characteristics that have been associated with either pricing anomalies or risk premia. To the extent that rational managers aim to gain exposure to these characteristics, and these correlate with past (weighted) yearly stock returns, including such characteristics as controls allows us to more accurately identify overreaction.

We first include size and book to market ratios as controls in Equation (1). Small and value firms, which are respectively measured by a small firm size and high book to market ratio, have been widely documented as having historically high abnormal returns (Fama and French, 1993). We follow up by also including asset growth and operating profitability as measured in Cooper, Gulen, and Schill (2008) and Fama and French (2015), respectively. Regarding these characteristics, Cooper et al. (2008) document a negative relationship between past asset growth and future stock returns while Novy-Marx (2013) documents a positive relationship between firm profitability and future stock returns. We also include stock volatility over the past 12 months as highly volatile stocks have been associated with low expected returns (Ang, Hodrick, Xing, and Zhang, 2006). Finally, we include a stock's past one month return to control for short-term reversals, because stocks with low past one-month returns have high returns in the subsequent period (Lehmann, 1990).

An additional reason to include these characteristics is the documented preferences of institutional investors and their demand for certain characteristics. Gompers and Metrick (2001) document how several characteristics are determinants of stock institutional ownership, finding that institutions have a particular demand for larger firms. More recently, Koijen and Yogo (2019) document how different types of institutions vary in their demand for stock characteristics such as size, book to market, profitability, and investment (asset growth).

We furthermore control for flow-induced trading. Lou (2012) shows that flows in the presence of liquidity costs lead to disproportionately buying and selling of certain stocks over others. In particular, funds expand their stock holdings set rather than increasing them proportionally after inflows. Moreover, he shows that funds buy (sell) less of more illiquid stocks after inflows (outflows). We therefore follow Lou (2012) and include two measures of liquidity costs as controls (i) the percentage of all shares outstanding of a stock that is held by the fund and (ii) the effective bid-ask spread of a stock. We obtain the bid-ask spread for each stock from the Open Source Asset Pricing website (Chen and Zimmermann, 2022). We

also interact both measures of liquidity costs with contemporaneous fund flows. Unlike Lou (2012), we do not include fund flows and fund-level liquidity costs as separate independent variables, because they are accounted for by the time fixed effects in our regressions.

IA2.2. Investment Universe

In each quarter, the stocks that appear in the cross-section of trades in Equation (1) is based on a definition of the relevant investment universe of fund j that is close to Koijen and Yogo (2019). They propose that a fund's investment universe is made of stocks that the fund has held at any point in time in the previous 11 quarters plus the current quarter. We follow their approach, but also include in the set those stocks that the fund will start owning at some point in the subsequent 11 quarters. This definition of the investment universe accounts for the fact that a stock can enter the investment universe before fund managers' first purchase of that stock, and the decision not to yet purchase that stock at time t also contains useful information to measure fund managers' overreaction.

IA2.3. Alternative Overreaction Measures

We now show that our results are robust to alternative measurements of overreaction. We consider three such alternatives: (i) to further ensure that we are not picking up funds' momentum trading, we exclude momentum stocks from the sample when measuring overreaction;[1] (ii) to show that our way to measure funds' trading behavior does not drive our results, we measure overreaction in regressions where funds' trading in a stock is measured as the active change in portfolio weights by the fund, i.e., the change in weights that is obtained after accounting for the mechanical change in weights that occurs due to cross-sectional heterogeneity in stock returns;[2] (iii) to alleviate concerns that the use of decaying weights drives our results, we re-estimate our overreaction metric using past-year stock returns (i.e., we use the same weights for returns in the past four quarters).

[1] Following the literature on momentum, momentum stocks are those stocks that are in the top or bottom deciles of the distribution of cumulative stock returns at time $t-1$ (e.g., Jegadeesh and Titman, 1993; Fama and French, 1996).

[2] A simple quarter-on-quarter change in portfolio weights is not appropriate, because portfolio weights can change even in the absence of active managerial decisions depending on how a stock performs relative to the other stocks the fund owns. Therefore, to capture the component of funds' trading that is due to active buying or selling of stocks by fund managers, we define weight changes as:

$$\Delta w_{s,j,t+1} = w_{s,j,t+1} - \frac{(1 + r_{s,t\to t+1})}{(1 + r^{P}_{j,t\to t+1})} w_{s,j,t}, \tag{IA12}$$

where $r^{P}_{j,t\to t+1}$ is the total portfolio return for fund j in quarter $(t, t+1]$, and $r_{s,t\to t+1}$ is the stock-return over the same quarter.

Table IA3 replicates our main results using the alternative overreaction metrics. Panel A studies the relation between overreaction and managerial performance. Panel B asks whether overreaction is reduced in teams. Across all of the six distinct tests conducted in Panel A, the relation between overreaction and performance appears negative and statistically significant. More importantly, Panel B shows that the evidence that teams help alleviate overreaction is confirmed when using the alternative overreaction metrics presented in this section.

Likewise, to show that our results do not rely on the specific form of the investment universe that we choose for our main overreaction metric, we also provide robustness of the investment universe that is used to estimate β_j^X. In particular, in Table IA4 of the Appendix, we show that our results remain unaltered when we assume that the investment universe only incorporates those stocks that the investors start to hold in the following year, instead of the following 11 quarters. Additionally, we show that our results are robust to a specification where we abstract from defining the investment universe and only include actual managers' trades.

Finally, in our main specification of fund trades, overreaction is measured while controlling for flow-induced-trading. However, the analysis in Lou (2012) shows that the trading that is induced by inflows and outflows can be different. Therefore, in Table IA4, we show that our results are robust to asymmetric flow-induced trading. Specifically, in Equation (1) we include a dummy variable that indicates whether the fund received inflows from quarter t to $t+1$, $D_{j,t+1}^{inflows}$, interacted with the variables in $F_{s,j,t}$. The dummy $D_{j,t+1}^{inflows}$ equals one if $flow_{j,t+1} > 0$ and zero otherwise. Using this approach we re-estimate our overreaction metric and perform all our analyses again. The results stay the same.

IA3. Field Evidence: Robustness

This appendix serves three purposes: i) it describes several controls included in the analysis to distinguish overreaction from other measures that rely on past returns; ii) we provide additional tests that corroborate our interpretation of overreaction as a bias and iii); we rule out alternative explanations for the reduction in overreaction observed in teams.

IA3.1. Overreaction versus Momentum and the Disposition Effect

In all of our tests, we distinguish overreaction or extrapolation from the disposition effect by controlling for the disposition effect explicitly. We follow Odean (1998) and Cici (2012) and estimate the disposition effect for every fund in our sample in a given quarter. For each fund j in a given quarter t, we start by estimating a fund's proportion of realized gains (PRG) and proportion of realized losses (PRL) from its holdings:

$$PRG_{j,t} = \frac{RG_{j,t}}{RG_{j,t} + UNRG_{j,t}}, \quad PRL_{j,t} = \frac{RL_{j,t}}{RL_{j,t} + UNRL_{j,t}}, \tag{IA13}$$

where $RG_{j,t}$ is the number of realized gains, $UNRG_{j,t}$ the number of unrealized gains, $RL_{j,t}$ the number of realized losses, and $UNRL_{j,t}$ is the number of unrealized losses. The disposition effect is then calculated as the difference between these two proportions: $DISP_{j,t} = PRG_{j,t} - PRL_{j,t}$.

To define which positions of a fund are at a loss or gain, we need to define a cost basis for the purchase price of the stocks held by the fund. We follow the Odean (1998) approach by using the average purchase price weighted by the number of shares in each given purchase and corrected for stock splits. Given that we only observe share prices at the end of each quarter, we assume purchases and sales occur at the end of the quarter.

Similarly, to distinguish overreaction from momentum trading explicitly, our regressions control for a fund's momentum trading strategy (Jegadeesh and Titman, 1993). To this end, we estimate a standard 4-factor Fama-French Carhart (Carhart, 1997) model for each fund over a 5-year rolling window using monthly returns. We use the fund returns' loading on the momentum factor from this model, β_{MOM}, as a control in our tests.

To underscore the empirical differences between our measure of overreaction β_j^X and the disposition effect and momentum, Panel B of Table IA1 reports the correlations between them. Our overreaction metric has a small correlation with both the disposition effect (-13%) and momentum trading (15%) and, as we show in Table IA5, momentum and the disposition effect capture about 3% of the cross-sectional variation in overreaction.[3] These

[3]The low correlations between the overreaction metrics and other aspects of the fund's trading behavior are not driven by outliers, because we winsorize all the variables at the 0.5th and 99.5th percentile.

low correlations are to be expected. Overreaction and the disposition effect only have a small correlation in Liao et al. (2022), who examine the relation between the two in a large sample of retail investors. Similarly, the weak relation between overreaction and momentum is expected, given the conceptual and empirical differences between the two forms of trading behavior. In particular: (i) unlike momentum, which prescribes how to trade stocks that place in the top and the bottom of the cross-sectional return distribution, our measure of overreaction concerns a fund's trading behavior over the entire investment universe of the fund; (ii) whereas momentum trading is based on past-year cumulative returns, in our regressions we measure overreaction in the spirit of Greenwood and Shleifer (2014) as fund managers' response to a weighted sum of past stock returns with exponentially declining weights; (iii) in keeping with Barberis (2018)'s view of overreaction as a form of delayed overreaction, our overreaction metric captures investors' behavior in quarter $t+1$ based on returns realized a quarter earlier, while momentum trading suggests what stocks to buy at $t+1$ based on stock returns at that point in time.

Moreover, to further strengthen the interpretation of β_j^X as a gauge for fund managers' irrational behavior rather than momentum trading, we repeated our analysis in Table IA3 by explicitly removing momentum stocks, that is, the stocks that are in the bottom and top deciles of momentum sorts, from the cross-section of position changes that is used to estimate the overreaction metric. As we show there, the results of our analysis remain unaltered.

Finally, we conduct a spanning test to examine the ability of momentum and the Fama-French 5 factors to explain a long-short strategy that goes long the highest 20% extrapolative funds and short the lowest 20% extrapolative funds. In order to address the discrepancy between academic momentum and practical momentum, we also include as an explanatory variable the return on a long-short portfolio that goes long (short) in the funds with the highest (lowest) momentum factor exposures. Table IA6 shows that the momentum factor explains only a small fraction of the variation in returns to the overreaction strategy. Moreover, across all specifications, the factors do not span the long-short strategy, as the alpha is consistently negative and statistically significant across all specifications.

IA3.2. Overreaction and Mutual Fund Performance

IA3.2.1. Performance Metrics and Control Variables

One important part of the analysis concerns the relationship between a fund manager's overreaction and the fund's investment performance. To measure a mutual fund's performance, we consider three distinct metrics: (i) fund returns, (ii) whether a fund is a star fund in a given quarter, and (iii) fund flows. In the paper we cover (i), here we'll show additional results for (ii) and (iii).

Regarding the fund returns we focus on fund raw returns and benchmark adjusted returns (defined as the fund return minus the return of the primary prospectus benchmark) as well as three measures of risk-adjusted returns: CAPM, Fama-French 3-factor (Fama and French, 1993), and Fama-French 5-factor alphas (Fama and French, 2015). We consider returns gross of fees as these are the returns that are relevant for manager's compensation. Because gross returns are not directly observable in CRSP data, we follow past work (e.g. Fama and French, 2010) and add the most recently available expense ratio to fund net returns. To estimate a fund's risk-adjusted return at time t, we use a rolling window of the previous five years of monthly returns to estimate the fund's factor exposures. We then use these exposures to estimate the fund's risk-adjusted returns over the following year by subtracting the portion of fund returns that are the result of factor exposures from the fund's returns over that period. Using these factor exposures, we also estimate fund manager risk as the standard deviation of monthly alphas over the following 12-month period.

As an alternative measure of performance, we also use an indicator variable for whether a fund is considered a star fund in a given year. A star fund is one that ranks in the top 10% of yearly returns in its respective Morningstar category. This measure is particularly relevant for our analysis, because managers' compensation can be linked directly to the achievement of this star status (Ma et al., 2019).

Moreover, we also analyze how overreaction is related to mutual fund flows. Flows for fund j are defined as the dollar inflows or outflows in a year (DF_{jt}), as a percentage of yearly lagged fund size, ($TNA_{j,t-1}$):

$$flow_{j,t} = \frac{DF_{j,t}}{TNA_{j,t-1}}. \tag{IA14}$$

When estimating how overreaction affects mutual fund flows, we also control for the fund's CAPM alpha over the past year, because Barber, Huang, and Odean (2016) and Berk and Van Binsbergen (2016) document that CAPM alphas predict mutual fund flows.

When studying how mutual fund manager performance depends on overreaction, it is important to control for other variables that are related to both factors to ensure our results are not driven by omitted variables. To this end, we start by controlling for fund characteristics, such as expense ratios, fund size, and past fund flows. Expense ratios can be related to manager skill as motivated in Berk and Green (2004), because self-interested, skilled managers can raise fees to capture the benefits of their skill. The positive relationship between fund size and skill is well documented in Berk and van Binsbergen (2015) and recent empirical research has also established that funds have decreasing returns to scale (Pástor et al., 2015; McLemore, 2019). Furthermore, funds with large outflows can experience high trading costs due to fund liquidity constraints, which can affect fund performance (Coval

and Stafford, 2007).

We also control for manager characteristics, such as managers' experience and trading behavior, which can be related to the tendency to extrapolate returns and have been extensively documented as having a relationship with fund performance (Golec, 1996; Chevalier and Ellison, 1999).

To the extent that experienced managers are less likely to suffer from behavioral biases, manager experience accounts for this effect. We also control for trading behavior by including mutual fund turnover ratios from CRSP and the number of stocks held in the mutual fund portfolio, because high trading activity may impact performance as documented by Wermers (2000) and Cremers and Petajisto (2009). The age of the fund is typically negatively related to fund performance (e.g. Cremers and Petajisto, 2009), so we control for fund age since inception in our tests as well.

IA3.2.2. Additional Tests Overreaction and Fund Performance

In Table IA7, we confirm that the negative impact of overreaction on managerial financial outcomes extends to other determinants of managers' pay, such as fund flows (which determine fund size and fund revenues), and the achievement of top-fund status. In Columns 1 and 2, we find that extrapolative funds experience outflows relative to other funds. This result holds both when analyzing percentage flows and when using a ranking measure for fund flows, that takes a higher value for the funds that receive larger inflows. In Columns 3 and 4, we investigate whether overreaction hinders funds' ability to attain a top-fund status. Following Guercio and Tkac (2008), we measure top performance as a dummy variable equal to one if a fund's raw return gross of fees ranks in the top 10% (Column 3) or 5% (Column 4) of the performance distribution in a fund's Morningstar style classification over the following year. Once again, the evidence points to a negative relationship between overreaction and fund managers' outcomes, because overreaction predicts a lower probability of achieving top-fund status.

IA3.3. Beliefs versus Preferences

We now lay out the results of additional tests to rule out that our overreaction metric captures preferences rather than beliefs. Throughout the paper, we interpret the documented reduction of overreaction in teams as evidence that teams curb biases in beliefs. We adopt this interpretation because the most common interpretation in the literature is that overreaction arises from biases in beliefs (e.g. Barberis and Shleifer, 2003; Barberis et al., 2018; Liao et al., 2022). More recently, Jiang, Liu, Peng, and Yan (2022) link overreaction and the larger weight placed on recent returns to investor memory, further strengthening the

interpretation that biased cognition and biased recall of past data, rather than preferences, drive overreaction. However, our measure of overreaction is based on trades, as opposed to direct survey data on beliefs. So it is possible that our tests speak to how teams aggregate investor preferences rather than biases in beliefs.

While it is possible that preferences influence our tests, there are several considerations that make this alternative interpretation difficult. First, in a recent paper Liao et al. (2022) argue that overreaction entails a dependence of trading on past returns that is very specific to biases in beliefs, namely, that past returns enter trading decisions with a decaying weight. Following this insight, all of our main results use decaying weights, and we select the weights assigned to past returns based on models of extrapolation that Greenwood and Shleifer (2014) calibrate to direct survey data on investor beliefs. This empirical design choice helps to strengthen the interpretation that overreaction is belief-based and that our results reflect the reduction of biases in beliefs in teams.

To further strengthen the case that our evidence captures the reduction of biases in beliefs, we follow Da et al. (2021) and Liao et al. (2022), who argue that initial buys are more likely to be linked to beliefs rather than to preferences.[4] Therefore, we repeat our team regressions by relying only on first-purchases to measure fund manager's overreaction.[5] Table IA8 of the Appendix shows that the conclusion that overreaction is reduced in teams is robust to this alternative measurement.

We also note that to the extent conditional skewness correlates with past returns, preferences for skewness as in Barberis and Huang (2008) could explain investors' buying of stocks on the back of good returns. With respect to this explanation, Boyer, Mitton, and Vorkink (2010) run predictive regressions of future skewness on a set of lagged stock characteristics. They find that idiosyncratic volatility is the strongest predictor of future skewness in the cross-section. Since we control for idiosyncratic volatility when we estimate fund managers' overreaction, our measures of overreaction are unlikely to proxy for skewness preferences. Moreover, Boyer et al. (2010) also show that past returns do not predict skewness after controlling for volatility, making it even more difficult to explain extrapolative trades with

[4]Initial purchases also help address the concern that overreaction captures prospect-theory-based preferences with narrow framing as in Barberis and Huang (2001). In the model, good stock returns lower stock-specific loss aversion and increase the desire to purchase a stock, because of the cushioning role that the past stock returns provide. This cushioning role is only present if the investor previously held the stock. As a result, initial purchases, which by definition offer no cushioning, allow for addressing this preference-based explanation.

[5]Liao et al. (2022) use a non-parametric overreaction metric based on retail investors' initial buys. In our setting, this non-parametric approach would not allow to control for institutional factors such as flow-induced trading and investment mandates that are important when studying mutual funds. Our regression-based approach allows to rely only on initial buys, while at the same time controlling for relevant alternative factors.

skewness preferences.[6]

To further rule out that our overreaction metric captures preferences rather than beliefs, we perform additional tests. We argue that if our overreaction metric reflects managers' preferences, then the managers we identify as extrapolators likely trade-off lower expected payoffs (see Table 4) for other valued features of the payoff distribution. We test three possible trade-offs: i) a risk-return trade-off where extrapolators sacrifice performance but achieve less volatile compensation, ii) a skewness-preferences trade-off where lower performance on average is associated with a positively skewed performance distribution (e.g. Barberis and Huang, 2008), and iii) hedging motives, whereby extrapolators fare better than non-extrapolators in bad states of the economy (e.g. Campbell and Cochrane, 1999; Guvenen, Ozkan, and Song, 2014).

These three potential trade-offs can be investigated empirically by extending the scope of the tests presented earlier. Specifically, to test whether those who we identify as extrapolators achieve lower payoffs in exchange for lower compensation volatility, we study the relation between overreaction and the volatility of fund performance, measured as the standard deviation of monthly benchmark adjusted returns and FF5 alphas over the next 12 months. Since managers' compensation is tied to their performance, lower performance volatility translates into a lower volatility of managers' compensation. Similarly, when testing whether overreaction leads to extremely positively skewed payoffs, we ask whether overreaction helps funds rank among the very top of their Morningstar style category in the following year (i.e., top 5 funds or top 10 funds), as this placement coincides with extremely positive financial outcomes. Finally, to investigate whether overreaction generates hedging benefits, we study the performance of funds whose trades are extrapolative vis-a-vis other funds in bad states of nature, either identified by a negative return to the market or by the recessionary state of the economy in that year.

In Panel A of Table IA9, we test whether overreaction is associated with lower return volatility. We find no evidence that overreaction leads to significantly lower volatility of returns. In Panel B, we test whether overreaction places managers at the very top of the performance distribution, but we find no evidence that this is the case; rather, we find evidence that the opposite is true. In Panel C, we analyze the relationship between overreaction and fund performance, but zoom in on years in which the CRSP weighted-market index was negative. We find that overreaction does not lead to better performance during times of poor market returns. When repeating the analysis in Panel D for years corresponding to NBER recessions, we find insignificant effects of overreaction on performance. However,

[6] An investor with cognitive limitations may misinterpret the relation between past returns and skewness, and could mistakenly believe that past returns predict skewness. However, this biased belief about skewness aligns with our explanation of overreaction stemming from biases in beliefs.

when comparing the results in Panel D with the earlier results of Table 4, extrapolative funds achieve worse performance in recession years as they do on average. So, we conclude that overreaction does not lead to better hedging properties.

Overall, these results support the view that the negative relation between overreaction and performance stems from managers' biases in expectations formation, not preferences.

IA3.4. Additional Tests Learning and Experience

IA3.4.1. Entering as a Solo versus Team Asset Manager

We argue that the learning argument described in Section 2.7.2 is natural when fund managers first work individually and then join a team. On the other hand, if working in a team precedes working individually, learning would be achieved during the years of team management, and thus learning predicts that one should observe a lower tendency to extrapolate in single-managed funds than in team-managed funds. In our sample, 65 belong to the case in which solo precedes team management for all members of the team, whereas 54 belong to the case in which team precedes solo management.[7] We flag all teams that belong to the former case using a dummy D_j^{ES} and all the teams that belong to the latter case using a dummy D_j^{ET}. We then estimate the following regressions:

$$\begin{aligned}\hat{\beta}_j^{TM} &= \alpha + \delta_0\hat{\beta}_j^{CF} + \delta_1\hat{\beta}_j^{CF} \times D_j^E \times D_j^l + \delta_2\hat{\beta}_j^{CF} \times D_j^E + \delta_3\hat{\beta}_j^{CF} \times D_j^l \\ &+ \delta_4 D_j^E \times D_j^l + \delta_5 D_j^E + \delta_6 D_j^l + \delta_7 C_j + \epsilon_j, \qquad l = ES, ET. \end{aligned} \tag{IA15}$$

The coefficient δ_1 on the double interaction term is the main coefficient of interest. When $l = ES$, the coefficient captures whether the transmission of overreaction is different for teams whose managers start out as single managers (i.e., $D_j^{ES} = 1$) versus other teams. Learning predicts that δ_1 is negative, i.e., compared to other teams, teams in which all managers had the opportunity to learn from their individual experiences prior to joining the team should exhibit the smallest transmission of overreaction to teams. Conversely, the coefficient δ_1 when $l = ET$ tests whether the attenuation of overreaction is different for teams whose managers start out as part of a team. A learning story predicts that δ_1 is positive. To see why, suppose that a fund manager learns that overreaction is harmful while working in a team, and later moves to a solo-managed fund. Compared to other managers, these managers are likely to exhibit a weaker overreaction when they move on to work individually, thus rendering a positive coefficient δ_1.

Panel B of Table IA10 reports the results of the estimation when $l = ES$ and Panel

[7] For the remaining teams, some members first operate solo and then move on to teams, whereas the opposite occurs for the other members.

C when $l = ET$. In both panels, the triple interaction term is not statistically significant, indicating no evidence of differences in the attenuation of overreaction as predicted by a learning explanation. Although the lack of statistical significance could result from the small sample size, the findings overall do not align with a learning explanation. The coefficient δ_1 is positive in Panel C, whereas learning predicts a negative δ_1 in Panel C. Overall, these results do not support learning as a plausible explanation for our finding.

IA3.4.2. Reduction of Overreaction in Teams of Non-Learners

If learning outside of the team is the driver of our result, we expect no attenuation of overreaction in the teams composed of non-learner managers. To identify non-learners, we first estimate fund managers' overreaction separately for the first half (the early sample) and the second half (the late sample) of the sample period in which these managers managed a fund alone. We then select the non-learners as the managers who extrapolate in both subsamples.[8]

Figure IA1 presents scatter plots that relate team-level overreaction to team members' average overreaction at the individual level. The figure portrays this relation for all teams (Panel A), the teams whose members extrapolate on average (i.e., extrapolative teams, Panel B), and for the extrapolative teams that consist of the non-learners and their teams only (Panel C). Panels A and B present in graphical form the evidence on the attenuation of overreaction in teams that we present in Table 3. More importantly, a comparison between Panel B and C reveals that overreaction is attenuated both in teams of learners and in teams of non-learners in virtually the same way. In particular, as in the full sample, teams attenuate the overreaction of non-learners as much as they do for other managers. Overall, this result confirms the robustness of our finding to a learning story. As a matter of fact, the finding that teams alleviate biases even among managers that display little tendency to learn when operating alone suggests that teams may be more effective at curbing biases than individual experience.

IA3.5. Alternative Channels Team Effect

IA3.5.1. Alternative Construction of Counterfactuals

Our main statistical counterfactual measure in Table 3 assigns equal weight to each team member. As this assumes that all managers carry the same weight in a team's decision,

[8]To validate our definition of non-learners, we compare the overreaction metric of the early and late sample and find for this group an average value of 0.19 for the early sample and an average value of 0.24 for the late sample. A difference in means test reveals that these values are statistically indistinguishable from each other (t-stat $= -1.29$).

we also perform tests where we assume that some team members have greater influence in team decisions. Formally, we compute three alternative team counterfactuals as weighted averages of team members' sensitivity to past returns, where managers' weights are based on: (i) quarters of industry experience; (ii) the number of funds managed; (iii) the aggregate size of the funds managed prior to team formation. We present these results in Table IA11 and find that they remain unchanged relative to the main measure, showing that individual overreaction is attenuated in teams.

IA3.5.2. Manager Workload and Bounded Rationality

Prior work shows that individuals make decisions with limited cognitive resources, and that the use of heuristics and the appearance of biases (like overreaction) are more likely when these resources are depleted. Thus, the evidence that biases are reduced when managers operate in teams may result from a loosening of such constraints in team-managed funds compared to solo-managed funds.

A loosening of the bounded rationality constraints could partly result from the team itself, as the efficient division of labor minimizes the burden on each manager. If this is the case, it is indeed appropriate to attribute the reduction of cognitive biases to the team as an organizational arrangement. However, the loosening of the bounded rationality constraints could also take place for reasons that are not intrinsically rooted in teams, but rather stem from differences in workload between solo and team management. For instance, if the managers in our sample oversee larger and more complex portfolios at the time in which we measure their individual-level overreaction, but their overall workload is systematically lower at the time in which they operate as part of a team, the constraints of bounded rationality can be more binding in solo-managed funds as opposed to team-managed funds. If this is the case, the reduction of overreaction could also be achieved in single-managed funds if these managers were asked to manage less complex and smaller portfolios.

To investigate whether systematic differences in workload between solo-managed funds and team-managed funds explain our result, we approximate the workload of manager i or team j at a given point in time as the TNA that the manager or team oversees across all portfolios. We compute the workload of the team as the time-series average of TNA during the team's existence. For the counterfactual workload, we take the equally weighted time-series average of the team members' workload when they operate individually. In other words, we assume that a team divides its TNA equally among its managers. Finally, we compute the difference in workload when the managers operate in team j versus when they operate in the counterfactual team, which we define as $\Delta Workload_j$. The summary statistics in Panel A of Table IA12 highlight the importance of this robustness check. We observe that

managers have lighter workloads when managing in teams compared to when they manage individually. On average, managers in team-managed funds oversee \$0.71 billion less in TNA than when managing funds individually.

If workload differences explain our findings, the reduction in overreaction should be more pronounced when workload declines during team management. Thus, we estimate the following regression:

$$\begin{aligned}\hat{\beta}_j^{TM} &= \alpha + \delta_0\hat{\beta}_j^{CF} + \delta_1\hat{\beta}_j^{CF} \times D_j^E \times \Delta Workload_j + \delta_2\hat{\beta}_j^{CF} \times D_j^E + \delta_3\hat{\beta}_j^{CF} \times \Delta Workload_j \\ &+ \delta_4 D_j^E \times \Delta Workload_j + \delta_5 D_j^E + \delta_6 \Delta Workload_j + \delta_7 C_j + \epsilon_j. \end{aligned} \tag{IA16}$$

If the differential workload of managers drives our results, we expect coefficient δ_1 to be statistically significant and positive. In words, if a decrease in workload during team-management explains the attenuation of overreaction in teams, then overreaction ought to be reduced more in teams where $\Delta Workload_j$ is low, i.e. teams where such a reduction in workload is more pronounced. Panel C of Table IA12 shows that δ_1 is statistically indistinguishable from zero. We take this as evidence that managers' reduction in workload once they join the team does not drive our results.

As an alternative proxy for workload, we use the number of stocks that the manager oversees in any of the portfolios he manages, be that alone or in a team. We start by constructing a time series of the total number of stocks the manager has in their investment universe at any given quarter t.[9] To construct the time series, we compute the total number of stocks managers have in their investment universe in all the solo-managed funds they manage at that point in time, plus the proportional fraction of the investment universe of the team-managed funds these managers co-manage. When allocating stocks in the investment universe of a team to one of the team managers, we assume that a manager who holds a given stock in his solo-managed funds will also be overseeing investments in that stock when operating as part of a team. Finally, we carefully avoid double counting overlapping stocks across multiple funds a manager oversees simultaneously. For example, if a manager simultaneously manages a solo-managed fund and a team-managed fund, and if both funds hold an Apple stock, then the Apple stock would only count once towards the manager's workload. Formally, we define the workload as:

$$Workload_{i,t} = \sum_{j \in S_{i,t}} U_{j,t} + \sum_{j \in C_{i,t}} \max\left[\frac{U_{j,t}}{NumManagers_{j,t}} - Overlap_{i,j,t}, 0\right], \tag{IA17}$$

where $U_{j,t}$ is the number of stocks in the investment universe of fund j at time t, $S_{i,t}$ is the

[9] Here we use the same definition of the investment universe as the one we use when we estimate our overreaction metric, see Section 2.

set of solo-managed funds that manager i is in charge of at time t, $C_{i,t}$ is the set of funds that manager i co-manages at time t, $NumManagers_{j,t}$ is the total number of managers for fund j at time t, and $Overlap_{i,j,t}$ is the number of stocks from the team-managed fund j that overlap with the investment universe of all the solo-managed funds of manager i at time t. We include a max operator to ensure that additional fund assignments do not reduce a manager's workload.

We obtain the workload of the members of team j when operating in the team and when operating in the counterfactual team composed of solo-managed funds, by taking time-series averages of the workload metric for each manager during that team's existence and take the average across managers. Finally, we compute the workload difference between managers operating in team j and in the counterfactual team, which we define as $\Delta Workload_j$.

Table IA13 summarizes the results and shows that δ_1 is statistically indistinguishable from zero. Thus, using this alternative workload measure, we conclude that differences in workload between team-managed and solo-managed funds cannot explain our findings.

IA3.5.3. Style Migrations

An additional alternative interpretation of our findings is that the attenuation of overreaction is driven by team members systematically migrating between styles when transitioning from single to team management. If differences in overreaction exist between different styles, then a migration between two styles could generate evidence consistent with our findings. We address this alternative interpretation in three ways. First, we add style fixed effects of the teams to our regressions. Second, we add stock characteristics such as market-to-book and size to our estimation of managers' overreaction. Adding stock controls mitigates the impact of differences in overreaction across managers are due to funds' size-based and value-based style classifications. Third, we formally test whether our results of bias reduction in teams are due to style migrations by performing a similar regression as in Equation IA16, where we replace $\Delta Workload_j$ with D_j^{SM}, and D_j^{SM} is an indicator variable equal to one in the case of a style migration for team j. In order to construct this indicator variable, we identify the prevalent style of the team as well as the prevalent style among team members when they manage alone. We consider all nine style classifications available in Morningstar. Out of 308 teams, 73 experienced style migrations, or 24% of the teams (Panel A, Table IA12). If the style migrations drive our results, we expect the coefficient δ_1 to be statistically significant and negative. Panel B of Table IA12 shows that δ_1 is statistically indistinguishable from zero. We take this as evidence that style migrations do not drive our results.

IA3.5.4. Mutual Fund Family Switches

A final alternative interpretation of our finding is that the attenuation of overreaction is driven by team members systematically switching mutual fund families when transitioning from single to team management. If fund families have different policies with respect to the abatement of overreaction, then fund managers switching between mutual fund families could generate evidence consistent with our findings.

To address this concern, we first identify the mutual fund family as the one where managers spent most of their time during their solo and team management years, respectively. A mutual fund family switch is then defined as a difference between the mutual fund family of the team and the family of the managers when they operate individually.[10] We then estimate a regression with a similar double-interaction term as in Equation (IA16), where we replace $\Delta Workload_j$ with D_j^{FS}, and where D_j^{FS} indicates a fund family switch for team j.

Table IA12 shows that a switch occurs in 40% of the cases. If mutual fund family switches drive our results, we expect the coefficient δ_1 to be statistically significant and negative. Panel D of Table IA12 shows that δ_1 is statistically indistinguishable from zero. We therefore conclude that switches in mutual fund families cannot explain our result.

[10]For instance, if manager A and B work primarily for Vanguard during their solo years, but as a team the managers work for Fidelity, then we define this case as a discrepancy between the mutual fund family of the team and the solo managers. Likewise, if manager A during its solo years as well as the team primarily work for Vanguard, but manager B works primarily for Fidelity during its solo years, then we also define this case as a family switch.

IA4. Field Evidence: IV Simulations – Transmission Overreaction to Teams

In this section, we simulate data to demonstrate how our instrumental variables (IV) approach is able to produce unbiased estimates of the transmission of individual behavior to the team. We find that our IV estimator successfully deals with the errors-in-variables problem, whereas a standard OLS estimator produces estimates that are biased towards zero.

In the final section of this appendix, we also use our manager stock holdings dataset to show that our main results are not driven by one lucky draw. Given that our IV approach depends on a random split of manager holdings, we show a distribution of IV estimates as a robustness check to confirm our evidence that teams dampen bias transmission of individual behavior. After producing 1,000 random draws and re-estimating our main results, we find that our conclusions remain unchanged.

IA4.1. Data Generating Process

We simulate data to demonstrate that measurement error has a downward bias on the OLS estimates and that our instrumental variable (IV) approach delivers unbiased estimates of the transmission coefficient. We start by simulating the returns of 100 stocks over 40 quarters, which matches the average number of holdings and duration of the mutual fund portfolios that we observe in our sample. Returns for each quarter are simulated with a single factor structure:

$$r_{s,t} = \gamma_s r_{m,t} + \varepsilon_{s,t}, \tag{IA18}$$

where γ_s is the factor exposure of stock s to the single factor $r_{m,t}$, and $\varepsilon_{s,t}$ is the idiosyncratic return of stock s. For each stock, we randomly draw γ_s from a Normal distribution $\mathcal{N}(1, 0.5)$, factor returns $r_{m,t}$ from a Normal distribution $\mathcal{N}(0.0175, 0.1)$, and the idiosyncratic return from a Normal distribution $\mathcal{N}(0, 0.15)$. For the single factor return, these parameter choices translate to an average yearly return of $\mu = 7\%$ with annualized volatility $\sigma = 20\%$. To draw a parallel with our empirical setting, we then calculate the rolling one-year returns of each stock to generate manager trades based on a stock's past return over the last year: $r_{s,t-4\to t}$.

We simulate the trades of 600 managers according to our main empirical specification, where managers change their holdings by using past yearly stock returns:

$$trade_{s,i,t+1} = \beta_i^{SM} r_{s,t-4\to t} + e_{s,i,t+1}, \tag{IA19}$$

where β_i^{SM} equals the true β_X of manager $i = 1, \ldots, N$, $r_{s,t-4\to t}$ the annual past return of stock s, and $e_{s,i,t+1}$ the noise term.

To simulate the manager trades, we draw the true β_X from the empirical distribution of the β_j^X in our sample. Formally, we assume that the individual manager betas are drawn from a normal distribution with mean $\mu_\beta = -0.05$ and standard deviation $\sigma_\beta = 0.5$. This means that we have both managers that extrapolate from past returns as well as managers that take contrarian positions relative to past stock returns. To keep transactions with a similar distribution as in our empirical setting, we then simulate the noise term as a standard normal distribution with mean $\mu_e = 0$ and standard deviation $\sigma_e = 1$. As these sets of parameters do not result in a large measurement error bias, we also use $\sigma_e = 2$ and $\sigma_\beta = 0.2$ to generate more measurement error for illustrative purposes.

We then move on to generate teams of managers and simulate the team trades. We assume that managers are teamed up in non-overlapping pairs such that managers 1 and 2 form a team, 3 and 4, and subsequently managers 599 and 600. Each team trades in the same way as the average of the individual managers, except that their reliance on past returns is mitigated with a fraction κ_0. Formally, we define the changes in the holdings of team $j = 1, \ldots, J$ and its corresponding team beta as:

$$trade_{s,j,t+1} = \beta_j^{TM} r_{s,t-4\to t} + e_{s,j,t+1}, \tag{IA20}$$

where

$$\beta_j^{TM} = \alpha + \kappa_0 \bar{\beta}_j^{SM} + \upsilon_j, \tag{IA21}$$

$$\bar{\beta}_j^{SM} \equiv \frac{\sum_{i=1|j}^{2} \beta_{i|j}^{SM}}{2}, \tag{IA22}$$

where $\beta_{i|j}^{SM}$ is the overreaction of manager i that is in team j and υ_j is a noise term with a $\mathcal{N}(0, 0.1)$ distribution.

Equation (IA21) implies that the team beta is the average beta of the managers composing the team scaled by the transmission coefficient κ_0. When $\kappa_0 = 1$, there is full transmission of heuristic rules as the team manages the portfolio as if each individual manager were trading independently managed portfolios. However, when $\kappa_0 < 1$, team-managed portfolios have a lower tendency to use past stock returns as information for their trades, which can be in an extrapolative or contrarian manner. The opposite is true if $\kappa_0 > 1$, this means that the team exacerbates the usage of past returns as information for future trades.

We also model the transmission of heuristic rules that is conditional on the team composition, namely depending on whether the average team extrapolates or performs contrarian trades:

$$\beta_j^{TM} = \alpha + \kappa_0 \bar{\beta}_j^{SM} + \kappa_1 \bar{\beta}_j^{SM} \times D_j + \kappa_2 D_j + \upsilon_j, \tag{IA23}$$

such that D_j is an indicator variable equal to one when $\bar{\beta}_j^{SM} > 0$. This means that the transmission coefficient will be different for teams that are on average contrarian (κ_0) and those that are on average extrapolative ($\kappa_0 + \kappa_1$).

IA4.2. OLS Estimation

Using the simulated solo manager and team trades, we estimate the respective betas from the simulated data, $\bar{\hat{\beta}}_j^{SM}$ and $\hat{\beta}_j^{TM}$. For the OLS regressions, we regress the estimated team overreaction on the average estimated individual overreaction:

$$\hat{\beta}_j^{TM} = c + \delta_0 \bar{\hat{\beta}}_j^{SM} + \epsilon_j. \tag{IA24}$$

To test for the transmission of heuristic rules depending on team composition, we instead estimate:

$$\hat{\beta}_j^{TM} = c + \delta_0 \bar{\hat{\beta}}_j^{SM} + \delta_1 \bar{\hat{\beta}}_j^{SM} \times \hat{D}_j + \delta_2 \hat{D}_j + \epsilon_j. \tag{IA25}$$

It is important to note that when estimating the regression above, we condition $\hat{D}_j$ on $\bar{\hat{\beta}}_j^{SM}$, because in our empirical setting we are not aware of the true nature of the team, but instead infer the team nature from estimates based on the data (i.e. we condition on the empirically estimated average single beta to determine the nature of the team).

IA4.3. IV Estimation

We now move to the simulations for our instrumental variables approach following Jegadeesh et al. (2019) to solve the error-in-variables problem we face. For the IV method, we estimate overreaction for each manager by splitting the sample used to estimate Equation (IA19), obtaining two separate estimates of β_i^{SM} for each manager, $\hat{\beta}_i^{SM,1}$ and $\hat{\beta}_i^{SM,2}$. We split the sample such that for each manager and quarter we randomly split the holdings into two equally sized samples of stocks, such that $\hat{\beta}_i^{SM,1}$ and $\hat{\beta}_i^{SM,2}$ are estimated using two sets of 50 stocks throughout the 40 quarters. For these two disjoint data samples, we calculate two sets of average overreaction, $\bar{\hat{\beta}}_j^{SM,1}$ and $\bar{\hat{\beta}}_j^{SM,2}$. Because we estimate overreaction on disjoint samples, their measurement errors are uncorrelated. As a result, we can use overreaction of sample one (two) as instrument for sample two (one). Formally, the first and second stage of the IV method are as follows:

$$\begin{aligned} &\text{1st stage: } \bar{\hat{\beta}}_j^{SM,1} = \alpha + \lambda_0 \bar{\hat{\beta}}_j^{SM,2} + u_j, \\ &\text{2nd stage: } \hat{\beta}_j^{TM} = c + \delta_0 \bar{\hat{\beta}}_j^{SM,1,pred} + \epsilon_j, \end{aligned} \tag{IA26}$$

where *pred* indicates the predicted values from the first stage regressions.

Because the measurement errors of $\bar{\hat{\beta}}_{1,j}^{SM}$ and $\bar{\hat{\beta}}_{2,j}^{SM}$ are uncorrelated, so will $\bar{\hat{\beta}}_{2,j}^{SM}$ and ϵ_j, meaning that δ_0 from Equation (IA26) will be an unbiased estimator of κ_0 from our true data generating process. To estimate the IV method conditional on the team composition, we follow a similar procedure where we also instrument for the indicator variable and interaction term using the estimates from the disjoint sample. Because these two additional variables are also estimated from the data, they suffer from measurement error too. Formally, we estimate:

$$\text{1st stage: } \begin{cases} \bar{\hat{\beta}}_j^{SM,1} & = \alpha_1 + \lambda_{1,0}\bar{\hat{\beta}}_j^{SM,2} + \lambda_{1,1}\bar{\hat{\beta}}_j^{SM,2} \times \hat{D}_j^2 + \lambda_{1,2}\hat{D}_j^2 + u_{1,j} \\ \bar{\hat{\beta}}_j^{SM,1} \times \hat{D}_j^1 & = \alpha_2 + \lambda_{2,0}\bar{\hat{\beta}}_j^{SM,2} + \lambda_{2,1}\bar{\hat{\beta}}_j^{SM,2} \times \hat{D}_j^2 + \lambda_{2,2}\hat{D}_j^2 + u_{2,j} \\ \hat{D}_j^1 & = \alpha_3 + \lambda_{3,0}\bar{\hat{\beta}}_j^{SM,2} + \lambda_{3,1}\bar{\hat{\beta}}_j^{SM,2} \times \hat{D}_j^2 + \lambda_{3,2}\hat{D}_j^2 + u_{3,j} \end{cases} \tag{IA27}$$

$$\text{2nd stage: } \hat{\beta}_j^{TM} = c + \delta_0\bar{\hat{\beta}}_j^{SM,1,pred} + \delta_1\bar{\hat{\beta}}_j^{SM,1,pred} \times \hat{D}_j^{1,pred} + \delta_2\hat{D}_j^{1,pred} + \epsilon_j.$$

In all regressions, we take into account the issue of weak instruments as, for instance, raised in Stock and Yogo (2005). To this end, we exclude instruments that have a t-statistic in the first stage that is below 4.05.

IA4.4. OLS and IV Results Simulations

To show that the OLS estimator is biased whereas the IV estimator is not, we run 1,000 simulations and compare the distribution of the coefficient estimates for both methodologies. The simulations are run such that the true parameters reflect our null hypothesis: $\kappa_0 = 1$ and $\kappa_1 = 0$. We start with the OLS and IV estimates of the true team effect in Equation (IA21). The results are depicted in Panel A of Table IA14 where we observe that the OLS estimate gives an average coefficient of 0.986, with a standard deviation equal to 0.021, and we reject $\delta_0 = 1$ in 18% of the cases at a 95% significance level. On the other hand, the IV estimator leads to an unbiased estimate of δ_0 and we find an average coefficient equal to 1.000 with a standard deviation equal to 0.022. For the IV estimator, we are only able to reject the null of a full transmission in 5% of the cases. When estimating the model with interaction terms, we also find similar results in Panel B of Table IA14, where the IV estimator provides us with an unbiased estimate for δ_0. We also run simulations with different parameter values to exacerbate the measurement error bias to illustrate the effectiveness of the IV estimator: whereas the OLS estimate gets more downward-biased if measurement error increases, the IV estimator gives an unbiased estimate equal to the true parameter $\kappa_0 = 1$.

One additional concern is that the measurement error could bias the coefficient of the interaction term as well. However, given that we test the null of $\kappa_1 = 0$, the bias would work against us because of the well-known fact that biases resulting from (uncorrelated) mea-

surement error tend to shrink estimates towards zero.[11] When simulating the transmission coefficient under the null $\kappa_1 = 0$, Panel B of Table IA14 shows that we do not falsely reject the null using both the OLS and IV estimators.

[Place Appendix Table IA14 about here]

IA4.5. Multiple IV Draws

The IV methodology used in our analysis relies on a random draw from a subset of manager holdings. To ensure that the IV results from our main analysis are not driven by one particular draw, we randomly draw 2,000 disjoint samples of the holdings from each manager in a given quarter using the same methodology of the main analysis (Section 2.4-2.5). In particular, for each of these draws, we run the same IV regression as the one specified in Column 6 of Table 3.

In Figure IA2, we show the distribution of the coefficient on $\hat{\beta}_j^{CF}$ for the 2,000 draws where we exclude draws that produce weak instruments.[12] Recall that our null hypothesis of full transmission of heuristics from individuals to the team implies that this coefficient equals 1. The figure shows that the distribution lies slightly below 1, with a mean coefficient equal to 0.80 and a standard deviation of 0.28.

[Place Appendix Figure IA2-IA3 about here]

We also perform the same analysis as before on the sub-samples of teams for which the team members are on average contrarians and for those which are on average extrapolators. In other words, we run the same IV regression as the one specified in Column 8 of Table 3 for multiple draws. In Figure IA3, we find that the distribution of the transmission coefficients is closer to zero for extrapolative teams as opposed to contrarian teams, consistent with our main analysis. The average transmission coefficient for contrarian teams equals 1.01 with a standard deviation of 0.47. These estimates imply that we cannot reject the null of a full transmission for contrarians. On the other hand, the average transmission coefficient for extrapolators equals 0.48 with a standard deviation of 0.25. These estimates show that we reject the null of full transmission but cannot reject the null of no transmission at all.

[11] In unreported simulations, we confirm that the measurement error works against us when simulating data with $\kappa_1 = -1$.

[12] According to Nelson and Startz (1990), to ensure that instruments are not weak, the correlation between the instrument and instrumented variable, $\rho_{xz} >> \frac{1}{\sqrt{N}}$. For an instrument to be included, we require that ρ_{xz} is at least 7 times greater than $\frac{1}{\sqrt{N}}$. This rule also yields a very similar result to requiring a t-statistic of 4.05 in the first stage regression as suggested in Stock and Yogo (2005) as a rule to screen out weak instruments.

IA5. Field Evidence: Cost of Overreaction – Back-of-the-Envelope Analysis

To measure the impact that overreaction has on managers' wealth accumulation, we will proceed in two steps. First, we are going to rely on insights from Ibert et al. (2018) to measure the impact that overreaction has on a manager's labor income in a given year. Second, we are going to accumulate the documented impact over the career length of a manager.

Ibert et al. (2018) use a sample of Swedish mutual fund managers for whom they can observe labor income, fund revenue, and fund performance in excess of the benchmark. With these data, they estimate in Table 4 of their paper the following relation between log labor income (L), fund revenues (Rev), and fund performance (R^{abn}):

$$\begin{aligned} log(L_t) = 0.13 log(Rev_t) + 0.253 log(1 + R_t^{abn}) + 0.586 log(1 + R_{t-1}^{abn}) \\ + 0.583 log(1 + R_{t-2}^{abn}) + 0.274 log(1 + R_{t-3}^{abn}) \end{aligned} \tag{IA28}$$

For convenience, we write $Rev_t = TNA_{t-1} \times ExpenseRatio_t$, where, following Ibert et al. (2018), we take revenue to be independent of fund returns, as these returns appear in the model by themselves. Hence, we can rewrite fund revenues as:

$$\begin{aligned} Rev_t &= TNA_{t-2}(1 + flow_{t-1}) \times ExpenseRatio_t \\ &= TNA_0 \prod_{k=0}^{t-2} (1 + flow_{t-1-k}) \times ExpenseRatio_t. \end{aligned} \tag{IA29}$$

where TNA_0 represents the size of a fund at inception that, as we show below, has no impact on our calculations. Combining Equations (IA28) and (IA29), we can write log labor income of a manager as:

$$\begin{aligned} log(L_t) = \sum_{k=0}^{3} w_k log(1 + R_{t-k}^{abn}) + 0.13 \sum_{k=0}^{t-2} log(1 + flow_{t-1-k}) + \\ 0.13 log(ExpenseRatio_t) + 0.13 log(TNA_0), \end{aligned} \tag{IA30}$$

where w_k represents the coefficient on the performance in year k on log labor estimated in Ibert et al. (2018).

To gauge the impact of overreaction on fund managers' labor income, we need to express fund performance, fund flows, expense ratios, and initial TNA of a fund as a function of overreaction. Following our evidence of the role of overreaction for returns and flows:

- We assume a linear relation between overreaction, i.e., β_X and a fund's abnormal return:

$$R_t^{abn} = R_0^{abn} + \beta^X R^{abnX} + \epsilon_t^R,$$

where we take a manager's overreaction β^X as constant, we are going to treat the underperformance that results from overreaction (R^{abnX}) also as a constant, and the error term ϵ_t^R captures all variation in performance that is unrelated to overreaction.

- Similarly, following our empirical results on the conditioning role of overreaction for fund flows, we model flows as a linear function of overreaction:

$$flow_t = flow_0 + \beta^X flowX + \epsilon_t^f.$$

- We assume that expense ratios and initial assets under management (TNA_0) are independent of overreaction.

After discarding terms that are independent from overreaction, we obtain:

$$log(L_t) = \sum_{k=0}^{3} w_k log(1 + \beta^X R^{abnX}) + 0.13 \sum_{k=0}^{t-2} log(1 + \beta^X flowX). \tag{IA31}$$

Since all quantities are in logs, we can interpret the equation above as describing the percentage-change impact on labor income of a given percentage change in abnormal return and flows:

$$\Delta L_t(\%) = \sum_{k=0}^{3} w_k \Delta R_{t-k}^{abn}(\%) + 0.13 \sum_{k=0}^{t-2} \Delta flows_{t-1-k}(\%). \tag{IA32}$$

Having expressed returns and flows as a function of a manager's overreaction, we can then write changes in labor income in the t^{th} year of the manager's career as:

$$\Delta L_t(\%) = \Delta\beta^X R^{abnX(\%)} \sum_{k=0}^{3} w_k + 0.13 \Delta\beta^X flowX(\%) \sum_{k=0}^{t-2} 1. \tag{IA33}$$

Using the estimates of R^{abnX} (-0.355%) and flowX(%) (-3.658%) from our performance analysis of Table 2 and 3, and Equation (IA30) based on the estimates in Ibert et al. (2018), we finally obtain:

$$\Delta L_t(\%) = \Delta\beta^X (1.696(-0.355\%)) + 0.13(-3.658\%)(t-1). \tag{IA34}$$

With this equation, we can estimate the loss of income and wealth that a fund manager who extrapolates past returns faces. Table IA15 reports the results. In the table, we assess

the losses that are due to overreaction using a number of statistics. In Panel A, we report the average percentage reduction in yearly salary that is due to overreaction. Panel B reports the cumulative reduction, expressed relative to the average yearly salary. In other words, the panel reports how many years of salary are foregone due to overreaction. Panel C and D measure the losses that are due to overreaction by means of a present-value calculation, whereby losses are discounted to the present using either the historical risk free rate (3.24%) or the market return (9.98%), both from Ken French's website.

The main parameters of the back-of-the-envelope analysis are selected as follows:

- To select the number of years a mutual fund managers operates in the industry, we form an educated guess by estimating the difference between the average age at retirement for college graduates and the age at which college graduates enter the workforce. For the former, we use 66 years, as per statistics from Center for Retirement Research at Boston College. As for the latter, we set the age of college degree to 28. This is a conservative estimate, since bachelor and master-level degrees are often completed earlier. These statistics set at 38 years the length of the career of a mutual fund manager. Of course, a mutual fund manager may have worked in another position before becoming a mutual fund manager. Similarly, the mutual fund manager could decide to retire earlier, or he may decide to work as a manager for a longer period. We consider all these possibilities by analyzing the sensitivity of our results to a choice of a shorter work spell of 30 years (e.g., early retirement or a few years in another position) or a longer one of 45 years (late retirement).

- When selecting the relevant change in overreaction $\Delta\beta^X$, multiple choices can be contemplated. The first possible choice is to use a one-standard deviation increase in β^X. As we show in Table IA1, this increase corresponds to 0.6. The next is to consider the simple difference between the average β^X among extrapolators and contrarians. The left panel of Figure 1 indicates that this difference is between 0.37 and 0.60 depending on whether we use full-sample estimates of overreaction or a recursive estimate. We choose conservatively, and pick a value of 0.4. Similarly, one can choose the difference between the average β^X in the top and the bottom quintile of β^X. The right panel of Figure 1 indicates that this difference is between 0.78 and 1.23, depending on whether we use full-sample or recursive estimates of β^X. Once again, we make a conservative selection, and pick 0.8. Finally, one can also be interested in a more modest change in overreaction, and for completeness we include in our analysis also a smaller reduction in β^X of only 0.1.

Other parameters that enter our estimations are:

- The yearly salary of a representative manager who does not extrapolate; we set this to one hundred thousand dollars, and results are fairly insensitive to this dollar amount, since our calculations are based on Ibert et al. (2018), and these in turn concern the growth in salary, rather than the level of the salary itself.[13]

- The discount rate, which affects Panels C and D only. We have experimented with both a risk-free discount rate and the historical market return. Results are reported for both cases.

The results indicate that overreaction can have a sizable impact on managers' labour income. For brevity we concentrate on the results for a 1-standard deviation increase in β^X. We also focus on a 38-year career length, but results with other values of the career span are similar. We find that overreaction leads to a 5% reduction in yearly salary on average during the work life of a manager. Cumulated over the entire work life, such a loss equates to more than 2 yearly salaries, relative to what the manager would have earned on average, had he not extrapolated. Such a reduction in yearly labor income corresponds to a 2.7% to 3.9% reduction in the present value of all manager's income, depending on the selected discount rate.

Of course, the opportunity cost of the loss that is due to overreaction extends beyond the sheer loss of salary. This is because the manager who receives labor income may smooth consumption by investing a portion of his labor income in financial markets. So, when a manager's income is reduced, he does not only miss the income, but he also gives up the profits of the foregone investment. Figure IA4 quantifies such foregone wealth in three possible scenarios. In Panel A, we assume a manager would reinvest all his foregone salary into the risk-free security. In Panel B, we assume full reinvestment in the market. Finally, in Panel C we assume reinvestment in the market portfolio, but we cap the reinvestment rate at 50%.

The figure reinforces that losses due to overreaction can be substantial. For brevity we focus on the intermediate case, a career length of 38 years, one standard deviation $\beta^X = 0.60$, and a market reinvestment of 50%. The figure shows that overreaction causes an asset manager to forego wealth that is close to ten times the yearly manager's income.

Overall, this back-of-the-envelope analysis indicates that overreaction can have a large impact on an asset manager's financial outcomes.

[13] We have also experimented with an alternative scenario where we use Ibert et al. (2018) estimates, jointly with average fund returns, average flows, and average salaries increases over the career of a reference manager reflecting nationwide statistics about salary increases with seniority. Results remain qualitatively similar to the ones presented in Table IA15.

IA6. Experimental Evidence: Details on the Design

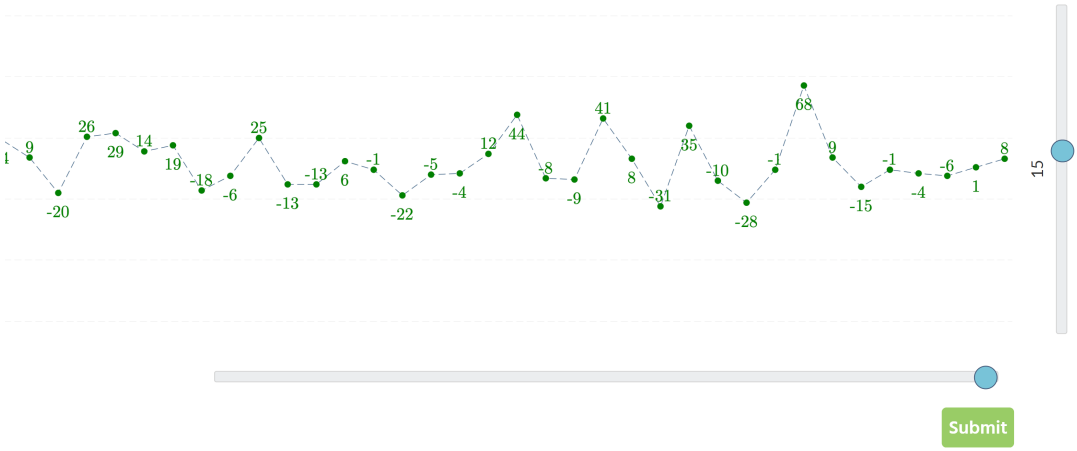
All the participants in the experiment belong to the US-based pool of respondents recruited by Prolific. Show-up fees are \$2.70, \$5.25, \$6.30, and \$3.30 for the Individual, Group, Internal Reflection, and Self-Selection treatments. These show-up fees were determined based on the median time spent by participants on pilot studies run prior to the actual experiments, to match a \$9 hourly rate (data from the pilots is not included in the final dataset). In practice, the total compensation is higher, as it also includes the bonus earned for the accuracy of the predictions.

The experiment begins with a series of instruction slides, which differ based on which treatment the participant is allocated to. The instructions describe the task to be performed with examples accompanied by screenshots that help participants familiarize with it, as well as an illustration of the payoff function. We also include a table with a series of examples to show how different scores translate into dollar compensation. The participants have the option to go back and forth between slides before beginning the task. At the end of the slide deck, the participants are asked whether they want to go back to the first slide and review the instructions again or begin the task. In the latter case, they move to the next stage of the experiment.

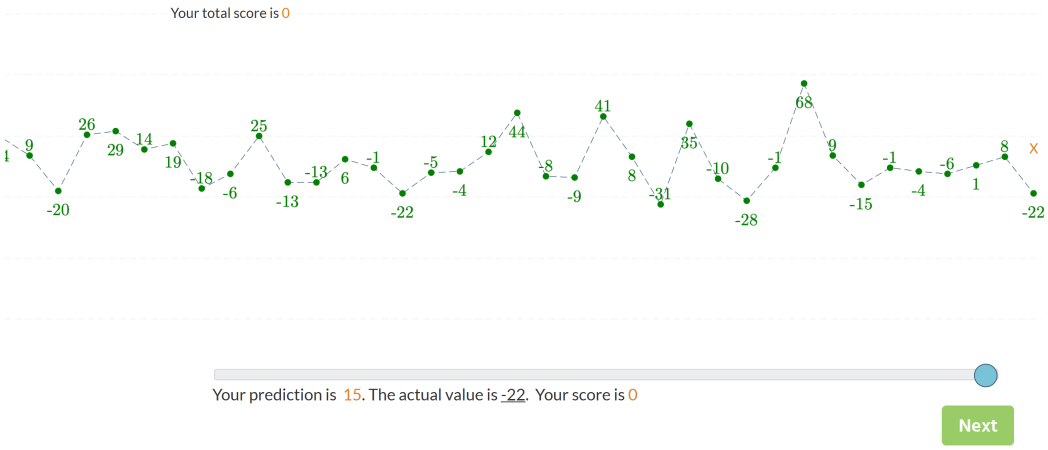
If the participant is allocated to the Individual treatment, the task begins immediately. Participants in the Group, Internal Reflection, and Self-Selection treatments are placed in a "waiting room" before being paired with another participant. The task performed differs depending on the treatment.

IA6.1. Individual Treatment

Participants are shown a series of stock returns of a hypothetical stock. There is a horizontal slider that allows them to scroll back to the beginning of the series, which initially includes forty return realizations. The vertical slider allows them to make their prediction for the next period's return by moving the marker. To avoid any anchoring effects, no marker is shown until they click on the vertical slider.



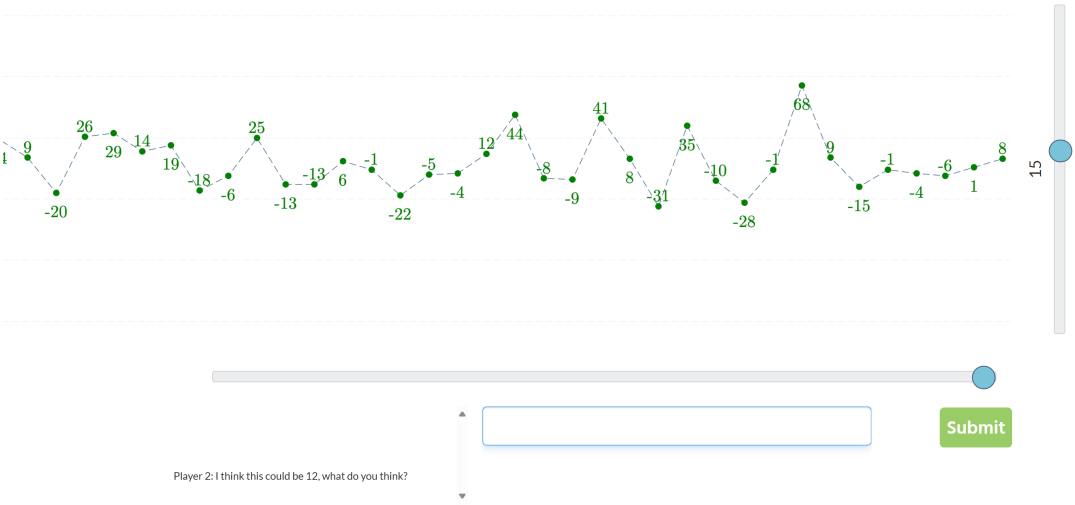
Once they have made a selection, they can click on the "Submit" button and move to the next frame. The correct answer is revealed to them and they are informed of their score.



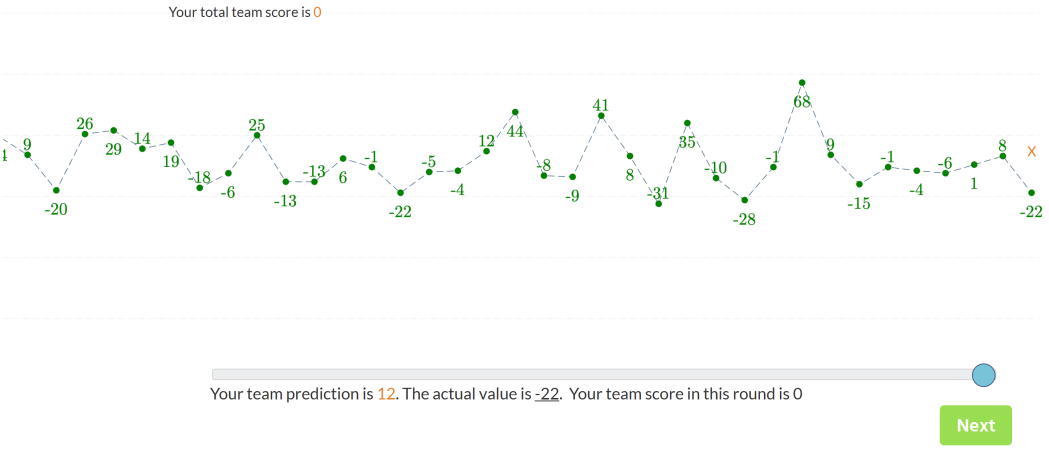
The total score over all the rounds is reported in the top left corner. After clicking on the "Next" button the participants proceed to the next round. They are shown the same series as before, expanded to include the most recent realization (5 in this case) and they have to again guess the next period's return.

IA6.2. Group Treatment

The task is identical to the one performed in the Individual treatment, except that the two participants need to coordinate via a chat box. The "Submit" button becomes active only if their predictions coincide, allowing them to move to the next round.



The correct answer is then revealed to participants, along with the score for the current round and the total score. Once both participants click the "Next" button, they proceed to the next round.



IA6.3. Internal Reflection Treatment

In this experimental arm, each round is repeated twice. The first prediction task is identical to the one performed by participants in the Individual treatment. After having made their selection, participants are shown the same series again, except that now they have to coordinate with the other participant via the chat box, as in the Group treatment. They are also informed of the other participant's prior selection.



Your choice is 21. Player 1's choice is -20. Now make the team choice

sounds good!

Submit

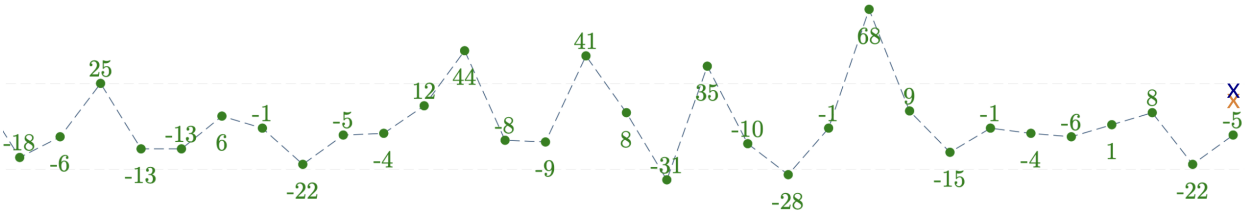
Player 2: should we pick 25?

Player 1: how about 15?

Participants are then revealed the correct answer, the score for the current round, and the total score, separately for the choices made individually and for those made in the team. They are also informed of the other participant's score for the round. Once they both click on the "Next" button they can move to the next round.

Your total individual score is 0

Your total team's score is 75



The actual value is -5. Your individual score is 0. Player 1's individual score is 25.

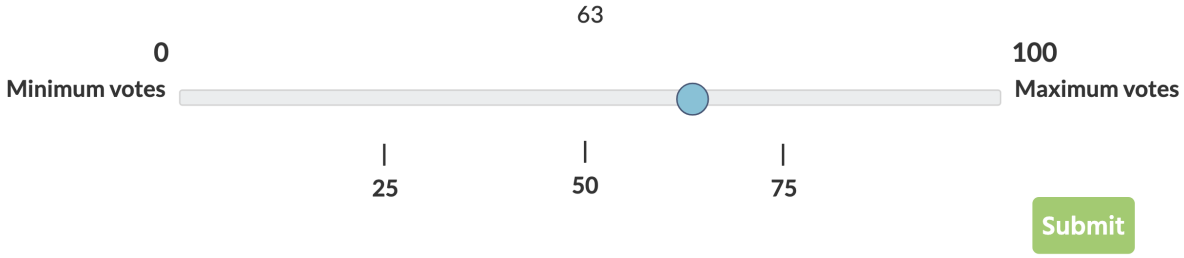
Your team's forecast is 15 and your team's score is 0

Next

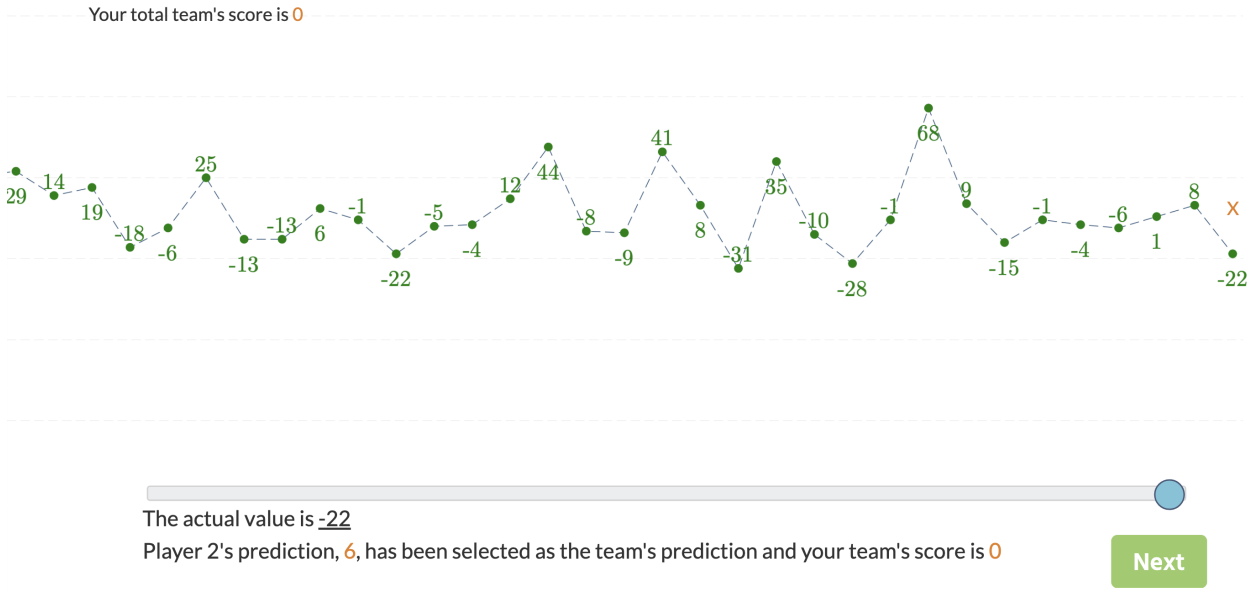
IA6.4. Self-Selection Treatment

In this experimental arm, the participants first perform a prediction task, identical to the one they perform in the Individual treatment. After having made their selection, they are asked to "vote" on their prediction. They are asked to move the marker on the horizontal slider, which corresponds to a vote from 0 to 100.

> How many votes do you want to submit for your own prediction? The prediction of the participant with the highest number of votes will be selected as the team's prediction.



Participants are then displayed the score for the task, and are told whether their prediction was selected as the team prediction, or their teammate's. In the presence of vote ties, the team prediction is the average of the two individual predictions. Non-integer averages are randomly rounded upward or downward.



For each of the three treatments, the task is repeated for twenty rounds. The participants

then move to the survey stage, which is the same independent of the treatment arm.

IA6.5. Survey Questions

Participants are first shown the following message:

> **Thank you for taking the time to complete our Participants Demographics Information Survey.**
>
> In order to ensure that our participants represent a diversity of backgrounds in race, gender, ethnicity, age, socio-economic statuses and more, we are collecting the following information to better direct our recruiting efforts within the lab.
>
> Information you provide in this survey will remain anonymous and will be used by lab personnel only. If we use this information for any of our analyses, as with all your data, the information provided here will be anonymized.

They are then asked to answer a series of questions, listed below.

What gender do you identify with?

- Male
- Female
- Other

How old are you?

In what region of the U.S. do you currently live?

- Midwest
- Northeast
- South
- West
- Do not live in the U.S.

What is the highest level of education you have attained so far?

- Less than a high school degree
- High school diploma (or equivalent)
- Some college but no degree (including academic, vocational, or occupational programs)
- Associate/Junior College Degree (including academic, vocational, or occupational programs)
- Bachelor's Degree (For example: BA, BS)
- Master's Degree (For example: MA, MBA, MS, MSW)
- Doctoral Degree
- Other (please specify)

What is your current employment status?

- ○ Working full-time (for someone or self-employed)
- ○ Working part-time (for someone or self-employed)
- ○ Not working, but would like to work
- ○ Temporarily laid off
- ○ On sick-leave or other leave
- ○ Permanently disabled or unable to work
- ○ Retiree or early retiree
- ○ Student, in school or in training
- ○ Homemaker
- ○ Other (please specify)

Do you consider yourself Hispanic, Latino, or of Spanish origin?

- ○ Yes
- ○ No

Please choose one or more races with which you identify:

- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black or African American
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White
- ☐ Other

Which of the following describes you best in terms of political views?

- ○ Very Conservative/Republican
- ○ Conservative/Republican
- ○ Moderate
- ○ Liberal/Democrat
- ○ Very Liberal/Democrat

Which of the following candidates are you most likely to vote for in the 2024 presidential election? terms of political views?

- ○ Joe Biden
- ○ Donald Trump
- ○ Third party candidate
- ○ Undecided/Prefer not to answer

What is your marital status?

- ○ Single (never married)
- ○ Living with partner
- ○ Married
- ○ Separated
- ○ Widowed
- ○ Divorced
- ○ I prefer not to answer

What is the 5-digit ZIP code of your primary residence (the place where you usually live)?

How many people usually live in your primary residence, other than yourself?

Do you own your current primary residence?

- Yes
- No

In which year did you become the owner of your current primary residence? *(if answered yes to the previous question)*

Have you ever traded the stock of a publicly listed company?

- Yes
- No

Have you ever held shares in a mutual fund?

- Yes
- No

Do you hold any stocks or other equity investments at this time, either through a retirement fund (401k etc.) or a personal portfolio?

- Yes
- No

Do you hold any stocks or other equity investments at this time, either through a retirement fund (401k etc.) or a personal portfolio?

- Yes
- No

Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please select a number between 1 and 7 where 1 means "absolutely unwilling to take risks" and 7 means "fully prepared to take risks."

Absolutely unwilling to take risks	1	2	3	4	5	6	7	Fully prepared to take risks
	○	○	○	○	○	○	○	

On a scale from 1 to 7, how would you rate your willingness to give up something that is beneficial for you today in order to benefit more from that in the future? 1 means "Completely unwilling to give up" and 7 means "Very willing to give up."

	1	2	3	4	5	6	7	
Completely unwilling to give up	○	○	○	○	○	○	○	Very willing to give up

On a scale from 1 to 7, would you say that most people can be trusted or that you need to be very careful in dealing with people?

	1	2	3	4	5	6	7	
Need to be very careful	○	○	○	○	○	○	○	Most people can be trusted

Which category represents the total combined pre-tax income of all members of your household (including you) during the past 12 months?

- Less than $10,000
- $10,000 - $19,999
- $20,000 - $29,999
- $30,000 - $39,999
- $40,000 - $49,999
- $50,000 - $59,999
- $60,000 - $69,999
- $70,000 - $79,999
- $80,000 - $89,999
- $90,000 - $99,999
- $100,000 - $149,999
- $150,000 - $199,999
- $200,000 or more

Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?

- More than $102
- Exactly $102
- Less than $102
- Do not know
- Refuse to answer

Please tell me whether this statement is true or false. "Buying a single company's stock usually provides a safer return than a stock mutual fund."

- TRUE
- FALSE
- Do not know
- Refuse to answer

If you're running a race and you pass the person in second place, what place are you in?

A farmer has 15 sheep and all but 8 died. How many are left?

Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?

How many cubic feet of dirt are there in a hole that is 3 feet deep × 3 feet wide × 3 feet long?

On a scale of 1 to 5, how interesting was the experiment today?

- 1 (Absolutely uninteresting)
- 2 (Somewhat uninteresting)
- 3 (Neither interesting, nor uninteresting)
- 4 (Somewhat interesting)
- 5 (Absolutely interesting)

On a scale from 1 to 7, how satisfied are you with the instructions of the experiment? 1 means "Completely unsatisfied" and 7 means "Completely satisfied."

	1	2	3	4	5	6	7	
Completely unsatisfied	○	○	○	○	○	○	○	completely Satisfied

Do you have comments about the instructions or the main experiment? Type them below.

Participant Answer...

Participants are informed of their total payoff, split between show-up fee, and bonus for accuracy.

IA6.6. Detecting Suspicious Observations

As discussed in the main text, after the data collection we realized that some participants had achieved implausibly good performances, in terms of accuracy of their predictions. This suggests that some participants were able to share the realized return series with each other,

or that the same user was using multiple accounts. This was a possibility we could not fully prevent given the online nature of the experiment.

In the Group and Internal Reflection treatments, the presence of abnormal behavior can also be detected from the chat transcripts. For example, we recorded the following exchange for a team of players in the Internal Reflection Treatment:

Round 1:

Player 1: choose -22, i know about stocks

Player 2: sounds good

Round 2:

Player 1: choose anything i choose

Player 2: ok

From then on, the two participants keep giving the correct prediction, achieving a perfect score in the team prediction task (with Player 1 obtaining a perfect score in the individual rounds as well).

Fortunately, detecting anomalous observations is not difficult and they can fairly easily be spotted and removed from the final sample. We proceed as follows. First, we simulate 10,000 series of 20 guesses. To be conservative, we assume that the simulated participants "know" that the original return series follows a normal distribution with mean 0 and standard deviation equal to 20. From this simulation, we obtain an estimate of the probability that a participant makes 0, 1, 2, ..., 20 correct predictions.

We find that 0, 1, and 2 correct guesses occur in 74.64%, 22.16%, and 2.91% of our simulations. 4 (0.02% of the simulations) or more correct guesses are considered near-impossible and removed from the sample. This filter led us to remove 2 participants from the Individual treatment, 5 teams from the Group treatment, 22 teams from the Internal Reflection treatment, and 7 teams from the Self-Selection treatment.

3 correct guesses occur in 0.27% of our simulations; thus, while unlikely, they are not entirely implausible. In our entire sample, only five teams (two in the Group treatment and three in the Self-Selection treatment) make three correct guesses. To detect whether either of these two teams' performances may be due to cheating, we also obtain empirical distributions of the dollar payoff and of the number of predictions with the correct sign. In other words, we recognize that individuals may achieve superior performance by making predictions that are systematically *close* to the target return.

Of the two Group treatment teams, one achieves a payoff of \$2.28 and correctly guesses the sign of the next return 20 out of 20 times. In only one of our simulations we observe a payoff over \$2, and we never observe a simulation where the sign was always correctly guessed. We therefore remove this team. The other team achieves a payoff of \$1.09, a good but not exceptional performance in our sample, and the sign is correctly guessed in only 10

out of 20 rounds. We therefore keep this team in our final sample.

Based on this \$2 threshold, we additionally remove one individual participant, two teams in the Group treatment, and two teams in the Self-Selection treatment, all of which have obtained a payoff higher than \$2. In total, our filtering procedure leads to the removal of 3 individuals in the Individual treatment, 9 teams in the Group treatment, 22 teams in the Internal Reflection treatment, and 9 teams in the Self-Selection treatment.

While we are confident that we have detected all the cases where a participant may have been aware of the correct answers, we can also adopt more stringent filters. For example, removing all the participants who earned more than \$1.5 (which results in the removal of 6 individuals in the Individual treatment, 4 teams in the Group treatment, and 2 teams in the Internal Reflection treatment) leads to virtually identical results.

As discussed in the main text, we had targeted a sample size of 250 participants for the Individual treatment and 500 participants (250 teams) for the other treatments. Thus, our targeted final sample size was 1,750, as opposed to the 1,512 we ultimately use. Aside from the cleaning we just described, we lose several participants as they are "repeated study takers." A participant may start the study without completing it, for example because either they or their teammate experience internet problems. Although we explicitly tell them in the instruction slides not to re-attempt a submission if they are unable to complete, and that we would be willing to pay partial compensation for their time, many do it anyway. Yet, if they complete a submission on a second attempt, we consider it invalid and do not include it in the sample.

The attrition is especially severe in the case of multiuser studies. Suppose, for example, that participants A and B join a team and, for some reason, are unable to complete the study. If A joins a new participant C and B joins a new participant D, and both teams complete the study, we consider all these four submissions invalid. A and B cannot be included in the sample because they have already been exposed to the task, but also C and D, as they have been paired with participants who have a partial knowledge of the task. However, all four participants are rewarded, even though they do not enter the final sample, raising the cost of the experiment, which is why we ultimately chose to fall slightly short of our targeted sample sizes. Once we exclude repeated experiment takers we are left with our final sample of 1,512 observations of participants with a *single* attempt (248, 456, 405, and 403 for the Individual, Group, Internal Reflection, and Self-Selection treatments, respectively).

IA7. Experimental Evidence: Additional Results

This appendix presents additional results from the experiment omitted from the main text for brevity. We start by presenting summary statistics on key variables by treatment. Table IA16 presents summary statistics for compensation, mean absolute error, mean squared error, and predictions made by the participants in the four treatment groups.

We then present Table IA17, which replicates the analysis of Table 9 in the main text, using the alternative overreaction proxy β_w as the dependent variable. Specifically, we estimate an alternative model of belief overreaction:

$$\widehat{E}_t x_{i,t+1} = \alpha_w + \beta_{w,i} \sum_{j=1}^{4} w_j x_{t+1-j} + \varepsilon_{i,t} \tag{IA35}$$

where $w_j \equiv \frac{\lambda^{j-1}}{\sum_{j=1}^{4} \lambda^{j-1}}$. Based on estimates in Greenwood and Shleifer (2014), we pick a value of the decay parameter that is micro-founded in surveys, and set $\lambda = 0.56$. We then repeat the analysis in Equation (8) in the main text using this alternative overreaction metric. We find that teams attenuate overreaction by 55% and a large fraction of this effect is explained by self-selection.

In Table IA18, we further investigate the team interactions in the Group treatment, and present suggestive evidence that team interaction (i) can serve as a tool for self-selection, and (ii) can counteract the presence of extreme beliefs.

Before going over the results, we describe the data collection steps we implemented to carry out the analysis. We first go over all the chat transcripts and manually extract the "first guess," namely the first prediction proposed by any of the two players. It turns out that the first guess is a key metric, as it coincides with the final team prediction 84.5% of the time. Because of that, we identify the participant who makes the first guess as the team "leader" for that particular task.

We then check whether this first guess is "challenged" by the other player, typically resulting in a different final prediction, and construct a "challenged first guess" dummy. A challenged guess is formally defined by one where either the final prediction is different from the first guess or where the participant that did not make the first guess proposes at least one alternative prediction.

In Columns 1 and 2 of Table IA18 we test the self-selection channel in the Group treatment. Here we make use of information on the first guess, focusing on *who* makes it. Our hypothesis is the following: if a participant has done very well in making a prediction, the other team member should be more likely to "opt out," and leave her the choice of making the next prediction as well. Conversely, poor predictions should lead to a role reversal, with

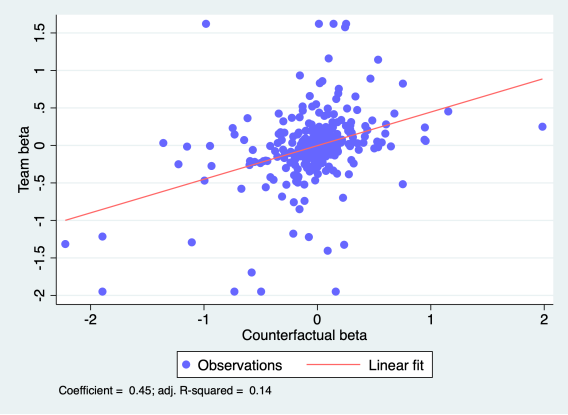
the other team member taking the leading role. In other words, we are hypothesizing that participants *self-select* in or out of the leading role in this "committee," as in Enke et al. (2023), but dynamically, as their confidence level is affected by their performance.

Our proxy for "taking the leading role" is making the first guess. We construct the "repeat first guess" dummy, equal to one if the participant who makes the first guess at time t is the same who made it at $t-1$. Then, we regress this dummy on the absolute error of their first prediction at $t-1$ (divided by the sample standard deviation for ease of interpretation). Our hypothesis is that, if Player 1 took the leading role in round $t-1$, and the forecast error turned out to be small, she is going to be more likely to take the lead in round t as well. This is precisely what we find. In the specification of Column 2, which includes the usual vector of control variables, we find that a one-standard deviation increase in the forecast error reduces the likelihood of having the same participant making the first guess by six percentage points. This effect is statistically significant at the 1% level.

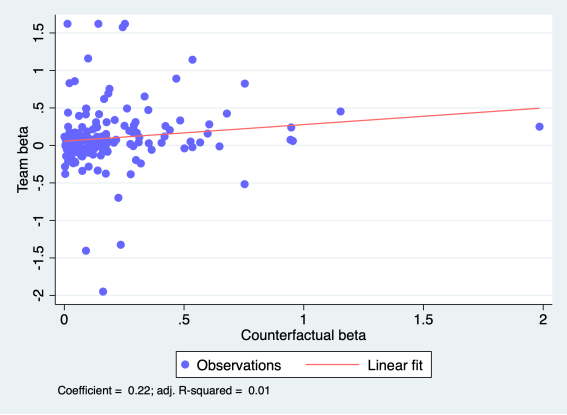
In Columns 3 and 4 of Table IA18, we then test whether the first guess is more likely to be challenged when it is further away from the conditional mean of our time series, which is always 0, and corresponds to the rational prediction benchmark. That is, does the team member that does not take the first initiative still influence the final prediction of the team and correct potential mistakes of the teammate?

We estimate a linear probability model where the dependent variable is a challenged first guess dummy and the key coefficient of interest is the absolute value of the first guess, as extracted from the chat transcripts (again standardized for ease of interpretation). We find that a one standard-deviation increase in the first guess-error measure leads to a 2.5 percentage point increase in the likelihood of a "challenge," relative to an unconditional probability of 18%. Again, the coefficient is significant at the 1% level.

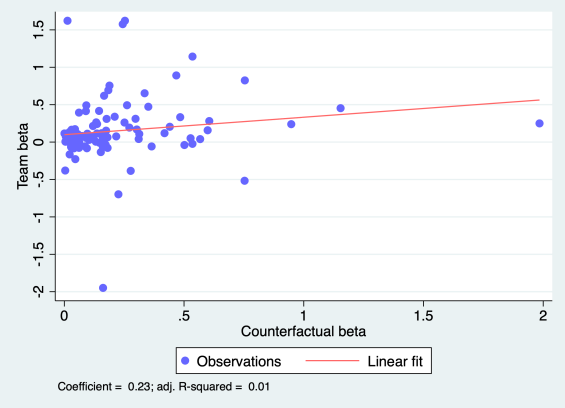
Figure IA1. **Field Evidence – Overreaction in Teams for Non-learner Managers**: This figure plots the team-level overreaction metric ($\hat{\beta}_j^{TM}$) against the average overreaction that team members display when managing a single-managed fund ($\hat{\beta}_j^{CF}$). The plot is presented for all teams (Panel A), all extrapolative teams (Panel B), and the extrapolative teams that consist of "non-learner" managers (Panel C). A team is defined as extrapolative if its members extrapolate on average (i.e., $\hat{\beta}_j^{CF} > 0$). The non-learners sample is defined as the set of managers who extrapolate in both the first half and the second half of their tenure as individual fund manager.



(a) Panel A: All Teams



(b) Panel B: Extrapolators



(c) Panel C: Extrapolative Non-Learners

Figure IA2. **Field Evidence – IV Simulations**: This figure shows the distribution of transmission coefficients (δ_0) that we obtain for different random samples using our IV methodology based on Equation (5), excluding the interaction term. The mean of the transmission coefficient equals 0.80 with a standard deviation of 0.28. The transmission coefficient estimated in the main specification in Column 6 of Table 3 equals 0.85.

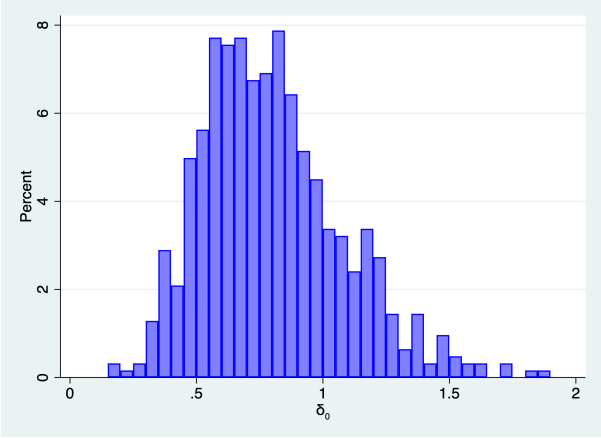
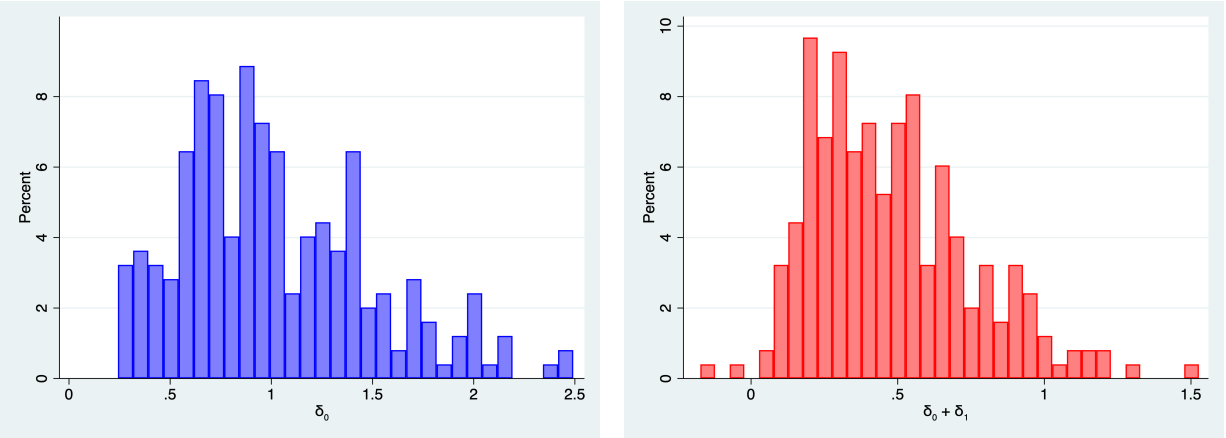


Figure IA3. **Field Evidence – IV Simulations Contrarians versus Extrapolators**: This figure shows the distribution of transmission coefficients for contrarians (δ_0) and extrapolators ($\delta_0+\delta_1$) that we obtain for different random samples using our IV methodology based on Equation (5). The average transmission coefficient for contrarians equals 1.01 with a standard deviation of 0.47 and for extrapolators the average transmission coefficient equals 0.48 with a standard deviation of 0.25. The transmission coefficient estimated in the main specification in Column 8 of Table 3 equals 1.17 for contrarians and 0.50 for extrapolators.



(a) Contrarians (b) Extrapolators

Figure IA4. **Field Evidence – Overreaction and Foregone Wealth Accumulation with Reinvestment**: This figure shows the foregone wealth accumulation due to overreaction, after accounting for the missed reinvestment opportunities. In every year during the work life of a manager, we measure the manager's salary loss due to overreaction. We then assume the manager would have reinvested this portion of salary either in full (Panel A and C), or only in part (Panel B, 50% reinvestment rate). We contemplate either the reinvestment at the risk-free rate (Panel A), or the reinvestment in the market portfolio (Panel B and C). Each panel reports the cumulative end-of-work-life value of the investment losses that are due to overreaction, in multiples of average yearly salary.

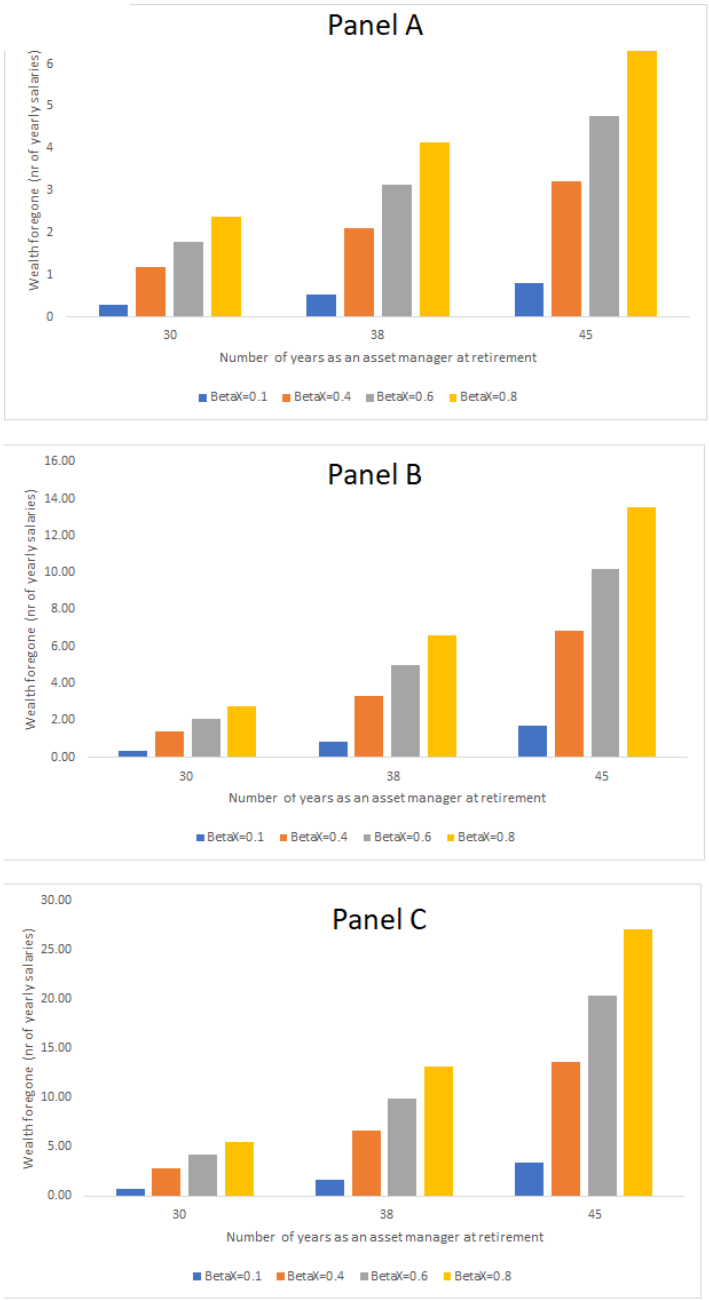


Table IA1. **Field Evidence – Summary Statistics Active Domestic US Equity Funds**: This table summarizes the full sample fund overreaction metrics and factor exposures, and time-series averages of the remaining variables for the domestic US equity fund sample for the period 1980Q1-2018Q4. β_j^X is the fund's overreaction metric defined in Section 2.3; β_j^X *Dummy* is equal to 1 when the fund's β_j^X is positive; β_j^X *(No Mom)* is the same metric estimated excluding momentum stocks; β_j^X *(Weight Change)* is the same metric estimated with portfolio weight changes on the left-hand side of Equation (1); β_j^X *(Past Year Ret)* is the same metric estimated using past 1 year realized stock returns on the right-hand side of Equation (1). *Fund TNA* is the fund's total net asset value (millions); *Team-Managed* is a dummy equal to 1 if a fund is team-managed; *Number of Managers* is the number of managers managing a fund; *Manager Experience* is the experience of the managers (quarters); *Number of Stocks* is the number of stocks held by a fund; *Fund Age* is the age of the fund since inception (years); *Expense Ratio* is the fund's total expense ratio (p.p.); *Fund Turnover* is the CRSP turnover ratio of a fund; β_{MOM} is the fund's exposure to the momentum factor in a Fama-French-Carhart (FFC) model; *Disposition* is the fund-level disposition effect; *Fund Return* is the fund's yearly raw return; *Benchmark Adj. Return* is the fund's yearly return in excess of the fund's primary prospectus benchmark; *CAPM Alpha* is the cumulative CAPM yearly alpha; *FF3 Alpha* is the cumulative Fama-French (FF) 3-factor yearly alpha; *FF5 Alpha* is the cumulative FF 5-factor yearly alpha; *Flow* is the yearly fund inflow; *Benchmark Adj. Volatility* is the volatility of monthly benchmark-adjusted returns over 12 months; *FF5 Alpha Volatility* is the volatility of monthly FF5 alpha over 12 months. The performance metrics are estimated on gross-of-fee returns (p.p.). All variables are winsorized at the 0.5th and 99.5th percentile. Panel B reports pairwise correlations of the aforementioned overreaction metrics, and their correlation with momentum, disposition effect, and factor loadings from a FF5 model.

Panel A: Summary statistics

Mutual fund characteristics	Mean	St. Dev.	5th Pct.	Median	95th Pct.
β_j^X	-0.01	0.60	-0.51	0.00	0.37
β_j^X Dummy	0.50	0.50	0	1	1
β_j^X (No Mom)	-0.01	0.46	-0.57	0.00	0.49
β_j^X (Weight Change)	-0.01	0.62	-0.50	0.00	0.38
β_j^X (Past Year Ret)	0.00	0.10	-0.09	0.00	0.07
Fund TNA	1357.18	4072.71	22.44	361.75	5435.25
Team Managed	0.68	0.36	0	1	1
Number of Managers	2.76	2.16	1	2	6
Manager Experience	32.78	15.08	13.52	30.37	61.90
Number of Stocks	96.44	159.73	26	60	253
Fund Age	14.86	11.59	4.67	11.81	38.04
Expense Ratio	1.19	0.39	0.64	1.16	1.81
Fund Turnover	0.80	0.54	0.20	0.69	1.81
β_{MOM}	0.01	0.10	-0.14	0.01	0.20
Disposition	-0.02	0.06	-0.13	-0.01	0.08
Mutual fund performance					
Fund Return	9.80	5.17	1.18	10.24	16.10
Benchmark Adj. Return	0.32	2.38	-3.48	0.29	4.11
CAPM Alpha	0.12	3.07	-4.51	0.16	4.42
FF3 Alpha	-0.35	2.35	-4.07	-0.29	3.01
FF5 Alpha	-0.12	2.54	-3.80	-0.25	3.88
Flow	3.50	23.54	-21.47	-0.82	39.67
Benchmark Adj. Volatility	1.53	1.67	0.61	1.31	2.80
FF5 Alpha Volatility	1.31	1.73	0.59	1.13	2.15

Panel B: Correlation table

β_j^X	1										
β_j^X (No Mom)	0.73	1									
β_j^X (Weight Change)	0.97	0.69	1								
β_j^X (Past Year Ret)	0.90	0.58	0.87	1							
β_{MOM}	0.15	0.16	0.13	0.10	1						
Disposition	-0.13	-0.13	-0.12	-0.09	-0.60	1					
β_{MKT}	-0.03	0.03	-0.02	-0.02	-0.11	0.01	1				
β_{SMB}	0.00	0.01	0.00	0.00	0.24	0.00	0.17	1			
β_{HML}	-0.16	-0.20	-0.15	-0.15	-0.59	0.52	-0.10	-0.08	1		
β_{RMW}	-0.08	-0.09	-0.07	-0.07	-0.27	0.26	-0.09	-0.08	0.62	1	
β_{CMA}	-0.13	-0.17	-0.13	-0.15	-0.21	0.29	-0.16	-0.19	0.44	0.46	1

Table IA2. **Field Evidence – Fund Characteristics in Solo-managed Funds versus Team-managed Funds**: This table compares the characteristics of the team with its statistical counterfactual, i.e. the average characteristics of the team members when they operate individually. *Fund TNA* is the team's (counterfactual's) time-series average total net asset value of the funds they manage; *Manager Experience* is the team's (counterfactual's) time-series average experience; *Number of Stocks* is the team's (counterfactual's) time-series average of the number of stocks held; *Fund Age* is the team's (counterfactual's) time-series average age of the funds since inception that they manage; *Expense Ratio* is the team's (counterfactual's) time-series average total expense ratio; β_{MOM} is the team's (counterfactual) time-series average momentum exposure; *Disposition* is the team's (counterfactual's) time-series average measured disposition effect. We report the results for all teams combined (Panel A), for contrarian teams (Panel B), and for extrapolative teams (Panel C). A team consists mainly of contrarians if $\hat{\beta}_j^{CF} \leq 0$ and of extrapolators if $\hat{\beta}_j^{CF} > 0$.

Panel A: All Teams

	CF	Team	CF - Team	t-stat	Obs.
Fund TNA	1310.48	1215.03	95.44	0.63	308
Manager Experience	30.79	32.34	-1.54	-1.51	308
Number of Stocks	73.49	79.65	-6.16	-2.47	308
Fund Age	16.91	16.22	0.70	0.93	308
Expense Ratio	1.25	1.28	-0.03	-1.66	308
β_{MOM}	0.02	0.02	0.00	-0.68	308
Disposition	-0.02	-0.02	-0.01	-1.44	308

Panel B: Contrarians

	CF	Team	CF - Team	t-stat	Obs.
Fund TNA	1331.30	1126.35	204.95	1.13	143
Manager Experience	30.60	33.46	-2.86	-1.87	143
Number of Stocks	69.51	70.66	-1.15	-0.42	143
Fund Age	15.91	14.87	1.05	1.14	143
Expense Ratio	1.23	1.30	-0.07	-2.48	143
β_{MOM}	0.00	0.00	0.00	-0.15	143
Disposition	-0.02	-0.01	-0.01	-1.88	143

Panel C: Extrapolators

	CF	Team	CF - Team	t-stat	Obs.
Fund TNA	1292.43	1291.89	0.54	0.00	165
Manager Experience	30.97	31.37	-0.40	-0.30	165
Number of Stocks	76.94	87.44	-10.50	-2.63	165
Fund Age	17.78	17.39	0.39	0.34	165
Expense Ratio	1.26	1.26	0.00	0.12	165
β_{MOM}	0.03	0.04	-0.01	-0.88	165
Disposition	-0.03	-0.03	0.00	-0.22	165

Table IA3. **Field Evidence – Alternative Overreaction Metrics**: This table shows robustness to three alternative ways to measure overreaction: i) exclusion of momentum stocks; (ii) using a weight change measure to identify portfolio changes; and (iii) using realized past stock returns. In Panel A, we repeat our analysis from Table 4 on the linkage between overreaction and fund performance using these alternative overreaction metrices. In Panel B, we use the alternative overreaction metrics to repeat the analysis in Table 3 on how overreaction is transmitted from individual team members to teams. For the *No Momentum* overreaction metric, we re-estimate managers' overreaction after excluding stocks that are part of the momentum strategy (Jegadeesh and Titman 1993, 2001). In each month t, a stock is classified as part of the momentum strategy in month t, and hence removed from the estimation of managers' β_j^X, if its cumulative 11-month return between the end of month $t-12$ and the end of $t-1$ is in the top or the bottom 10% of the cross-sectional distribution of stock returns. For the *Weight Change* overreaction metric, we use active weight changes as specifcied in Equation (IA12) as the main dependent variable in Equation (1). For the *Past 1 Year Return* overreaction metric we use realized past year returns for the main independent variable $r_{s,t-4\to t}$ in Equation (1). More details on the construction of the alternative overreaction metrics are in Section IA2.3. Standard errors are in brackets, and in Panel A, we cluster standard errors by quarter and at the fund level. Significance: ***99%, **95%, *90%.

Panel A: Fund Performance

	No Momentum		Weight Changes		Past 1 Year Return	
	BM Adj. Return	**FF5 Alpha**	**BM Adj. Return**	**FF5 Alpha**	**BM Adj. Return**	**FF5 Alpha**
	(1)	(2)	(3)	(4)	(5)	(6)
β_j^X	-0.251**	-0.199***	-0.257*	-0.374***	-1.331*	-1.394***
	[0.106]	[0.075]	[0.148]	[0.107]	[0.684]	[0.457]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time × Style FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67628	67623	67634	67629	67634	67629
Adj. R-squared	0.1012	0.2252	0.101	0.2255	0.1011	0.2253

Panel B: Transmission of overreaction from solo to team-managed funds

	No Momentum Stocks		Weight Changes		Past 1 Year Return	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}_j^{CF}$	0.8793***	1.1708**	0.7657***	0.9021***	0.6469***	0.8733***
	[0.2325]	[0.5086]	[0.0958]	[0.1898]	[0.1215]	[0.3011]
$\hat{\beta}_j^{CF} \times D_j^E$	-0.8477***	-1.0344**	-0.6501***	-0.8052***	-0.3644**	-0.7548***
	[0.2585]	[0.4283]	[0.1079]	[0.1730]	[0.1596]	[0.2827]
D_j^E	0.0728	0.5436	0.0395	0.2083	0.0051	0.0263
	[0.0887]	[0.4279]	[0.0483]	[0.1589]	[0.0165]	[0.0441]
Style fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Team controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	307	307	307	307	307	307
Adj. R-squared	0.1553	0.0734	0.3165	0.2124	0.1896	0.0895
Hypothesis testing:						
p-value $H_0 : \delta_0 = 1$	0.604	0.737	0.015	0.606	0.004	0.674
p-value $H_0 : \delta_0 + \delta_1 = 0$	0.774	0.573	0.025	0.134	0.006	0.377
p-value $H_0 : \delta_0 + \delta_1 = 1$	0.000	0.000	0.000	0.000	0.000	0.000

Table IA4. **Field Evidence – Alternative Assumptions Overreaction Metric**: This table shows robustness to some of the assumptions that we make to measure overreaction in our main analysis. Specifically, in the analysis labeled *Horizon Universe* we re-estimate managers' overreaction using an investment universe that only incorporates the stocks that the investor start to hold within the next year, but we keep all the stocks the investor has held in the previous 11 quarters. For the analysis labeled *No Universe*, we estimate overreaction based only on actual trades. This approach effectively takes the definition of the appropriate investment universe off the empiricists' hands. Finally, in the analysis labeled *Separate Inflows/Outflows*, we relax an implicit assumption in Section 2.3, namely, that flow-induced trading is symmetric for inflows and outflows. To better incorporate the evidence in Lou (2012), we model the impact of inflows and outflows on a fund's trading behavior by incorporating on the right hand side of Equation (1) a dummy variable that indicates whether the fund received inflows in quarter $t+1$, $D^{inflows}_{j,t+1}$, interacted with the stock-level variables defined in $F_{s,j,t}$ (Section 2.3, Equation (1)). In Panel A, we repeat our analysis from Table 4 on the linkage between overreaction and fund performance using these alternative overreaction metrices. In Panel B, we use the alternative overreaction metrics to repeat the OLS regressions of Table 3 on how overreaction is transmitted from individual team members to teams. Standard errors are in brackets, and in Panel A, we cluster standard errors by quarter and at the fund level. Significance: ***99%, **95%, *90%.

Panel A: Fund Performance

	Horizon Universe		No Universe		Separate Inflows/Outflows	
	BM Adj. Return	**FF5 Alpha**	**BM Adj. Return**	**FF5 Alpha**	**BM Adj. Return**	**FF5 Alpha**
	(1)	(2)	(3)	(4)	(5)	(6)
β_j^X	-0.315**	-0.278***	-0.205***	-0.183***	-0.365***	-0.358***
	[0.121]	[0.085]	[0.075]	[0.053]	[0.129]	[0.091]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time × Style FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67634	67629	67634	67629	67634	67610
Adj. R-squared	0.1013	0.2254	0.1013	0.2255	0.1013	0.2255

Panel B: Transmission of overreaction from solo to team-managed funds

	Horizon Universe		**No Universe**		**Separate Inflows/Outflows**	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}_j^{CF}$	0.5155***	0.4850***	0.5312***	0.5099***	1.2978***	1.1819***
	[0.0770]	[0.0771]	[0.0701]	[0.0703]	[0.1496]	[0.1506]
$\hat{\beta}_j^{CF} \times D_j^E$	-0.2793**	-0.2308**	-0.2516**	-0.2408**	-0.9663***	-0.8581***
	[0.1187]	[0.1154]	[0.1221]	[0.1179]	[0.1838]	[0.1823]
D_j^E	0.1532**	0.0823	0.1772	0.1687	-0.0596	-0.0662
	[0.0707]	[0.0708]	[0.1269]	[0.1243]	[0.0591]	[0.0590]
Style fixed effects	No	Yes	No	Yes	No	Yes
Team controls	No	Yes	No	Yes	No	Yes
Observations	308	307	308	307	307	306
Adj. R-squared	0.2141	0.2616	0.2298	0.2708	0.2608	0.2983
Hypothesis testing:						
p-value $H_0 : \delta_0 = 1$	0.000	0.000	0.000	0.000	0.047	0.228
p-value $H_0 : \delta_0 + \delta_1 = 0$	0.009	0.004	0.006	0.006	0.002	0.002
p-value $H_0 : \delta_0 + \delta_1 = 1$	0.000	0.000	0.000	0.000	0.000	0.000

Table IA5. **Field Evidence – Explanatory Power of Momentum Trading and the Disposition Effect for Overreaction**: This table estimates regressions of our overreaction metrics on a fund's momentum trading behavior and on a fund's disposition effect. We measure momentum trading as the loading of a fund's return on the momentum factor (Carhart, 1997). We measure the disposition effect following Odean (1998) and Cici (2012) (see Section IA2.1 for more details). Following the analyses elsewhere in the paper, we show the results for four distinct measures of overreaction: i) the main one, that uses all stocks traded by a fund; ii) an alternative one (No Momentum), that excludes momentum stocks; (iii) an alternative metric that uses weight changes as the way to measure funds' trading behavior (Weight Changes); and (iv) an alternative metric for overreaction that relies on past 1-year returns without using the structural estimates of extrapolators' memory in Greenwood and Shleifer (2014) (Past Year Return). More details about these metrics are in Section IA2.3. Standard errors are in brackets. Significance: ***99%, **95%, *90%.

	Main Specification	**No Momentum**	**Weight Changes**	**Past Year Return**
	(1)	(2)	(3)	(4)
β_{MOM}	0.674***	0.583***	0.575***	0.073***
	[0.097]	[0.117]	[0.106]	[0.017]
Disposition	-0.637**	-0.394*	-0.670**	-0.083
	[0.323]	[0.213]	[0.337]	[0.053]
Constant	-0.034*	-0.026**	-0.029	-0.004
	[0.018]	[0.012]	[0.018]	[0.003]
Observations	2056	2056	2056	2056
R-squared	0.0296	0.0289	0.0235	0.0142
Adj. R-squared	0.0286	0.0279	0.0226	0.0132

Table IA6. **Field Evidence – Spanning Tests of Extrapolative Funds**: This table shows regressions of quarterly returns of a long-short portfolio of extrapolative funds on the Carhart (1997) momentum factor, the Fama and French (2015) 5 factors, and a long-short portfolio of funds that goes long the 20% of funds with the highest momentum factor loading and short the 20% with the lowest momentum factor loading ($\beta_{FundMOM-LS}$). The long-short portfolio of extrapolative funds goes long the 20% most extrapolative funds and short the 20% least extrapolative funds. The factor exposures and the overreaction metric are measured as in Table 4. Standard errors are in brackets. Significance: ***99%, **95%, *90%.

	Long-Short Extrapolative Funds		
	(1)	(2)	(3)
α	-0.203***	-0.137**	-0.181***
	[0.065]	[0.067]	[0.063]
β_{MOM}	0.064***		0.035**
	[0.010]		[0.015]
$\beta_{FundMOM-LS}$		0.205***	0.122**
		[0.032]	[0.052]
β_{MKT}	0.044***	0.038***	0.042***
	[0.010]	[0.009]	[0.010]
β_{SMB}	-0.061***	-0.044**	-0.049***
	[0.017]	[0.017]	[0.017]
β_{HML}	-0.061***	-0.067***	-0.060***
	[0.020]	[0.019]	[0.018]
β_{RMW}	-0.001	0.050***	0.028
	[0.018]	[0.018]	[0.020]
β_{CMA}	-0.084***	-0.049**	-0.067***
	[0.027]	[0.023]	[0.023]
Observations	114	114	114
Adj. R-squared	0.7241	0.7291	0.7473

Table IA7. **Field Evidence – Additional Performance Metrics**: This table shows the linkage between overreaction and future fund flows and achieving top status. In Column 1, the dependent variable is fund flows, defined as the fund's net dollar flows over the following year scaled by the prior period TNA. In Column 2, we perform a similar regression but with the fund's ascending rank of fund flows by Morningstar style over the following year (i.e. lowest ranked fund has an value of 1). Controls are the same as in Table 4, in addition to lagged CAPM alpha over the previous year (an explanation of this additional control is in Appendix IA2). In Column 3 and 4, the dependent variable is an indicator variable that equals one if the fund ranks in the top 10% or 5% of funds of its Morningstar style category over the following year (based on raw fund returns). Controls are the same as in Table 4. Standard errors are clustered by quarter and at the fund level, and reported in brackets. Significance: ***99%, **95%, *90%.

	Flow	**Flow Rank**	**Top 10% Fund**	**Top 5% Fund**
	(1)	(2)	(3)	(4)
β_j^X $(t-1)$	-3.658***	-6.274***	-0.021***	-0.015***
	[1.271]	[1.753]	[0.007]	[0.005]
Expense Ratio $(t-1)$	-2.035*	-9.503***	0.023***	0.022***
	[1.123]	[2.382]	[0.008]	[0.006]
Log Fund TNA $(t-1)$	-9.225***	-18.678***	-0.022***	-0.014***
	[1.277]	[2.025]	[0.007]	[0.005]
Log Fund TNA^2 $(t-1)$	0.551***	1.349***	0.002***	0.001**
	[0.091]	[0.165]	[0.001]	[0.000]
Log Fund Age $(t-1)$	-3.590***	-4.745***	-0.002	-0.001
	[0.625]	[1.468]	[0.004]	[0.003]
Fund Turnover $(t-1)$	-0.877	-8.444***	0.011*	0.007*
	[0.775]	[1.423]	[0.006]	[0.004]
Log N Stocks $(t-1)$	0.701	3.157***	-0.028***	-0.019***
	[0.457]	[1.025]	[0.003]	[0.002]
Avg. Manager Exp. $(t-1)$	0.299	0.924	0.002	0.002
	[0.315]	[0.625]	[0.002]	[0.001]
β_{MOM} $(t-1)$	-7.97	7.84	-0.134***	-0.094***
	[5.764]	[8.771]	[0.043]	[0.027]
Disposition $(t-1)$	-4.499**	-19.355***	0.023*	0.018*
	[2.168]	[3.999]	[0.013]	[0.009]
CAPM Alpha $(t-1)$	1.654***	2.922***		
	[0.115]	[0.181]		
Flow $(t-1)$			0.000	0.000
			[0.000]	[0.000]
Time × Style FE	Yes	Yes	Yes	Yes
Observations	57632	57632	67634	67634
Adj. R-squared	0.0777	0.4403	0.0124	0.0148

Table IA8. **Field Evidence – Overreaction and Initial Buys**: In this table, we estimate our overreaction metric on initial buys only and repeat the OLS regressions of Table 3 on how overreaction is transmitted from individual team members to teams. We define initial buys as the stocks that are bought by the managers for the first time. To reduce the impact of noise in our estimates, we construct the counterfactual overreaction metric $\hat{\beta}_j^{CF}$ by pooling all observations from the single managers that comprise the team and re-estimate our overreaction metric, rather than taking the average of each individual manager's overreaction metric within the team. Team-level controls include the time-series average log TNA of the team-managed portfolio, the time-series average log experience of the team members, the time-series average log of fund age of the team-managed portfolio, the time-series average of the exposure to momentum of the team, the time-series average of the disposition effect of the team, and team's style dummies. Significance: ***99%, **95%, *90%.

	Initial buys			
	(1)	(2)	(3)	(4)
$\hat{\beta}_j^{CF}$	0.1108	0.1132	0.4345	0.5651*
	[0.1113]	[0.1138]	[0.2802]	[0.2882]
$\hat{\beta}_j^{CF} \times D_j^E$			-0.5955*	-0.8074**
			[0.3280]	[0.3413]
D_j^E			0.146	0.1408
			[0.2257]	[0.2342]
Style fixed effects	No	Yes	No	Yes
Team controls	No	Yes	No	Yes
Observations	307	306	307	306
Adj. R-squared	0	-0.0108	0.0087	0.0064
Hypothesis testing:				
p-value $H_0 : \delta_0 = 1$	0.000	0.000	0.045	0.132
p-value $H_0 : \delta_0 + \delta_1 = 0$			0.346	0.185
p-value $H_0 : \delta_0 + \delta_1 = 1$			0.000	0.000

Table IA9. **Field Evidence – Overreaction and Investor Preferences**: In Panel A, we regress future annual fund return volatility (i.e., the volatility of a fund's monthly benchmark-adjusted returns and FF5 Alpha) on funds' lagged overreaction. In Panel B, we estimate a linear probability model in which the dependent variable is a dummy that is equal to one if a fund ranks among the top 10 or top 5 funds of its Morningstar style in a given year, and the main independent variable is the lagged fund's overreaction. In Panel C, we regress fund investment performance (benchmark-adjusted returns and FF5 alpha) over the following year on lagged funds' overreaction, but restrict the analysis to the years in which the CRSP weighted market index is negative at the time in which the dependent variable is measured. In Panel D, we repeat the analysis of Panel C, but now restrict the sample to years of NBER recessions, defined as years in which 6 or more months are part of a NBER recession. Regressions also control for style-quarter fixed effects and include the fund-level controls of Table 4. Standard errors are clustered by quarter and at the fund level, and reported in brackets. Significance: ***99%, **95%, *90%.

	Panel A: Managerial Risk		Panel B: Extreme Payoffs	
	BM Adj. Vol	**FF5 Alpha Vol**	**Top 10 Fund**	**Top 5 Fund**
	(1)	(2)	(1)	(2)
β_j^X	-0.027	0.005	-0.016***	-0.008***
	[0.029]	[0.027]	[0.004]	[0.003]
Controls	Yes	Yes	Yes	Yes
Time × Style FE	Yes	Yes	Yes	Yes
Observations	54951	57958	67634	67634
Adj. R-squared	0.0753	0.0455	0.047	0.0346
	Panel C: Downside Performance		Panel D: NBER Recessions	
	BM Adj. Return	**FF5 Alpha**	**BM Adj. Return**	**FF5 Alpha**
	(1)	(2)	(3)	(4)
β_j^X	-0.168	-0.463*	-0.474	-0.599
	[0.373]	[0.248]	[0.640]	[0.424]
Controls	Yes	Yes	Yes	Yes
Time × Style FE	Yes	Yes	Yes	Yes
Observations	11726	11723	5321	5321
Adj. R-squared	0.1242	0.256	0.166	0.3055

Table IA10. **Field Evidence – Entering as a Solo versus Team-based Asset Manager**: In Panel A, we show summary statistics for the entry measures as specified in Section IA3.4.1. In Panels B and C, we estimate a double interaction regression of the form: $\hat{\beta}_j^{TM} = \alpha + \delta_0 \hat{\beta}_j^{CF} + \delta_1 \hat{\beta}_j^{CF} \times D_j^E \times R_j + (...) + \delta_7 C_j + \epsilon_j$, where R_j represents alternative interaction terms for our robustness tests. The main coefficient of interest in these regressions is δ_1, which measures if the attenuation of overreaction is stronger or weaker for teams whose members load more on the R_j characteristic. In Panel B, $R_j = D_j^{ES}$, a dummy variable that is equal to one if all members of team j started off as solo managers. In Panel D, $R_j = D_j^{ET}$, a dummy variable that is equal to one if all members of team j started off in team-managed funds. In addition to the reported additional regressors, team-level controls include the time-series average log TNA, the time-series average log experience of the team members, the time-series average log of fund age, the time-series average of the exposure to momentum, the time-series average of the disposition effect, and style fixed effects. Standard errors are in brackets. Significance: ***99%, **95%, *90%.

Panel A: Summary Statistics

All Teams	Obs.	Mean	St. Dev.	Min	Max
D_j^{ES}	308	0.21	0.41	0	1
D_j^{ET}	308	0.18	0.38	0	1
Contrarian Teams	Obs.	Mean	St. Dev.	Min	Max
D_j^{ES}	143	0.21	0.41	0	1
D_j^{ET}	143	0.14	0.35	0	1
Extrapolative Teams	Obs.	Mean	St. Dev.	Min	Max
D_j^{ES}	165	0.21	0.41	0	1
D_j^{ET}	165	0.21	0.41	0	1

	Panel B: Enter Single		Panel C: Enter Team	
	(1)	(2)	(3)	(4)
$\hat{\beta}_j^{CF}$	0.9742***	0.9181***	0.9405***	0.8578***
	[0.1053]	[0.1068]	[0.1159]	[0.1166]
$\hat{\beta}_j^{CF} \times D_j^E \times R_j$	1.0816*	0.7937	0.3102	0.1992
	[0.6227]	[0.6180]	[0.3569]	[0.3518]
$\hat{\beta}_j^{CF} \times D_j^E$	-0.7489***	-0.6709***	-0.7379***	-0.6446***
	[0.1538]	[0.1529]	[0.1717]	[0.1705]
$\hat{\beta}_j^{CF} \times R_j$	-0.8292**	-0.9372***	-0.1319	-0.0591
	[0.3553]	[0.3490]	[0.2434]	[0.2360]
$D_j^E \times R_j$	-0.0467	0.1042	-0.0115	-0.0158
	[0.1488]	[0.1483]	[0.1572]	[0.1552]
D_j^E	0.0463	0.0036	0.0257	0.0015
	[0.0633]	[0.0644]	[0.0626]	[0.0619]
R_j	0.0066	-0.0683	-0.07	-0.0705
	[0.1095]	[0.1085]	[0.1212]	[0.1197]
Style fixed effects	No	Yes	No	Yes
Team controls	No	Yes	No	Yes
Observations	308	307	308	307
Adj. R-squared	0.282	0.312	0.2628	0.2931

Table IA11. **Field Evidence – Alternative Counterfactuals**: In this table we estimate the transmission of overreaction from solo managers to their respective teams using different measures for the counterfactual $\hat{\beta}_j^{CF}$. Whereas in our main specifications $\hat{\beta}_j^{CF}$ is the simple average of overreaction of each individual team member, here we use weighted averages based on each team members' quarters of experience, number of individual funds managed, and size of individual funds managed, all measured at the time of team formation. We estimate the following regression as in Table 3: $\hat{\beta}_j^{TM} = \alpha + \hat{\beta}_j^{CF}(\delta_0 + \delta_1 D_j^E) + \delta_2 D_j^E + \delta_3 C_j + \epsilon_j$. Team controls ($C_j$) are the time-series average log TNA of the team-managed portfolio, the time-series average log experience of the team members, the time-series average log of fund age of the team-managed portfolio, the time-series average of the exposure to momentum of the team, the time-series average of the disposition effect of the team, and style fixed effects, and are included as reported. Column (1), (3) and (5) estimate a simpler model without any interaction term. Thus, in these columns, the coefficient δ_0 captures the transmission of all individual-level return-based trading behavior (be it extrapolative or contrarian) to the team. Column (2), (4) and (6) estimate the full model. In these columns, we analyze separately the transmission of individual contrarian behavior (δ_0), the transmission of extrapolative behavior ($\delta_0 + \delta_1$), and the difference between the two (δ_1). Standard errors are in brackets. Significance: ***99%, **95%, *90%.

	Experience		**Number of Funds**		**Size of Funds**	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}_j^{CF}$	0.4941***	0.8664***	0.5109***	0.8621***	0.4879***	0.8944***
	[0.0640]	[0.1111]	[0.0639]	[0.1051]	[0.0627]	[0.1063]
$\hat{\beta}_j^{CF} \times D_j^E$		-0.6657***		-0.6382***		-0.6762***
		[0.1428]		[0.1402]		[0.1383]
D_j^E		0.0178		-0.0299		-0.0535
		[0.0562]		[0.0575]		[0.0567]
Style fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Team controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	307	307	307	307	307	307
Adj. R-squared	0.2398	0.2895	0.2491	0.2945	0.2419	0.2952
Hypothesis testing:						
p-value $H_0 : \delta_0 = 1$	0.000	0.230	0.000	0.155	0.000	0.322
p-value $H_0 : \delta_0 + \delta_1 = 0$		0.028		0.034		0.016
p-value $H_0 : \delta_0 + \delta_1 = 1$		0.000		0.000		0.000

Table IA12. **Field Evidence – Style, Workload, and Fund Families**: In Panel A, we show summary statistics for the style, workload, and fund family measures as specified in Sections IA3.5.3 – IA3.5.4. In Panels B to D, we estimate a double interaction regression of the form: $\hat{\beta}_j^{TM} = \alpha + \delta_0 \hat{\beta}_j^{CF} + \delta_1 \hat{\beta}_j^{CF} \times D_j^E \times R_j + (...) + \delta_7 C_j + \epsilon_j$, where R_j represents alternative interaction terms for our robustness tests. The main coefficient of interest in these regressions is δ_1, which measures if the attenuation of overreaction is stronger or weaker for teams whose members load more on the R_j characteristic. In Panel B, $R_j = D_j^{SM}$, a dummy variable that is equal to one if a style migration occurs, i.e., the style classification of team j does not match the prevailing style classification of the funds managed individually by the members of the team. In Panel C, $R_j = \Delta Workload_j$, i.e., the difference in workload faced by the managers of team j when managing as part of that team, and the workload these same managers face when managing a fund alone (further details are in Section IA3.5.2). In Panel D, $R_j = D_j^{FS}$, a dummy variable that is equal to one if the mutual fund family of team j does not match with the mutual fund family under which managers operate when solo (i.e. a family switch). In addition to the reported additional regressors, team-level controls include the time-series average log TNA, the time-series average log experience of the team members, the time-series average log of fund age, the time-series average of the exposure to momentum, the time-series average of the disposition effect, and style fixed effects. Standard errors are in brackets. Significance: ***99%, **95%, *90%.

Panel A: Summary Statistics

All Teams	Obs.	Mean	St. Dev.	Min	Max
D_j^{SM}	308	0.24	0.43	0	1
$\Delta Workload_j$	308	-0.71	2.28	-23	8
D_j^{FS}	308	0.40	0.49	0	1
Contrarian Teams	Obs.	Mean	St. Dev.	Min	Max
D_j^{SM}	143	0.20	0.40	0	1
$\Delta Workload_j$	143	-0.78	1.93	-12	3
D_j^{FS}	143	0.41	0.49	0	1
Extrapolative Teams	Obs.	Mean	St. Dev.	Min	Max
D_j^{SM}	165	0.27	0.44	0	1
$\Delta Workload_j$	165	-0.65	2.56	-23	8
D_j^{FS}	165	0.39	0.49	0	1

	Panel B: Style Migrations		Panel C: Workload		Panel D: Mutual Fund Family	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}_j^{CF}$	0.9601***	0.8982***	0.8277***	0.7825***	0.8303***	0.6614***
	[0.1046]	[0.1061]	[0.1578]	[0.1576]	[0.2080]	[0.2074]
$\hat{\beta}_j^{CF} \times D_j^E \times R_j$	0.2146	0.3279	0.0557	0.0253	0.1117	-0.117
	[0.5480]	[0.5478]	[0.1308]	[0.1287]	[0.3369]	[0.3388]
$\hat{\beta}_j^{CF} \times D_j^E$	-0.6769***	-0.6263***	-0.5684***	-0.5193***	-0.6655***	-0.4671*
	[0.1543]	[0.1530]	[0.1948]	[0.1927]	[0.2468]	[0.2481]
$\hat{\beta}_j^{CF} \times R_j$	-0.7110*	-0.7600**	0.0557	-0.065	0.12	0.2525
	[0.3673]	[0.3635]	[0.1308]	[0.1163]	[0.2389]	[0.2391]
$D_j^E \times R_j$	0.0766	0.0744	-0.0829	0.0081	-0.0825	-0.0812
	[0.1474]	[0.1455]	[0.1187]	[0.0319]	[0.1198]	[0.1169]
D_j^E	0.0261	0.0104	0.0068	-0.0058	0.0569	0.0406
	[0.0637]	[0.0642]	[0.0329]	[0.0284]	[0.0775]	[0.0765]
R_j	0.0513	-0.008	-0.01	-0.0389	0.0788	0.1129
	[0.1132]	[0.1136]	[0.0292]	[0.1480]	[0.0888]	[0.0869]
Style fixed effects	No	Yes	No	Yes	No	Yes
Team controls	No	Yes	No	Yes	No	Yes
Observations	308	307	308	307	308	307
Adj. R-squared	0.285	0.3099	0.2627	0.2929	0.265	0.2974

Table IA13. **Field Evidence – Stocks Overseen as Proxy for Workload**: In this table we perform the same analysis as in Table IA12 for our workload measure $Workload_j$: $\hat{\beta}_j^{TM} = \alpha + \delta_0 \hat{\beta}_j^{CF} + \delta_1 \hat{\beta}_j^{CF} \times D_j^E \times Workload_j + (...) + \delta_7 C_j + \epsilon_j$, where we measure the change in workload as the difference in workload faced by managers of team j when managing as part of that team, and the workload these same managers face when managing a fund alone. Workload is proxied by the total number of stocks that are overseen by the managers when they work in teams versus when they work individually, see Internet Appendix **??** for details. The main coefficient of interest is δ_1, which measures if the attenuation of overreaction is stronger or weaker for teams whose members experience a larger difference in workload when they manage as part of a team versus in their solo-managed funds. In addition to the reported additional regressors, team-level controls include the time-series average log TNA, the time-series average log experience of the team members, the time-series average log of fund age, the time-series average of the exposure to momentum, the time-series average of the disposition effect, and style fixed effects. Standard errors are in brackets. Significance: ***99%, **95%, *90%.

	Workload	
	(1)	(2)
$\hat{\beta}_j^{CF}$	0.9576***	0.8955***
	[0.1000]	[0.1024]
$\hat{\beta}_j^{CF} \times D_j^E \times Workload_j$	0.0089**	0.0069
	[0.0038]	[0.0047]
$\hat{\beta}_j^{CF} \times D_j^E$	-0.7848***	-0.7242***
	[0.1625]	[0.1619]
$\hat{\beta}_j^{CF} \times Workload_j$	-0.0113***	-0.0092***
	[0.0029]	[0.0029]
$D_j^E \times Workload_j$	0.0006	0.0003
	[0.0006]	[0.0006]
D_j^E	-0.0138	-0.0333
	[0.0602]	[0.0608]
$Workload_j$	-0.0005	-0.0003
	[0.0005]	[0.0005]
Style fixed effects	No	Yes
Team controls	No	Yes
Observations	308	307
Adj. R-squared	0.2627	0.2929

Table IA14. **Field Evidence – Transmission of Overreaction to Teams IV Simulations:** In this table we present the OLS and IV results from 1,000 independent simulations. Panel A (B) shows the results without (with) the interaction terms. In the first column we report the true parameters, whereas in the preceding columns we present the estimated parameters using OLS and IV procedures respectively for three different sets of parameters, σ_β and σ_e. We report average estimates, the standard deviations and average t-statistics over the 1,000 simulations. For the δ_0 parameter, we also provide the percentage of times we reject the null, H_0: $\delta_0 = 1$ at a 95% significance level.

Panel A: Team Effect

		$\sigma_\beta = 0.5$, $\sigma_e = 1$		$\sigma_\beta = 0.5$, $\sigma_e = 2$		$\sigma_\beta = 0.2$, $\sigma_e = 1$	
	True Coefficient	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
Avg. (α)	0	-0.001	0.000	-0.003	0.000	-0.004	0.000
Std. (α)		0.007	0.008	0.010	0.011	0.008	0.008
Avg. (T_α)		-0.094	0.003	-0.252	0.004	-0.527	-0.001
Avg. (δ_0)	1	0.986	1.000	0.947	1.001	0.920	1.001
Std. (δ_0)		0.021	0.022	0.033	0.034	0.055	0.059
Avg. (T_{δ_0})		49.513	46.618	34.664	30.444	19.213	17.457
% Rejected H_0: $\delta_0 = 1$		0.175	0.053	0.586	0.054	0.499	0.057

Panel B: Interaction Terms

		$\sigma_\beta = 0.5$, $\sigma_e = 1$		$\sigma_\beta = 0.5$, $\sigma_e = 2$		$\sigma_\beta = 0.2$, $\sigma_e = 1$	
	True Coefficient	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
Avg. (α)	0	-0.001	0.000	-0.003	0.002	-0.004	0.002
Std. (α)		0.009	0.012	0.013	0.020	0.010	0.014
Avg. (T_α)		-0.073	0.003	-0.196	0.106	-0.412	0.100
Avg. (δ_0)	1	0.986	1.000	0.947	1.005	0.920	1.009
Std. (δ_0)		0.028	0.034	0.042	0.056	0.068	0.091
Avg. (T_{δ_0})		35.927	30.284	25.137	18.833	14.732	11.277
% Rejected H_0: $\delta_0 = 1$		0.131	0.052	0.387	0.041	0.357	0.042
Avg. (δ_1)	0	0.000	0.001	-0.001	0.032	-0.001	0.079
Std. (δ_1)		0.081	0.112	0.112	0.194	0.229	0.434
Avg. (T_{δ_1})		-0.004	0.004	-0.006	0.125	-0.004	0.143
Avg. (δ_2)	0	0.000	0.000	0.000	-0.023	0.000	-0.022
Std. (δ_2)		0.034	0.058	0.048	0.111	0.039	0.094
Avg. (T_{δ_2})		0.005	-0.008	0.007	-0.175	0.004	-0.201

Table IA15. **Field Evidence – Back of the Envelope**: This table shows back-of-the-envelope statistics concerning the impact of overreaction on fund returns and flows. More details are in Section IA5.

Panel A: Average yearly reduction in labour income

Years of work	Increase in Overreaction 0.1	0.4	0.6	0.8
30	0.72%	2.86%	4.26%	5.65%
38	0.90%	3.57%	5.31%	7.02%
45	1.06%	4.19%	6.21%	8.20%

Panel B: Cumulative reduction in labour income relative to yearly salary

Years of work	Increase in Overreaction 0.1	0.4	0.6	0.8
30	0.22	0.86	1.32	1.78
38	0.34	1.37	2.09	2.80
45	0.48	1.90	2.89	3.88

Panel C: PV cumulative reduction in labour income as a fraction of PV total income (risk-free rate)

Years of work	Increase in Overreaction 0.1	0.4	0.6	0.8
30	0.49%	1.96%	2.93%	3.88%
38	0.67%	2.67%	3.97%	5.26%
45	0.82%	3.25%	4.84%	6.40%

Panel D: PV cumulative reduction in labour income as a fraction of PV total income (market rate)

Years of work	Increase in Overreaction 0.1	0.4	0.6	0.8
30	0.41%	1.62%	2.43%	3.22%
38	0.46%	1.81%	2.71%	3.60%
45	0.48%	1.91%	2.86%	3.79%

Table IA16. **Experiment – Compensation, Forecast Errors and Forecast Dispersion**: This table presents descriptive statistics for several variables collected in the experiment for the Individual, Internal Reflection, Self Selection, and Group treatments. We report means, medians, standard deviations, and number of observations of the bonus compensation earned for the forecast accuracy (in dollars), the mean absolute error (the absolute value of the difference between the prediction and the return realization), the mean squared error (the squared difference between the prediction and the return realization), and the prediction.

Variable	Mean	Median	Std. Dev.	Obs.
Individual				
$ Compensation	0.910	0.900	0.245	248
Mean Absolute Error	22.757	22.075	4.972	248
Mean Squared Error	829.703	742.675	396.046	248
Prediction	3.664	2.000	20.205	4960
Internal Reflection				
$ Compensation	0.939	0.940	0.239	405
Mean Absolute Error	22.310	21.550	5.058	405
Mean Squared Error	804.468	710.750	414.257	405
Prediction	4.353	3.000	20.094	8,100
Self-Selection				
$ Compensation	0.879	0.880	0.282	403
Mean Absolute Error	24.628	23.100	7.942	403
Mean Squared Error	1025.241	804.450	757.805	403
Prediction	4.990	4.000	25.121	8060
Group				
$ Compensation	0.958	0.950	0.247	456
Mean Absolute Error	20.812	20.450	3.669	456
Mean Squared Error	673.239	634.150	233.314	456
Prediction	4.217	2.000	17.007	9,120

Table IA17. **Experiment – Using Alternative Overreaction Metrics**: In this table we present results from regressions of the overreaction coefficient on dummy variables that equal one if participants are assigned to the Group (G), Internal Reflection (IR), or Self Selection (SS) treatments, and zero otherwise. The dependent variable is the alternative overreaction coefficient estimated from Equation (IA35). We also report the differences $\widehat{\gamma_G}-\widehat{\gamma_{IR}}$, $\widehat{\gamma_{SS}}-\widehat{\gamma_{IR}}$, and $\widehat{\gamma_G}-\widehat{\gamma_{SS}}$, with corresponding standard errors, where $\widehat{\gamma_j}$ is the coefficient associated to the treatment dummy $j = G, IR, SS$. In the two bottom rows, we report summary statistics for the dependent variable for participants in the Individual treatment. Column 2 includes demographics controls, namely ethnicity dummies; 1-to-7 measures of trust in others, patience, and risk tolerance; age; a college graduate dummy; an employed dummy; a 1-to-6 measure of political opinions, ranging from 1 for very conservative to 6 for very liberal; Biden or Trump dummies for the favorite presidential candidate; a single dummy; the number of people living in the household. In Column 3 we further include financial sophistication controls, namely dummies equal to one if the participant holds any stocks, has traded in stocks in the previous year, or holds shares in a mutual fund; a 0-to-4 score assigned on the basis of a cognitive reflection test; a 0-to-3 score assigned on the basis of a financial literacy test. In Column 4, we include US region and income category fixed effects. Standard errors are in brackets and are clustered at the team level (which corresponds to the single participant in the case of the Individual treatment). Significance: ***99%, **95%, *90%.

	(1)	(2)	(3)	(4)
Group	-0.214***	-0.227***	-0.228***	-0.227***
	[0.064]	[0.064]	[0.064]	[0.064]
Internal Reflection	-0.068	-0.062	-0.066	-0.060
	[0.060]	[0.061]	[0.061]	[0.062]
Self-Selection	-0.195***	-0.177**	-0.184**	-0.176**
	[0.073]	[0.072]	[0.072]	[0.072]
Controls:				
Demographics	No	Yes	Yes	Yes
Fin. Sophistication	No	No	Yes	Yes
Region and Income FE	No	No	No	Yes
Observations	1,512	1,512	1,512	1,512
R-squared	0.013	0.032	0.035	0.051
Avg. $\tilde{\beta}_w$ Individual	0.411	0.411	0.411	0.411
Std. Dev. $\tilde{\beta}_w$ Individual	0.746	0.746	0.746	0.746
$\widehat{\gamma_G}-\widehat{\gamma_{IR}}$	-0.147***	-0.164***	-0.161***	-0.167***
$se(\widehat{\gamma_G}-\widehat{\gamma_{IR}})$	[0.056]	[0.057]	[0.057]	[0.057]
$\widehat{\gamma_{SS}}-\widehat{\gamma_{IR}}$	-0.127*	-0.115*	-0.118*	-0.116*
$se(\widehat{\gamma_{SS}}-\widehat{\gamma_{IR}})$	[0.067]	[0.066]	[0.066]	[0.066]
$\widehat{\gamma_G}-\widehat{\gamma_{SS}}$	-0.020	-0.049	-0.043	-0.051
$se(\widehat{\gamma_G}-\widehat{\gamma_{SS}})$	[0.070]	[0.069]	[0.069]	[0.068]

Table IA18. **Experiment – The Moderating Role of External Screening for Team Forecasting**: In this table we present results regarding the communication and self selection between team members assigned to the "Group" treatment. In the first two columns, the dependent variable is "repeat first guess," a dummy equal to one if the participants who make the first guess in round t is the same who made it in round $t-1$. The key regressor is the absolute error of their first prediction at $t-1$ (divided by the sample standard deviation). In Columns 3 and 4, the dependent variable is "challenged first guess," a dummy equal to one if either the final prediction is different from the first guess or the participant that did not make the first guess proposes at least one alternative prediction. The key coefficient of interest is the absolute value of the first guess (divided by the sample standard deviation). In the two bottom rows, we report summary statistics for the dependent variable. The regressions in Columns 2 and 4 include the same demographics controls, financial sophistication controls, and region and income category fixed effects used in other tests in the paper. Standard errors are in brackets and are clustered at the team level. Significance: ***99%, **95%, *90%.

	Repeat first guess		Challenged first guess	
	(1)	(2)	(3)	(4)
$\lvert\text{FcastErrorFG}\rvert_{i,t-1}$	-0.061***	-0.062***		
	[0.009]	[0.009]		
$\lvert\text{FirstGuess}\rvert_{i,t}$			0.024***	0.025***
			[0.008]	[0.007]
Observations	4281	4281	4504	4504
R^2	0.013	0.0479	0.0032	0.026
Controls:				
Demographics	No	Yes	No	Yes
Fin. Sophistication	No	Yes	No	Yes
Region and Income FE	No	Yes	No	Yes
Avg. Dep. Var.	0.438	0.438	0.180	0.180
Std. Dev. Dep. Var.	0.496	0.496	0.384	0.384

References Internet Appendix

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*(1), 259–299.

Barber, B. M., X. Huang, and T. Odean (2016). Which factors matter to investors? evidence from mutual fund flows. *Review of Financial Studies 29*(10), 2600–2642.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 1, pp. 79–175. Elsevier.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2018). Extrapolation and bubbles. *Journal of Financial Economics 129*(2), 203–227.

Barberis, N. and M. Huang (2001). Mental accounting, loss aversion, and individual stock returns. *Journal of Finance 56*(4), 1247–1292.

Barberis, N. and M. Huang (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review 98*(5), 2066–2100.

Berk, J. B. and R. C. Green (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy 112*(6), 1269–1295.

Berk, J. B. and J. H. van Binsbergen (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics 118*(1), 1–20.

Berk, J. B. and J. H. Van Binsbergen (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics 119*(1), 1–23.

Boyer, B., T. Mitton, and K. Vorkink (2010). Expected idiosyncratic skewness. *Review of Financial Studies 23*(1), 169–202.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy 107*(2), 205–251.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Chen, A. Y. and T. Zimmermann (2022). Open source cross-sectional asset pricing. *Critical Finance Review 27*(2), 207–264.

Chevalier, J. and G. Ellison (1999). Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance. *Journal of Finance 54*(3), 875–899.

Cici, G. (2012). The prevalence of the disposition effect in mutual funds' trades. *Journal of Financial and Quantitative Analysis 47*(4), 795–820.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *Journal of Finance 63*(4), 1609–1651.

Coval, J. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*(2), 479–512.

Cremers, K. M. and A. Petajisto (2009). How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies 22*(9), 3329–3365.

Da, Z., X. Huang, and L. J. Jin (2021). Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics 140*(1), 175–196.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies'. *Journal of Finance 51*, 55–84.

Fama, E. F. and K. R. French (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance 65*(5), 1915–1947.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Golec, J. H. (1996). The effects of mutual fund managers' characteristics on their portfolio performance, risk and fees. *Financial Services Review 5*(2), 133–147.

Gompers, P. A. and A. Metrick (2001). Institutional investors and equity prices. *Quarterly Journal of Economics 116*(1), 229–259.

Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *Review of Financial Studies 27*(3), 714–746.

Guercio, D. D. and P. A. Tkac (2008). Star power: The effect of morningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 907–936.

Guvenen, F., S. Ozkan, and J. Song (2014). The nature of countercyclical income risk. *Journal of Political Economy 122*(3), 621–660.

Ibert, M., R. Kaniel, S. Van Nieuwerburgh, and R. Vestman (2018). Are mutual fund managers paid for investment skill? *Review of Financial Studies 31*(2), 715–772.

Jegadeesh, N., J. Noh, K. Pukthuanthong, R. Roll, and J. Wang (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics 133*(2), 273–298.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jiang, Z., H. Liu, C. Peng, and H. Yan (2022). Investor memory and biased beliefs: Evidence from the field. *Available at SSRN*.

Kim, M. (2020). Fund flows, liquidity, and asset prices. *Working Paper*.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics 105*(1), 1–28.

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies 25*(12), 3457–3489.

Ma, L., Y. Tang, and J.-P. Gómez (2019). Portfolio manager compensation in the us mutual fund industry. *Journal of Finance 74*(2), 587–638.

McLemore, P. (2019). Do mutual funds have decreasing returns to scale? Evidence from fund mergers. *Journal of Financial and Quantitative Analysis 54*(4), 1683–1711.

Nelson, C. and R. Startz (1990). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *The Journal of Business 63*(1), S125–S140.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance 53*(5), 1775–1798.

Pástor, L., R. F. Stambaugh, and L. A. Taylor (2015). Scale and skill in active management. *Journal of Financial Economics 116*(1), 23–45.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Chapter 5. Cambridge: Cambridge Univ. Press.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *Journal of Finance 55*(4), 1655–1695.