

DO GROUPS MAKE WISE DECISIONS?

Daniel Evans Georg Schneider

Version: October 26, 2026

[Click here for the latest version](#)

Abstract

Many important decisions are made jointly by groups. While groups can draw on and discuss diverse perspectives, they may also suffer from counterproductive dynamics, such as herding and gridlock. To investigate this tension, we study three features that distinguish joint from individual decision-making: *composition*, *communication*, and *collaborative aggregation*. Simple comparisons between joint and individual decisions cannot distinguish their effects. To separate them, we run an interactive online experiment in which subjects construct portfolios to maximize returns and minimize volatility in one of several randomly assigned decision modes; each mode differs from adjacent modes by a single feature. We find that *composition* drives the benefits of joint decision-making: simply pooling two individuals' independent submissions together reduces portfolio concentration and volatility. *Communication* attenuates these gains by inducing partners to converge on similar portfolios and to prioritize maximizing returns over managing risk. *Collaborative aggregation* has little effect on average, but improves diversification when partners hold differing beliefs. Disentangling these features reveals an offsetting cost of communication that limits the ability of groups to make full use of members' inputs.

JEL No. C92, D70, D91, G41, M54

Contact: Daniel Evans; daniel.evans@uni-bonn.de. Georg Schneider; georg.schneider@uni-bonn.de.

Acknowledgements: We thank Olga Balakina, Anke Becker, Thomas Dohmen, Benjamin Enke, Paul Grass, Luca Henkel, Tomáš Jagelka, Milena Jessen, Valentin Kecht, John List, Luca Michels, Farzad Saidi, Amelie Schiprowski, Philipp Schirmer, Matthias Sutter, Séverine Toussaert, and Florian Zimmermann for valuable discussion and comments.

Funding: Daniel Evans is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/2 - 390838866.

Ethics approval: The study was approved under the Gesellschaft für experimentelle Wirtschaftsforschung (GfEW) Institutional Review Board certificate no. gE1F4rmG (<https://gfew.de/ethik/gE1F4rmG>).

Pre-registration: This study was preregistered with the AEA RCT Registry under ID AEARCTR-0018014 (<https://doi.org/10.1257/rct.18014-1.1>).

1 Introduction

Many consequential economic decisions, including those of hiring committees, corporate boards, and mutual-fund management teams, are made jointly by groups rather than by single individuals. Groups are entrusted with these decisions because they can draw on, deliberate about, and aggregate a diverse array of information and judgments. However, these features can also impair decision quality: group members can hold redundant or conflicting inputs, deliberation can induce herding, and aggregation can amplify the influence of overconfident members. Because joint decision-making differs from individual decision-making in all of these respects at once, it is difficult to determine which features enhance decision quality and which harm it. We address this challenge with an experiment.

We characterize joint decision-making as a process in which a group communicates and reconciles individual judgments into a shared decision. This definition highlights *composition*, *communication*, and *collaborative aggregation* as three features that distinguish it from individual decision-making. Composition refers both to the size of a decision unit (a group as opposed to an individual) and to the mix of people in it. Communication concerns the exchange and discussion of inputs during group work. Finally, collaborative aggregation refers to the requirement that members agree on a single joint decision.

Our experiment separates the effects of each feature within a common decision environment. Subjects construct portfolios to maximize returns and limit volatility over a known evaluation window. We find that joint decision-making produces more diversified and less volatile portfolios than individual decision-making. This advantage arises even without interaction: the compositional effect of pooling two individuals' independent submissions together exceeds the total advantage of joint decision-making. However, communication partly offsets these gains. We detect no significant average effect of collaborative, unanimity-based aggregation relative to simple pooling, but it improves diversification when group members hold differing beliefs. Our results serve as a proof of concept that a decomposition of joint decision-making into its component features reveals offsetting effects that simple comparisons would miss. In particular, a naive comparison between joint and individual decisions that ignored composition would erroneously attribute a positive effect to communication.

Isolating these effects is typically infeasible because researchers rarely observe comparable decisions made with exogenous variation in each feature. Previous studies have advanced our understanding of joint decision-making in a range of contexts, often focusing on one feature while bundling two or more together (Blinder and Morgan, 2005; Charness and Sutter, 2012; Mellers et al., 2014; Atanasov et al., 2017; Radbruch and Schiprowski, 2023). Effects that are attributed to communication may therefore also reflect composition and collaborative aggregation. Designs that compare judgments elicited before and after group discussion potentially also confound the effect of discussion with learning and changes in effort.

A design that avoids these limitations must satisfy three criteria. First, it must elicit

comparable decisions made in decision modes ranging from individual to joint decision-making and allow for the mechanical aggregation of individual decisions into group decisions. The relevant modes include individual decision-making with no communication, individual decision-making with communication, and joint decision-making with communication and collaborative aggregation. Comparing independent individual decisions with their mechanically aggregated group decision identifies the compositional effects of changing the size of the decision unit and combining members' inputs. Comparing aggregated decisions from communicating and non-communicating groups estimates the effect of communication. Finally, the comparison of mechanically aggregated decisions from communicating groups with those made jointly by group members isolates collaborative aggregation.

Second, the design must assign decision modes and partners exogenously, with the nature and timing of the task held fixed. This prevents selection into modes and partners and ensures that estimated effects are not confounded by differences in prior task experience.

Third, interpreting between-mode differences in outcomes as effects on decision quality requires objective or theoretically grounded performance benchmarks. Additional desirable properties include scope for each feature to affect performance, the economic relevance of the decision problem, and access to rich process and choice data for studying mechanisms.

We now summarize our experimental design and explain how it satisfies the above criteria. In our setup, pairs of subjects construct portfolios of stocks from the S&P 500 in one of three randomly assigned decision modes; adjacent modes differ by one feature of interest. In (i) *Crowds*, there is no interaction between partners and each individual submits a portfolio that is mechanically pooled with their partner's. We use equal-weight pooling as the benchmark aggregation rule for estimating compositional effects because it is transparent and treats partners' inputs symmetrically. In (ii) *Panels*, partners communicate but still submit individual portfolios that are pooled using the same rule. Finally, in (iii) *Teams*, partners communicate and must agree on one joint portfolio. Comparisons within and between these modes isolate the effects of each feature as required by the first criterion.

An additional treatment, which we cross-randomize with decision mode, allows us to examine whether compositional effects depend on the combination of inputs within a pair. In particular, we assign each subject a partner whose thinking style is either similar to or different from their own. Partner assignment also generates within-pair variation in stock preferences and beliefs about the stock market. With the nature and timing of the task held fixed, our exogenous assignment of decision mode and partner satisfies the second criterion.

To evaluate performance across decision modes and pairings, we incentivize subjects to maximize the Sharpe ratios of their portfolios over three months. This metric is among the most widely used measures of risk-adjusted investment performance (Eling and Schuhmacher, 2007). To motivate effort across the entire performance distribution and reduce the appeal of extreme investing strategies, bonus payments increase progressively with a subject's performance ranking but are capped at \$10. The Sharpe ratio thus serves as the

objective performance measure required by the third criterion. Benchmarks of interest include the performance of the S&P 500 and of portfolios constructed by LLMs.

Our design also satisfies the additional desirable properties stated above. Compared with tasks featuring a narrow choice set or a readily verifiable correct answer, portfolio allocation offers more scope for heterogeneous information, beliefs, and approaches to affect performance. Prior literature has identified competing effects of each feature on decision quality.

Composition can affect decision quality directly, by broadening or narrowing the range of inputs available to a group, and indirectly, by changing how group members interact. The literature on the wisdom of the crowd (WoC) documents gains in performance from combining diverse, independent inputs into collective judgments, primarily in forecasting and estimation tasks (Surowiecki, 2005; Atanasov et al., 2017). More broadly, diverse groups can draw on a wider array of information and approaches (Hong and Page, 2004). However, differences between members can also create frictions that impede their ability to make effective use of these inputs (Lyons, 2017; Evans et al., 2025).

Communication can likewise improve or impair decision quality. While it enables the sharing of private information and the critical evaluation of suggested approaches (Cooper and Kagel, 2005; Goeree and Yariv, 2011), members may not make full use of the information from partners (Conlon et al., 2026). Moreover, communication can undermine the diversity of thought and reinforce shared biases (Lorenz et al., 2011; Bénabou, 2013).

The effect of collaborative aggregation is similarly ambiguous. It can improve upon fixed aggregation rules by allocating more weight to the views of sophisticated members (Tetlock and Gardner, 2016; Chan, 2021), but it can also be harmful if domineering and overconfident members exert the most influence (Anderson and Kilduff, 2009; Enke et al., 2023).

We develop a conceptual framework that models these trade-offs, motivates our empirical outcomes, and poses testable predictions. The first prediction is that portfolio pooling improves certain components of portfolio quality even without interaction. It further predicts that communication increases the similarity of partners' beliefs and preferred portfolios, and that this convergence can be beneficial if improvements in individual portfolio quality offset any reductions in portfolio diversification. Finally, it identifies a similar channel through which collaborative aggregation can improve decision quality. However, this effect is limited when partners' preferred decisions have already converged through communication.

Regarding economic relevance, the task resembles a real decision faced by retail investors, namely constructing portfolios from a large choice set under considerable uncertainty. This differs from common tasks in experimental and behavioral finance, including trading in experimental asset markets (Palan, 2013), forecasting asset returns (Barahona et al., 2026), and allocating funds between risky and safe assets (Imas, 2016). The realism of our task may also increase engagement among subjects interested in investing, including those drawn from the large online population of retail investors (Cookson et al., 2024).

The design also generates rich process and choice data. We elicit subjects' individually

preferred portfolios in an initial round before assignment to a group decision mode. These initial portfolios and the size of the choice set allow us to distinguish assets preferred by only one member from those preferred by both members. We can then measure influence and dominance by tracking which member's unique holdings are represented in the final group portfolio. Additionally, changes in portfolio similarity between rounds quantify convergence within each pair. These portfolio measures are richer than those from tasks with a binary or unidimensional choice set, which can obscure the unique contributions of each group member. We also collect chat logs from *Panels* and *Teams* for text analysis.

Under positive excess returns, subjects could improve their Sharpe ratio both by increasing returns and reducing volatility. Their choices and self-reports suggest that they were attentive to both components; we therefore study each one separately. Our preregistered expectation was that group structure would primarily affect the portfolio risk component, as captured by realized volatility and diversification. We also anticipated the competing behavioral channels through which each feature could reduce or increase portfolio risk.

We first compare the portfolios of *Teams* and individuals in order to estimate the combined effects of all three features. We find that *Teams* outperform individuals on several measures of portfolio risk. In particular, *Team* pairs are 19 percentage points more likely to reduce volatility relative to the average across their members' initially preferred portfolios. Their portfolios are also 0.50 standard deviations (SDs) less concentrated as measured by the Herfindahl–Hirschman Index (HHI) and 0.23 SDs less volatile than those of individuals, although the latter difference is significant only at the 10% level. Since realized volatility is sensitive to outliers and noise in return realizations, we report effects on multiple portfolio risk outcomes for each comparison. Consistent with our preregistered expectation, we detect no systematic differences in realized returns between decision modes.

To estimate the contribution of each feature, we compare the decision modes as outlined above. We identify a strong compositional effect of expanding the decision unit from one to two members and pooling their inputs. Compared to unpooled *Crowd* portfolios, pooled portfolios are 39 percentage points more likely to reduce volatility relative to the average across their members' initially preferred portfolios. They are also 0.89 SDs less concentrated and have 0.46 SDs lower realized volatility.

Next, we show that communication exacerbates portfolio risk. *Panel* pairs are 16 percentage points less likely than *Crowd* pairs to reduce volatility relative to the average across their members' initially preferred portfolios. The average *Panel* portfolio is also 0.43 SDs more concentrated and directionally more volatile.

Communication also changes subjects' behavior. In particular, relative to *Crowd* members, *Panel* members adopt portfolios that are much more similar to their partners' portfolios. This reflects *Panel* subjects replacing some assets and sectors that were unique to their initial portfolio with others that were held by their partner. They also increase portfolio weights on a handful of "winner stocks" with high historical returns.

Taken together, these findings are consistent with communication shifting subjects' strategies away from diversifying their portfolios toward selecting a narrower set of stocks for which they expect higher returns. Subjects' self-reports mirror this shift, with *Panel* subjects reporting greater attention to returns relative to risk than *Crowd* subjects. Their chat logs also reveal a heightened focus on assets with high historical returns.

We detect no significant effect of collaborative aggregation on portfolio returns, volatility, diversification, or attention to returns. In particular, its estimated effects on realized volatility and concentration are small and point in opposite directions for portfolio risk.

Consistent with the adverse effect of communication and the small estimated effect of collaborative aggregation, we find that *Teams* underperform *Crowds* on portfolio risk. Compared to *Team* portfolios, pooled *Crowd* portfolios are 21 percentage points more likely to reduce volatility relative to the average across their members' initially preferred portfolios. They are also 0.40 SDs less concentrated and are 0.22 SDs less volatile.

Our main conclusions remain robust to extending the evaluation window by nearly two months and winsorizing outcomes at the 1st/99th and 5th/95th percentiles. The precision of some individual estimates varies between specifications.

The underperformance of *Teams* relative to *Crowds* is surprising for three reasons. First, *Teams* could readily replicate the aggregation procedure used by *Crowds* by pooling their members' initially preferred portfolios. Second, subjects' individual choices suggest that they understood the rationale behind diversification. The failure of *Teams* to diversify at least as well as *Crowds* is consistent with a counterproductive shift in strategy induced by communication. Third, *Teams* could use communication and collaborative aggregation to reduce portfolio risk beyond the pooling solution. Although *Teams* outperform individuals working alone, they fail to take full advantage of their members' diverse inputs.

While increasing the intensity of interaction between partners has small or negative effects on average, their size and direction could vary with pair composition. Our primary focus is on within-pair variation in partners' cognitive inputs, including their initially preferred stocks, stock market beliefs, and thinking styles. In principle, cognitive diversity could amplify the diversification benefits of pooling. Whether it improves the quality of discussion and joint decisions is ambiguous, since differences in inputs and approaches may reduce the risk of groupthink while increasing that of conflict and miscommunication.

We find mixed support for the hypothesis that cognitive diversity matters for group performance. Pairs with lower overlap in initially preferred portfolios see larger gains from mechanical pooling. However, their subsequent choices largely offset these gains. Next, pairing subjects with differing stock market beliefs together is associated with lower portfolio risk. This association is strongest among *Teams*, consistent with disagreement being especially valuable when members must reconcile competing views into a single joint decision. Finally, within-pair differences in thinking styles have no detectable effect, suggesting that cognitive composition matters most in our setting when it captures domain-specific differences.

These findings relate the literature on the wisdom of the crowd to portfolio allocation. Our task resembles prior WoC tasks in two key respects. First, as with forecasting, the performance of individual and group decisions depends on uncertain future outcomes. Second, as with forecasting and estimation, combining diverse and independent inputs can improve decision quality even in the absence of communication. In our task, aggregation reduces idiosyncratic risk by broadening portfolios (Markowitz, 1952). In forecasting and estimation, it reduces idiosyncratic error. For both tasks, the gains from aggregation therefore depend on the diversity of the underlying inputs. Very similar portfolios or judgments leave little scope to reduce either idiosyncratic risk or idiosyncratic error, which motivates the emphasis on cognitive diversity in the literature on the WoC (Surowiecki, 2005).

We also contribute to the broader literature on diversity by using experimental methods to study three distinct dimensions of cognitive diversity. Prior studies rely on observational variation (Kang et al., 2022) or examine sociodemographic diversity (Hoogendoorn et al., 2013; Calder-Wang and Gompers, 2021). Studying cognitive diversity directly is important because sociodemographic differences are imperfect proxies for it (Kilduff et al., 2000).

Our findings additionally speak to the literature on individual and group financial decision-making. Previous studies have compared the performance of individually managed and team-managed funds (Bär et al., 2011; Adams et al., 2018), examined how teams and team composition affect bubble formation (Cheung and Palan, 2012; Levine et al., 2014), and investigated whether teams display biases such as the disposition effect (Rau, 2015) and the extrapolation of past performance (Barahona et al., 2026). These questions matter in practice because many investment funds are run by teams (Alternative Investment Management Association, 2019). While comparisons of individually managed and team-managed funds typically rely on observational variation, we use an experiment to identify the causal effect of the decision mode and provide evidence on mechanisms. The downside is that using retail investors as subjects limits the generalizability of our findings to professional investors.

Finally, we relate our findings to the household finance literature, where prior work has documented that retail investors tend to under-diversify and to neglect correlations between assets (Campbell, 2006; Goetzmann and Kumar, 2008). We partly replicate these patterns in our experimental data. More directly relevant to our setting, existing research on joint decision-making among couples documents that the traits and bargaining positions of spouses shape portfolio outcomes (Ke, 2021; Guiso and Zaccaria, 2023). Furthermore, even retail investors who decide alone may encounter the views of others on investing forums and social media (Cookson et al., 2024). Our experiment provides evidence relevant to these settings by isolating the causal contributions of communication and aggregation.

The remainder of the paper proceeds as follows. Section 2 presents the conceptual framework. Section 3 describes the experimental design. Section 4 describes the data and empirical strategy. Section 5 presents the results, and Section 6 concludes.

2 Conceptual framework

In this section, we present a framework that compares individual and joint decision-making. To decompose the difference between them, we add two intermediate regimes in which individual decisions are pooled, first without and then with communication. For each step from one regime to the next, we present general expressions for the gains and losses attributable to each feature of joint decision-making. We then apply the framework to the portfolio allocation context in order to motivate our choice of outcome measures and to generate predictions that we can test with our experimental data.

2.1 General framework

Setup and Individual regime. We consider two individuals, i and j , who face a common decision problem. Each chooses a decision x from a convex feasible set $\mathcal{X}$. Let θ denote the true state of the world and $u(x, \theta)$ the payoff from decision x in state θ . Individual i 's information about θ is captured by $\mathcal{I}_i$, with subjective density $\pi_i(\theta | \mathcal{I}_i)$, perceived precision $\gamma_i > 0$, and mean belief $\hat{\theta}_i = \mathbb{E}_i[\theta | \mathcal{I}_i]$. In the individual decision-making regime, individual i chooses the decision perceived to maximize their expected payoff:

$$x_i \in \arg \max_{x \in \mathcal{X}} \mathbb{E}_i[u(x, \theta) | \mathcal{I}_i]. \quad (1)$$

We assume throughout that expected payoffs are finite and continuous in the decision, and that an optimal decision exists in each regime.

Individuals and groups may have several strategies available for maximizing their expected payoff, each targeting different subcomponents of performance. To examine how the decision-making regime affects each subcomponent separately, let $v(x, \theta)$ denote a payoff-relevant outcome. In our application, u is the Sharpe ratio of a portfolio, while the main outcomes v capture expected excess returns, volatility, concentration, and other measures of portfolio risk (Section 2.2).

Crowd regime. In the first group regime, individuals i and j again make private decisions, but now with awareness that their decisions will be averaged together before evaluation. We assume that *Crowd* members make their decisions sincerely in this regime, behaving identically to how they would under individual decision-making.¹ The resulting group decision is

¹A sophisticated individual would instead maximize the expected payoff of the pooled decision:

$$x_i \in \arg \max_{x \in \mathcal{X}} \mathbb{E}_i \left[u \left(\frac{x + x_j}{2}, \theta \right) \middle| \mathcal{I}_i \right], \quad (2)$$

where $\mathcal{I}_i$ also contains any information available to individual i about their partner's behavior and beliefs, including higher order beliefs. Our experiment represents a particularly complex setting for strategic decision-making. Even assuming equal allocations to exactly ten stocks, subjects must choose from and anticipate their partner's choices from $\binom{500}{10} \approx 2.5 \times 10^{20}$ portfolios. Allowing for unequal allocations complicates this problem further. In a level- k framework (Nagel, 1995), sincere submissions correspond to level-0 reasoning, best responses to anticipated sincere submissions correspond to level-1 reasoning, and so on. However, even

denoted by $X^C = (x_i + x_j)/2$.

We define the pooling benefit as the difference between the outcome of the pooled decision $v(X^C, \theta)$ and the average outcome of the two individual decisions that produce it:²

$$B_v^C \equiv v(X^C, \theta) - \frac{v(x_i, \theta) + v(x_j, \theta)}{2}. \quad (3)$$

The average individual outcome corresponds to the expected outcome from implementing either individual's decision with equal probability. The sign and magnitude of the pooling benefit for a given outcome depend on the curvature of the outcome function. For any outcome that is linear in x , pooling preserves the average of the two individual outcomes and results in a pooling benefit of zero. For any outcome that is concave in x , the pooling benefit is instead weakly positive by Jensen's inequality; it is weakly negative for convex outcomes.

The pooling benefit is larger for concave outcomes when partners disagree more. Holding the pooled decision and the direction of the disagreement fixed, the size of the benefit is weakly increasing in their distance. Differences in beliefs, thinking styles, and other cognitive inputs are one source of such disagreement. This links the pooling benefit to pair composition.

Section 2.2 derives closed-form expressions for the pooling benefit for the outcomes in our application. B_v^C serves as a common benchmark against which to compare the performance of more interactive decision-making regimes.

Panel regime. In the next regime, the individuals i and j submit private, individual decisions as before, except that they engage in discussion first. Let m_j denote the information conveyed by individual j to their partner i and $L_i(m_j \mid \theta, \mathcal{I}_i)$ denote the likelihood that i assigns to receiving m_j if the true state is θ . We represent updating as Bayesian under the following subjective likelihood, which need not be accurate:

$$\pi_i^D(\theta) \propto L_i(m_j \mid \theta, \mathcal{I}_i) \pi_i(\theta \mid \mathcal{I}_i), \quad (4)$$

where the superscript D denotes beliefs after discussion. The new mean belief is $\hat{\theta}_i^D$.

We define the *Panel*-regime variables $\tilde{x}_i$, $\tilde{x}_j$, X^P , and B_v^P analogously to those introduced under previous regimes. We assume that *Panel* members decide sincerely after discussion,³

in simple settings, subjects typically exhibit no more than a few levels of reasoning (Crawford et al., 2013). Consistent with our assumption of sincere choice, *Crowd* members' portfolios in our experiment change little between Round 1, when they face individual incentives, and Round 2, under shared incentives (Section 5.6). We adopt this as our default assumption throughout the framework.

²In our experiment, this benefit can be calculated either using individual and pooled decisions from Round 1 or from Round 2. Our analyses use Round-2 decisions unless stated otherwise.

³As in the *Crowd* regime, a sophisticated individual would instead solve the analogue of equation (2) under their updated beliefs. While communication provides signals about partners' beliefs and preferences over actions, strategic choice also requires the anticipation of how they will respond to one's own expected submission. Prior evidence shows that subjects can fail to best respond even given their own stated beliefs (Costa-Gomes and Weizsäcker, 2008), and that uncertainty creates difficulties with contingent reasoning (Martínez-Marquina et al., 2019). Furthermore, holding beliefs fixed, strategic response under quadratic pay-

making the decision they individually prefer under their updated beliefs:

$$\tilde{x}_i \in \arg \max_{x \in \mathcal{X}} \mathbb{E}_i^D[u(x, \theta)], \quad (5)$$

where $\mathbb{E}_i^D$ is the expectation operator under individual i 's posterior. Discussion therefore changes decisions only by updating beliefs. We allow for the magnitude of updating to be asymmetric between partners and assume that it depends on the perceived precision of their beliefs. We denote $\varphi_i = \gamma_i/(\gamma_i + \gamma_j)$ as the perceived relative precision of individual i before discussion. These quantities can reflect an individual's confidence, as well as other characteristics, such as their thinking style. Suppose that after discussion, each individual shifts their mean belief part of the way toward their partner's:

$$\hat{\theta}_i^D = (1 - \kappa_i)\hat{\theta}_i + \kappa_i\hat{\theta}_j, \quad \kappa_i = d(1 - \varphi_i), \quad \kappa_j = d\varphi_i, \quad (6)$$

where $d \in [0, 1]$ measures how fully each individual updates. We use this rule as a simple representation of how relative precision could affect the weight individuals place on each other's views. A value of $d = 1$ implies full convergence in mean beliefs, while $d < 1$ means that individuals place additional weight on their own initial view; the individual with higher perceived precision is more influential for any given $d > 0$. Whether convergence in beliefs reduces the pooling benefit depends on the mapping between beliefs and decisions. In Section 2.2, we derive conditions under which it does for some payoff-relevant outcomes in our application.

Decomposition of communication effect. We now obtain the following decomposition:

$$v(X^P, \theta) - v(X^C, \theta) = \underbrace{\frac{1}{2} \sum_{h \in \{i, j\}} [v(\tilde{x}_h, \theta) - v(x_h, \theta)]}_{\text{change in average individual outcome}} + \underbrace{(B_v^P - B_v^C)}_{\text{change in pooling benefit}}. \quad (7)$$

The first term captures how belief updating affects the average individual outcome. When $v = u$, the first term is nonnegative in expectation if the updating individual holds a correct prior, correctly understands the process by which their partners' messages are generated, and updates using Bayes' rule (Blackwell, 1953). In other words, people who properly interpret their own information and that of others can expect to benefit from obtaining more information. These assumptions are demanding in many contexts, including in our application. If any of these assumptions do not hold, for example because individuals neglect the correlation in information between sources (Enke and Zimmermann, 2019), this term can be negative.

The second term is the change in the pooling benefit between the *Crowd* and *Panel*

offs predicts divergence in partners' submissions due to each individual attempting to offset their partner's anticipated choice. In our experiment, we instead observe strong convergence in *Panel* partners' portfolios between Round 1 and Round 2 (Section 5.4), suggesting that the belief updating channel dominates any strategic adjustment. These considerations support our assumption of sincere decision-making for this regime.

regimes. Its sign and magnitude again depend on the curvature of the outcome v and how discussion affects the extent of disagreement between partners. Intuitively, the errors in the two individuals' submissions may offset each other when each one works alone. Allowing them to communicate can make their errors more similar, which reduces the benefit of averaging even as it may improve the quality of their individual submissions.

Individuals making more similar decisions based on their updated beliefs can be beneficial if perceived precision is well-calibrated. At the same time, their decisions may converge in ways that reduce the pooling benefit. As such, communication trades off potential improvements in individual decision quality against possible reductions in the pooling benefit (Lorenz et al., 2011; Becker et al., 2017).

Team regime. In the final regime, the individuals i and j also communicate and hold the posteriors π_i^D and π_j^D after discussion. However, instead of submitting individual decisions that are pooled, they must now agree on one joint decision. We hold their posteriors fixed at those in the *Panel* regime and model collaborative aggregation as a bargaining problem; their preferred decisions $\tilde{x}_i$ and $\tilde{x}_j$ and potential pooling benefit B_v^P remain unchanged. The comparison between the *Panel* and *Team* regimes therefore isolates the effect of the aggregation procedure. Let u_i^0 denote individual i 's perceived payoff if no agreement is reached, net of any costs of conflict. We assume that there exists a feasible decision that gives both members a higher expected payoff than disagreement.

The individuals decide upon a joint submission according to the generalized Nash bargaining rule:

$$\begin{aligned} X^T \in \arg \max_{x \in \mathcal{X}} & \left[\mathbb{E}_i^D[u(x, \theta)] - u_i^0 \right]^\eta \left[\mathbb{E}_j^D[u(x, \theta)] - u_j^0 \right]^{1-\eta} \\ \text{s.t. } & \mathbb{E}_h^D[u(x, \theta)] \geq u_h^0, \quad h \in \{i, j\}, \end{aligned} \quad (8)$$

where X^T is the group decision in this regime and $\eta \in (0, 1)$ is i 's bargaining weight.

When expected payoffs are differentiable and concave, the bargaining solution maximizes a weighted average of the individuals' expected payoffs. The resulting weight on individual i 's expected payoff is denoted by λ , and determined as:

$$\lambda = \frac{\eta / [\mathbb{E}_i^D[u(X^T, \theta)] - u_i^0]}{\eta / [\mathbb{E}_i^D[u(X^T, \theta)] - u_i^0] + (1 - \eta) / [\mathbb{E}_j^D[u(X^T, \theta)] - u_j^0]}. \quad (9)$$

Under the assumptions stated in Section 2.2, the resulting joint decision will be a linear combination of the members' individually preferred decisions: $X^T = \lambda \tilde{x}_i + (1 - \lambda) \tilde{x}_j$. Unlike the pooling rule in the previous regimes, the aggregation procedure in the *Team* regime allows for the unequal weighting of partners' preferred decisions.

Decomposition of aggregation effect. Let $B_v^T \equiv v(X^T, \theta) - [\lambda v(\tilde{x}_i, \theta) + (1 - \lambda)v(\tilde{x}_j, \theta)]$ denote the *Team* pooling benefit, i.e., the difference between the outcome of the joint deci-

sion and the weighted average of the outcomes of the individually preferred decisions. We decompose the change in performance between the *Panel* and *Team* regimes as follows:

$$v(X^T, \theta) - v(X^P, \theta) = \underbrace{(\lambda - \frac{1}{2})[v(\tilde{x}_i, \theta) - v(\tilde{x}_j, \theta)]}_{\text{change in weighted average individual outcome}} + \underbrace{(B_v^T - B_v^P)}_{\text{change in pooling benefit}}. \quad (10)$$

The sign and magnitude of the first term depend on which individual receives greater weight. It is positive when the individual whose preferred decision yields a higher value of the outcome in the true state is more heavily weighted. The sign and magnitude of the second term again depend on the curvature of the outcome and where the joint decision lies relative to the individually preferred ones. If the outcome v is concave and X^T is a weighted average of $\tilde{x}_i$ and $\tilde{x}_j$, then B_v^T is weakly positive by Jensen's inequality. In Section 2.2, we derive conditions under which unequal weighting reduces B_v^T relative to B_v^P for some payoff-relevant outcomes.

The scope for collaborative aggregation to affect groups' decisions is limited when beliefs have already converged. In the extreme case, if individuals' preferred decisions are identical, bargaining and equal weighting produce the same outcome.

2.2 Application to portfolio choice

We now consider an application of the above framework to a setting where an investor aims to maximize the Sharpe ratio of a portfolio by choosing non-negative asset weights w that sum to one. This outcome is given by

$$S(w) = \frac{w' \mu}{\sqrt{w' \Sigma w}},$$

where μ is a vector of expected excess asset returns and Σ is the covariance matrix of asset returns. The denominator $\sqrt{w' \Sigma w}$ is the volatility of the portfolio.

In terms of the framework, the incentivized payoff is $u(w, \theta) = S(w)$. The decision made by individuals and groups is $x = w$ and the state of the world is $\theta = (\mu, \Sigma)$. Individual i 's beliefs about the relevant quantities are given by $\hat{S}_i(w)$, $\hat{\mu}_i$, and $\hat{\Sigma}_i$.

When $w' \mu > 0$, investors can improve the Sharpe ratio of their portfolio by increasing excess returns or reducing volatility, holding the other component fixed. We expect retail investors to have limited skill at choosing assets with higher expected returns (Odean, 1999). By contrast, volatility and related portfolio risk outcomes are potentially easier for them to forecast or control. The volatility of a portfolio is the square root of its variance $V(w) = w' \Sigma w$. Following Campbell et al. (1997), for portfolios with fixed weights, $V(w)$ satisfies

$$V(w) = \underbrace{\psi(w)}_{\text{concentration}} + \underbrace{\overline{\sigma^2}_{w^2}}_{\text{asset risk}} + \underbrace{\sum_{k \neq \ell} w_k w_\ell \text{Cov}(r_k, r_\ell)}_{\text{co-movement}}, \quad (11)$$

where $\psi(w) = \sum_k w_k^2$ is the Herfindahl–Hirschman Index (HHI) of the portfolio, r_k and σ_k^2 denote the return and variance of an asset k contained in the portfolio, and $\overline{\sigma^2}_{w^2} =$

$\sum_k w_k^2 \sigma_k^2 / \psi(w)$ is the average asset variance weighted by squared portfolio weights. Concentration as measured with the HHI is under the direct control of the investor, while asset variances can be estimated more accurately than expected returns (Merton, 1980). This decomposition motivates our choice of empirical risk outcomes to examine in Section 5.

Outcome properties. We now establish the relevant properties of each outcome that will allow us to generate predictions. First, expected excess returns $v(w, \theta) = w' \mu$ are linear in portfolio weights. Second, the negative transformation of portfolio volatility ($v(w, \theta) = -\sqrt{V(w)}$) is a concave function. Third, the negative transformations of portfolio variance ($v(w, \theta) = -V(w)$) and concentration ($v(w, \theta) = -\psi(w)$) are also concave functions. More specifically, they take the quadratic form $v(x, \theta) = b(\theta) + a(\theta)'x - \frac{1}{2}x'Qx$, with $Q = 2\Sigma$ and $Q = 2I$, respectively, where I is the identity matrix. These outcome properties allow us to predict the sign of the pooling benefit for each outcome separately.

Pooling. We now apply the framework from Section 2.1 to this setting. For simplicity, we assume throughout that portfolios' asset weights are fixed when they are constructed and do not drift with asset prices. Writing w_i and w_j for the portfolios of individuals i and j , we refer to the average outcome across these two portfolios as the individual benchmark. The pooling benefit is the improvement in an outcome achieved by the pooled portfolio $w^C = \frac{1}{2}(w_i + w_j)$ relative to this benchmark.

For expected portfolio returns $v(w, \theta) = w' \mu$, the pooling benefit is zero because this outcome is linear in portfolio weights (equation (3)); pooling preserves the average expected return of the two portfolios. By contrast, concavity predicts a weakly positive pooling benefit for portfolio variance, volatility, and concentration. Outcomes that are additionally quadratic in form allow for the following closed-form expression of the pooling benefit:

$$B_v^C = \frac{1}{8}(x_i - x_j)'Q(x_i - x_j). \quad (12)$$

Substituting $Q = 2\Sigma$ and $Q = 2I$ into equation (12) for $-V(w)$ and $-\psi(w)$, we obtain:

$$\begin{aligned} B_V^C &= \frac{1}{4}(\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j), \\ B_\psi^C &= \frac{1}{4}\|w_i - w_j\|^2 = \frac{\psi(w_i) + \psi(w_j) - 2\sqrt{\psi(w_i)\psi(w_j)}c_{ij}}{4}, \end{aligned} \quad (13)$$

where σ_i and σ_j are portfolio volatilities, ρ_{ij} is the correlation between their returns, and c_{ij} is the cosine similarity between portfolio weight vectors. Intuitively, holding individual portfolio volatilities fixed, reductions in portfolio variance from pooling are smaller when the two portfolios' returns are more correlated. Likewise, holding individual concentration fixed, pairs with more similar allocations see smaller reductions in concentration. Although volatility is not quadratic, its pooling benefit also has a closed-form expression: $B_\sigma^C = \frac{1}{2}(\sigma_i + \sigma_j) - \frac{1}{2}\sqrt{\sigma_i^2 + \sigma_j^2 + 2\rho_{ij}\sigma_i\sigma_j}$. This benefit is nonnegative and, when individual portfolio volatilities are held fixed, is smaller when the portfolios' returns are more correlated.

Prediction 1. *Pooling has no effect on a portfolio’s expected excess returns relative to the average of the two individual portfolios, but it weakly reduces variance, volatility, and concentration.*

To test this prediction using our experimental data, we can compare the outcomes of pooled *Crowd* portfolios against their individual benchmarks. The expected directions for other key outcomes are ambiguous without further assumptions. Since the asset weights of the pooled portfolio differ from those of each individual portfolio, whether the asset risk term increases or decreases depends on the riskiness of the assets that receive greater weight. Likewise, the co-movement component can rise or fall depending on the weighted covariance between the assets being combined relative to those held in each individual portfolio.

Communication. For the communication and aggregation results, we approximate the incentivized payoff with the mean-variance objective $u(w, \theta) = w'\mu - \frac{A}{2}w'\Sigma w$, where $A > 0$ is a common risk-aversion parameter. We further assume that partners hold fixed, shared beliefs about asset covariances Σ , represented by a positive definite matrix $\hat{\Sigma}$,⁴ and that only the budget constraint is binding. Each individual’s portfolio is then an affine function of their mean belief about expected returns μ . Under the updating rule in equation (6), discussion causes partners’ beliefs about expected returns to converge. Since both individuals’ preferred portfolios follow the same affine mapping from these beliefs, we obtain the following prediction about individuals’ portfolio choices:

Prediction 2. *Communication increases the similarity of partners’ portfolios.*

To test this prediction, we can compare the similarity of *Panel* partners’ portfolios with those of *Crowd* partners. The welfare implications of this convergence differ by outcome. For expected returns, the change in pooling benefit term from equation (7) is again zero because returns are linear in portfolio weights. The effect of communication on this outcome therefore depends entirely on whether the revised individual portfolios have higher expected returns than the pre-discussion portfolios. As argued above, we expect the ability of groups to identify assets with high expected returns through discussion to be limited. They may nonetheless regard these discussions as informative.

Convergence in beliefs about expected returns can reduce the size of the pooling benefit for portfolio variance and concentration even when we hold beliefs about Σ fixed. Equation (12) shows that the pooling benefit for quadratic outcomes depends on the degree of difference between partners’ individually preferred decisions. Under the common affine mapping between beliefs and choices, convergence in beliefs creates a proportionate reduction in the difference between partners’ portfolio weights and therefore also attenuates the pooling benefit.

⁴Allowing for disagreement and changes in beliefs about Σ would add another channel through which discussion could improve the risk profile of individually preferred portfolios. We discuss this channel informally below.

If we relax the assumption of fixed beliefs about Σ , communication could also improve the risk profile of individual portfolios by informing individuals about asset volatilities and between-asset correlations. They can use this information to avoid large exposure to risky assets or sets of highly correlated assets. However, our expectation is that subjects' discussions and belief updating will primarily revolve around returns. This is motivated by evidence that retail investors' asset choices are predicted by their return beliefs (Giglio et al., 2021). Furthermore, portfolio risk and diversification may be more complicated topics of discussion than forecasted returns or return narratives. This focus on perceived return advantages may draw attention away from opportunities to improve portfolio risk.

These considerations motivate the following prediction:

Prediction 3. *Communication does not systematically increase expected portfolio returns. Its effect on portfolio risk depends on whether improvements in individual portfolio risk are sufficient to offset any reduction in the pooling benefit.*

To test this prediction, we can check whether average individual portfolio quality differs between *Crowd* and *Panel* subjects. We can also compare the size of the pooling benefit between these conditions.

Collaborative aggregation. Finally, we examine how the switch in the aggregation procedure is expected to affect performance. We adopt the same assumptions from above regarding the mean-variance approximation, portfolio constraints, and subjects' beliefs, under which the *Team* decision is a weighted average of the individual decisions that each member would have made in the absence of bargaining. For a quadratic outcome v , the benefit of this weighted average is then $B_v^T = \frac{1}{2}\lambda(1 - \lambda)(\tilde{x}_i - \tilde{x}_j)'Q(\tilde{x}_i - \tilde{x}_j) = 4\lambda(1 - \lambda)B_v^P$, and equation (10) becomes

$$v(X^T, \theta) - v(X^P, \theta) = \underbrace{(\lambda - \frac{1}{2})[v(\tilde{x}_i, \theta) - v(\tilde{x}_j, \theta)]}_{\text{change in weighted average outcome}} - \underbrace{4(\lambda - \frac{1}{2})^2 B_v^P}_{\text{reduction in pooling benefit}}. \quad (14)$$

This equation illustrates a similar trade-off to the one that governs the effect of communication. Giving greater influence to better individual portfolios can increase expected returns or reduce portfolio risk. The pooling benefit is again zero for expected returns, while departures from equal weighting reduce its magnitude for variance and concentration. Any net improvement in risk outcomes would require the gain in quality to exceed the lost pooling benefit.

As with communication, we expect little systematic improvement in expected returns. Collaborative aggregation could reduce portfolio risk if greater influence is assigned to the individual with a better understanding of asset risk and diversification, although influence may reflect other characteristics, such as confidence in return expectations or thinking styles. Moreover, if communication has already produced similar portfolios or influence is approxi-

mately equal, then the magnitude of any aggregation effect will be limited. This motivates the following prediction:

Prediction 4. *Collaborative aggregation does not systematically increase expected portfolio returns. Its effect on portfolio risk is ambiguous, but its potential magnitude is limited by the amount of disagreement over preferred portfolios that remains after discussion.*

To test this prediction, we can compare outcomes between *Panel* and *Team* pairs.

3 Experimental design

Our data collection is divided into two parts. In the intake survey, we recruit participants for the main experiment, collect data on their thinking styles and demographics, and give them information necessary for understanding the portfolio allocation task. In the experiment, subjects complete the main investing task. Figure 1 summarizes the two stages and their constituent modules.

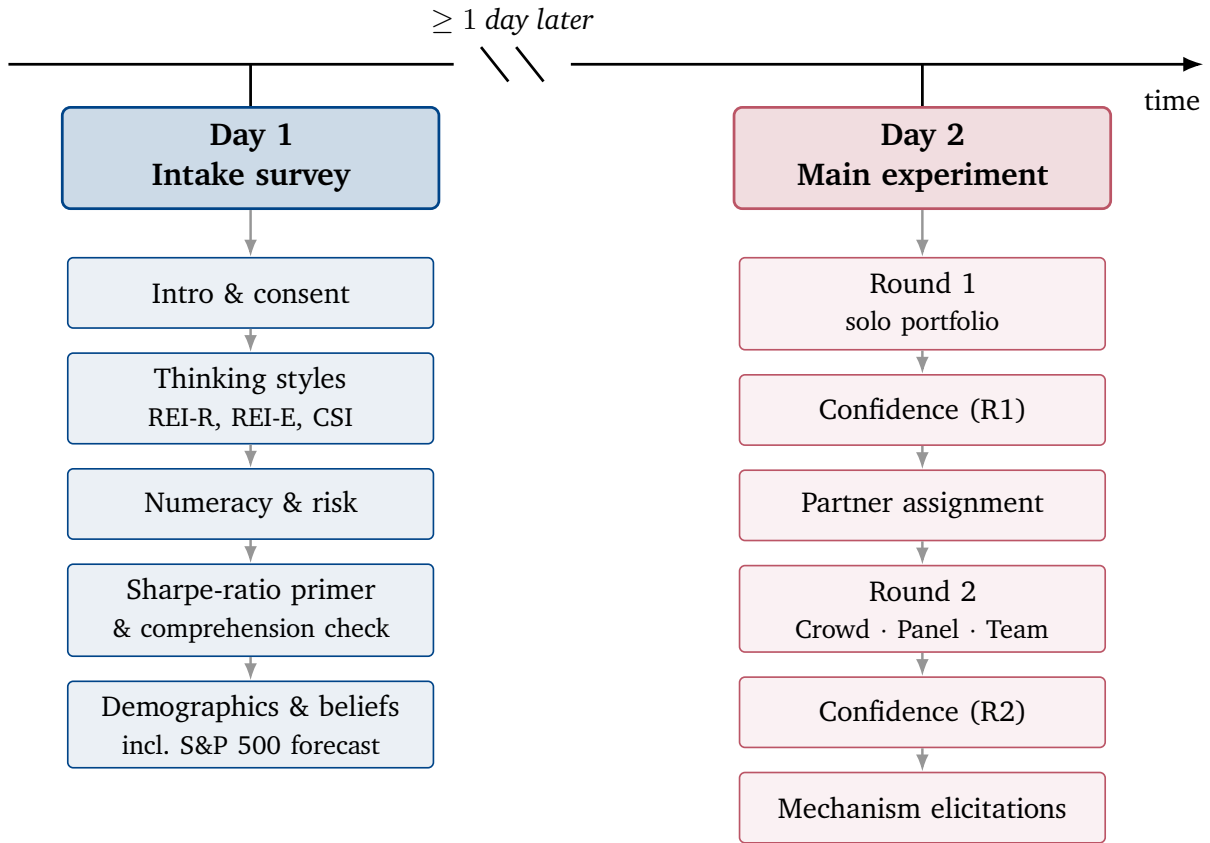
3.1 Intake survey

To recruit subjects who hold relevant background knowledge and who would likely be motivated to participate in our task, we explicitly advertise our study as an investing task on Prolific. Only subjects who self-report owning investments in their private holdings are eligible to sign up. Participants sign up via an intake survey that informs them of the date and time of the main experiment and elicits a commitment from them to participate (Figures A.13 through A.15).

Next, we elicit analytic and intuitive thinking styles using a psychometric instrument. We use 22 items taken from the *Rational-Experiential Inventory* (Norris and Epstein, 2011). The REI conceptualizes analytic and intuitive thinking as separate dimensions that can vary independently within the same person. Accordingly, we treat the two styles as distinct dimensions rather than opposite ends of a single continuum. To keep the survey short, we selected only questions with high factor loadings (absolute value > 0.3) on the respective rationality and intuition components in the PCA reported by Norris and Epstein (2011). The questions featured in the Rationality and Experientiality instruments are available in Appendix Figures A.18 and A.19, respectively. For robustness analyses, we also elicit subjects' one-dimensional thinking style scores on the Cognitive Style Index (Allinson and Hayes, 1996) (Figure A.20).

We then collect data on subject numeracy using the Berlin Numeracy Test (Cokely et al., 2012) (Figure A.21) and measure participant risk preferences (Figure A.22). We also include elicitations aimed at detecting AI-assisted responding, including an altered logic item designed to prompt an incorrect answer from LLMs.

Figure 1: Timeline of the experiment



Notes: This figure displays the timeline of the experiment. The left and right columns show the sequence of screens and elicitations that subjects encounter during the intake survey and the main experiment, respectively. After completing the intake survey, subjects return at least one day later for the main experiment. Treatment assignment occurs after Round-1 confidence is elicited, when participants are assigned to a partner whose thinking style is either similar to or different from their own and to the *Crowd*, *Panel* or *Team* decision mode.

The elicited psychometric measures exhibit patterns consistent with construct validity and convergent validity. The REI rationality score (REI-R) is positively associated with the Cognitive Style Index ($\beta = 0.39$, $p < 0.01$) and with numeracy ($\beta = 0.03$, $p < 0.05$) (Table A.1).

In one of the final modules of the intake survey, we provide subjects with information and questions on Sharpe ratios (Figures A.23 through A.27). Among other information, they learn from this material that diversification is generally a more appropriate strategy for maximizing Sharpe ratios than choosing “winner” stocks. In order to comply with Prolific’s rules regarding the use of comprehension checks, subjects who fail the comprehension check questions twice are still paid for taking the survey but are not invited to the main experiment.

In the final module, we elicit subject demographics and attitudes toward investing that may be relevant for interpreting behavior in the portfolio allocation task (Figures A.28–A.29). Most importantly, we elicit their forecast of the S&P 500 in 3 months’ time, which we use as a cognitive input in later analyses. The survey ends with a reminder about the date and time of the experiment (Figure A.30).

Figure 2: Round-1 portfolio interface

Time left for this task: 5:15

Investment task: Round 1

Please complete this task as if you have \$1,000 to allocate to stocks in the S&P 500. Choose your investment(s) by selecting a ticker name and the amount invested for **at least one**, and **at most ten** stocks.

We will calculate your portfolio's Sharpe ratio using **real returns data** starting from today through 3 months from today. You can earn up to \$10 by maximizing your Sharpe ratio. As such, your best strategy is to design a portfolio with high returns and low volatility. More info is available below.

► *More information on incentives*

To qualify for bonus payments, your portfolio must have **exactly \$1,000** invested. You cannot continue until the timer has ended. When time is up, the page will **automatically advance**, regardless of whether you have invested \$1,000.

Total invested: \$0 / \$1,000

Stock 1 Ticker:	Amount (\$):
	0
Stock 2 Ticker:	Amount (\$):
	0
Stock 3 Ticker:	Amount (\$):
	0

Notes: This figure displays the Round-1 portfolio interface that participants saw during the main experiment. The task was to allocate \$1,000 to between one and ten stocks from the S&P 500. Each “Stock x Ticker” field was a searchable dropdown list that displayed the company name, ticker, and GICS sector code of all stocks listed in the S&P 500. For brevity, this screenshot omits the fields for Stock Ticker 4 through Stock Ticker 10, which were present in the experimental interface.

3.2 Main investment experiment

In order to allow subjects to work with their group partner in real time, we offer a participation payment conditional on returning at a prespecified time to complete the main task. Subjects who return at the scheduled time are given information on the nature of the main task and their incentives (Figure A.31).

Round 1: solo portfolio. Shortly after giving consent, subjects proceed to the first investing task. In this round, they spend approximately 10 minutes working alone.⁵ Their task is to create a portfolio worth \$1,000 by selecting stocks from companies listed on the S&P 500.

⁵The Round-1 decision screen closed at a fixed time, 12 minutes after the start of the experiment. The additional two minutes were intended to give participants time to read instructions and to allow those who arrived slightly late to participate. Participants who reached the Round-1 decision screen early could therefore spend slightly more than 10 minutes on it.

This task is shown in Figure 2. They are required to select a stock ticker and amount invested for at least one and at most 10 individual companies. We enforce that all selected tickers are valid. Portfolios that do not sum to \$1,000 are treated as invalid and are ineligible for bonus payments.

They are incentivized to design a portfolio with a high Sharpe ratio. They are told that the Sharpe ratios of their portfolios will be calculated after 3 months using real return and volatility data and that they can receive a bonus payment of up to \$10 depending on how well their portfolio performs (Figure A.36). To implement this, we calculate and rank the realized Sharpe ratios of every eligible participant separately by round. We then randomly select 10% of participants to receive bonus payments and randomly choose either their Round 1 or Round 2 portfolio to be evaluated for payment. Selected subjects whose Sharpe ratios fall into the top decile of all subjects in the chosen round receive \$10, while those in the next decile receive \$9. This pattern continues through the bottom decile, where subjects receive \$1.

We use this incentive scheme to reduce the attractiveness of extreme investing strategies and to reward improvements in performance throughout the outcome distribution. We also elicit subjects' confidence (0–100) in the ability of their portfolio to generate a high Sharpe ratio (Figure A.37; Enke and Graeber (2023)) immediately after submission.

Round 2: group portfolio. To minimize attrition and inattentive participation in the group task, we restrict participation in Round 2 to subjects who followed instructions and successfully designed a portfolio in Round 1. Every participating subject in Round 2 is first assigned to a treatment condition (diversity or homogeneity) and then to a partner within that condition. Partner assignment is conducted using a greedy local-search algorithm that maximizes Euclidean distance in thinking styles between partners in the diversity condition and minimizes it in the homogeneity condition (visualized in Appendix Figure A.1).

The algorithm strikes a balance between performance and speed in pairing subjects in our live setting. The first step in the algorithm lists all possible pairings between subjects, calculates their distances, and makes greedy pairings starting with the lowest- and highest-distance pairs in the homogeneity and diversity conditions, respectively. Assignment continues with the best remaining feasible pair until all subjects are assigned.

Since this procedure is not likely to produce globally optimal matches, the next step in the algorithm aims to improve match quality while remaining computationally tractable. It does so by scanning through all possible partner swaps in a fixed order and accepting those that increase (decrease) the sum of a concave (convex) function of Euclidean distance in the diversity (homogeneity) condition as they are encountered. After a swap, both of the affected pairs remain eligible for additional swaps that improve the objective function in the same pass using the updated pairing of subjects. This swapping exercise is repeated with up to 50 passes or until no more such swaps are available. The concavity (convexity)

Figure 3: Round-2 team format descriptions by decision mode

Time left for team task: 1:15

Investment task: Round 2

Team description

In this round, you will again **edit a portfolio**. However, when you are done, it will be **pooled together** with your teammate's and re-allocated to \$1,000. You will both be paid based on how the pooled portfolio performs.

For example, if you invest \$100 in Stock A and your teammate invests \$100 in Stock B, then your pooled portfolio will have \$50 invested in Stock A and \$50 invested in Stock B.

You are **unable to communicate** with your teammate. You may view your **Round 1 portfolio** below.

(a) *Crowd*

Time left for team task: 1:15

Investment task: Round 2

Team description

In this round, you will again **edit a portfolio**. However, when you are done, it will be **pooled together** with your teammate's and re-allocated to \$1,000. You will both be paid based on how the pooled portfolio performs.

For example, if you invest \$100 in Stock A and your teammate invests \$100 in Stock B, then your pooled portfolio will have \$50 invested in Stock A and \$50 invested in Stock B.

You may **chat with your teammate** and view both of your **Round 1 portfolios** below.

(b) *Panel*

Time left for team task: 1:15

Investment task: Round 2

Team description

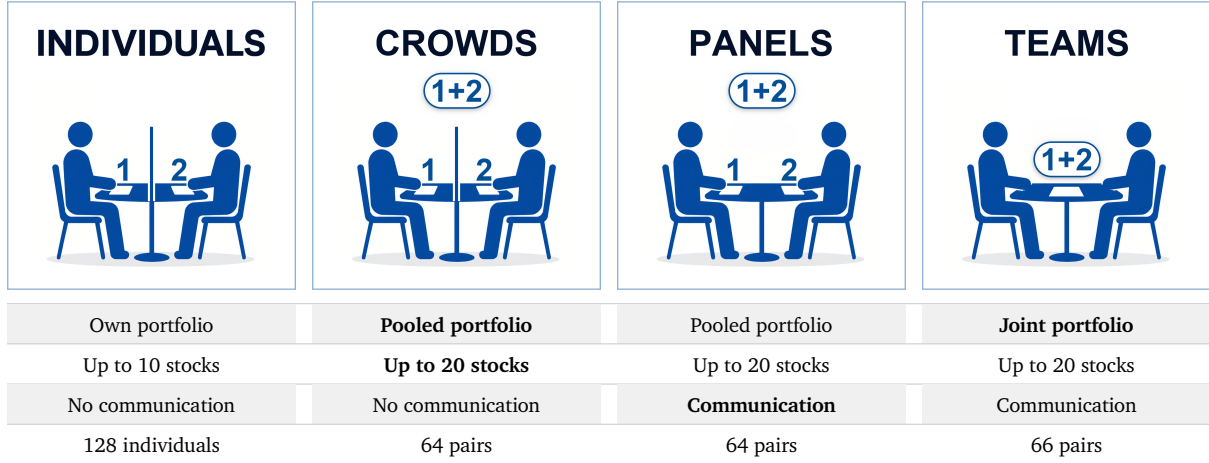
In this round, you and your teammate will **jointly edit a shared portfolio**. You will both be paid based on how this joint portfolio performs.

You may **chat with your teammate** and view both of your **Round 1 portfolios** below.

(c) *Team (joint)*

Notes: This figure displays the Round-2 team format descriptions shown to subjects in each decision mode. Panels (a), (b) and (c) show the descriptions for the *Crowd*, *Panel* and *Team* decision modes, respectively. For brevity, these screenshots omit the investing-task instructions and portfolio interface shown to subjects during Round 2.

Figure 4: Individual and group decision modes



Notes: This figure summarizes the differences between individual decision-making and our three group decision modes in Round 2. Individuals in Round 2 are *Crowd* members who work alone and whose unpooled portfolios provide the benchmark for individual performance. The group portfolios of *Crowd* and *Panel* members are formed by pooling partners' individual submissions together. In contrast, each *Team* submits one joint portfolio. *Panel* and *Team* members can communicate and view their partner's Round-1 portfolio. The stock counts in each mode refer to maximum possible portfolio sizes. The sample sizes refer to the number of valid portfolios submitted in each decision mode in Round 2.

of the distance function is intended to reduce the number of outlier pairs with abnormally high or low distance in each condition. If the number of eligible subjects were odd, the unmatched participant would be assigned to a solo "pair" and excluded from pair-level analyses. However, this did not happen.

We observe that the assignment algorithm successfully changed within-pair thinking-style distance: the Diversity treatment raises within-pair Euclidean distance by 14.43 points relative to a Homogeneity mean of 1.44, without changing pair-level mean scores on either style. It also separates the two arms across the full distribution (Appendix Figure A.1).

Separate from partner assignment, our other treatment manipulation targeted the group decision mode that a subject experienced in Round 2. The task in Round 2 is identical to Round 1, except that some aspects of the experience are modified as follows. Subjects are cross-randomized to one of three decision modes after partner assignment:

1. **Crowd:** subjects work alone but are paid according to the performance of a joint portfolio created by pooling⁶ their portfolio together with that of their assigned partner.
2. **Panel:** same as 1., except that subjects can chat with their partner and view their partner's Round 1 portfolio.
3. **Team:** same as 2., except that subjects design a shared portfolio using jointly editable input fields. Teammates must agree unanimously to submit a portfolio. To equalize

⁶We deliberately pool portfolios in the simplest way possible. The pooling algorithm is explained by informing participants that if they invest \$100 in Stock A and their partner invests \$100 in Stock B, then their pooled portfolio will have \$50 invested in Stock A and \$50 invested in Stock B. This approach is generalized to reallocate the total holdings to \$1,000.

the capacity for diversification between conditions, *Teams* are able to allocate portfolio holdings to up to 20 tickers.

Figure 3 shows the Round 2 portfolio interface separately by decision mode; full-page versions appear in Appendix Figures A.38 and A.39. We visualize individual decision-making and the three group decision modes in Figure 4. Round 2 lasts 10 minutes in all decision modes.

After Round 2 is complete, a number of outcomes of interest are immediately measurable, including indices of diversification, the individual assets and sectors that received the most allocations, and between-treatment differences in how teams discussed and aggregated individual inputs from Round 1. We also measure realized portfolio outcomes over the evaluation window running from June 2 to September 2, 2026.

Mechanism elicitations. To explore mechanisms, we elicit a set of measures after the main task is complete. These include their subjective confidence in their Round 2 portfolio (Figure A.40), explanations of their confidence ratings from both rounds (Figure A.41), an open-text elicitation about their decision-making procedure and team experience, and their relative attention to perceived returns versus risk when designing a portfolio (Figure A.42). In a final task, we give participants the option to reallocate between 0% and 100% of their Round 1 and Round 2 portfolios to an exchange-traded fund (ETF) that tracks the S&P 500 (Figures A.43 and A.44). Higher ETF reallocation indicates a stronger preference to replace the self-selected portfolio with the diversified index fund. The decision is incentivized: a random 10% of subjects are chosen to have their preferred ETF allocation implemented before their portfolio is evaluated.

LLM detection. We use the following techniques to reduce and detect AI-assisted responding among experimental subjects. First, we embed white text on white backgrounds instructing autonomous, LLM-based agents to exit the survey (a technique referred to as prompt injection). We also disable the copying of questions and the pasting of responses via JavaScript to make it more difficult for subjects to consult chat-based AI interfaces such as ChatGPT. Furthermore, we track how many times a participant leaves the browser tab during important questions and tasks. Finally, in the numeracy section of the intake survey, we embed questions designed to elicit atypical or incorrect answers from LLMs but that are easily solvable by humans.

4 Data and empirical strategy

In this section, we first describe the sample we collected and define the variables used in the analysis. We then conduct a balance check and explain the primary specifications that

comprise our empirical strategy.

4.1 Sample

Our sample fell short of the preregistered target of 600 participants in the main experiment due to lower-than-expected rates of subjects passing the comprehension check and returning to participate in the main experiment. Of the 753 individuals who completed the intake survey, 615 (82%) passed the comprehension check and were invited to the main experiment. Of those invited, 480 (78%) returned to participate. Of these 480, 388 (81%) arrived before the end of the first round and designed a valid Round-1 portfolio.

The 388 subjects who proceeded to Round 2 form 194 two-person pairs: 64 *Crowd*, 64 *Panel*, and 66 *Team*. Thinking-style composition is cross-randomized across decision modes, yielding 97 *Diversity* and 97 *Homogeneity* pairs. Because treatment assignment occurred only after the Round-2 analysis sample had been formed, the sample loss described in the previous paragraph cannot generate differential attrition across treatment arms. Four *Team* pairs did not reach unanimous agreement on a Round-2 portfolio before the time limit. For these pairs, we use the latest shared portfolio they had assembled on screen. All conclusions are robust to instead excluding these four *Teams*.

4.2 Variable definitions

In this section, we offer definitions of the key terms and variables used in Section 5. Most importantly, we introduce notation to distinguish between individual portfolios, aggregated portfolios produced by *Crowd* and *Panel* pairs, and joint portfolios produced by *Teams*. We also explain how our key outcome and explanatory variables are constructed.

Units of observation and benchmarks. We define the **individual Round-1 portfolio** as the portfolio submitted by an individual in Round 1 and y_i^{R1} as the value of an outcome y taken by the Round-1 portfolio of individual i . Following Section 2.2, the **Round-1 individual benchmark** is the average outcome of the two individual Round-1 portfolios submitted by the members of the pair (i, j) , calculated as $\bar{y}_{ij}^{R1} = \frac{1}{2}(y_i^{R1} + y_j^{R1})$. Meanwhile, the **pooled Round-1 portfolio** is the portfolio formed by combining the two individual Round-1 portfolios of pair (i, j) using equal-weight pooling. We write y_{ij}^{POOL} for the value of an outcome y taken by the pooled Round-1 portfolio of this pair.

We define the **individual Round-2 portfolio** as a portfolio submitted by an individual in Round 2 in the *Crowd* and *Panel* decision modes, with the outcome y_i^{R2} and the **Round-2 individual benchmark** $\bar{y}_{ij}^{R2}$ defined analogously. The **Round-2 group portfolio** is a new object in Round 2: it is either the pooled Round-2 portfolio of a *Crowd* or a *Panel* or the collaboratively aggregated Round-2 portfolio of a *Team*. We write y_{ij}^{R2} for the value of an outcome y taken by the group portfolio of pair (i, j) .

We next explain how we construct the realized return and volatility outcomes of each portfolio, as well as conceptually related outcomes like historical returns and portfolio concentration.

Realized and historical outcomes. We start by computing the simple daily returns r_t of each portfolio, using a buy-and-hold strategy with no rebalancing over the evaluation window from June 2 to September 2, 2026. Returns are calculated using dividend-adjusted S&P 500 closing prices from Yahoo Finance. To calculate daily *excess* returns, we subtract the daily risk-free rate $r_{f,t} = \text{DGS3MO}_t / (100 \times 252)$ from each day's returns, where DGS3MO is the 3-month Treasury constant-maturity yield. We then record the **realized excess return** variable as the annualized mean daily excess return of a portfolio ($100 \times 252 \overline{(r_t - r_{f,t})}$). **Historical excess returns** are calculated identically over the two-year historical window preceding June 2.

We set the indicator variable **higher excess return than R1** equal to one if a pair's Round-2 realized excess return exceeds its Round-1 individual benchmark, and zero otherwise. As an additional outcome variable, we calculate the **realized log gross return** of a portfolio as $100 \log(V_T/V_0)$, where V_0 and V_T are the initial and final portfolio values. The corresponding **historical log gross return** reports the same quantity over the preceding two-year historical window.

The **realized volatility** of a portfolio is then the annualized standard deviation of its daily returns ($100\sqrt{252} \text{SD}(r_t)$). The **lower volatility than R1** indicator equals one if a pair's realized volatility value falls below its Round-1 individual benchmark, and zero otherwise.

The incentivized **realized Sharpe ratio** variable is calculated as the ratio of our realized excess returns and realized volatility variables. The **higher Sharpe ratio than R1** indicator equals one if a pair's Round-2 realized Sharpe ratio exceeds its Round-1 individual benchmark, and zero otherwise.

We also record the conceptually related **historical asset risk** variable for each portfolio. To construct this variable, we denote a portfolio's vector of asset weights as w_k , where w_k is the share of the \$1,000 budget allocated to asset k . Historical asset risk is then calculated as $\sum_k w_k^2 \sigma_k^2 / \sum_k w_k^2$, where σ_k is the volatility in daily returns for ticker k over the historical two-year window. We divide by $\sum_k w_k^2$ to remove the overall scale of the diagonal variance component associated with portfolio concentration. Since the measure remains weighted by squared portfolio weights, it captures the weighted average of the variance of stocks in a portfolio. In regressions, we use the square root of this measure. We annualize this outcome and report changes in percentage points.

Risk and diversification outcomes. Given noise in market returns, we also use measures of portfolio risk and diversification of each portfolio that do not depend on the evaluation window. We define the **Herfindahl–Hirschman Index (HHI)** of a portfolio as $\sum_k (100 w_k)^2 =$

$10,000 \sum_k w_k^2$. It equals 10,000 for a single-asset portfolio and $10,000/K$ for equal weights over K stocks. Lower values indicate less concentrated holdings.

The **unique assets** and **unique sectors** variables record the numbers of distinct tickers and GICS sectors that receive a positive allocation in a portfolio. The **maximum weight** of a portfolio is the highest weight w_k allocated to any single asset k in a portfolio.

The **weighted average pairwise correlation (APC)** is the weighted average correlation of historical daily returns over the two-year window ending June 2, 2026, between all pairs of distinct tickers in the portfolio. The correlation between assets k and ℓ is weighted by $w_k w_\ell$. It is lower when holdings are spread across assets whose returns are uncorrelated.

Behavioral and secondary outcomes. We also record a number of additional variables that help us to understand the mechanisms behind treatment effects on our primary outcomes. **Portfolio similarity** represents the cosine similarity between the vectors of asset-level dollar weights in two portfolios. It equals 1 when the two portfolios coincide and 0 when they hold no stocks in common. A member’s **weight influence** is the cosine similarity between the vectors of asset-level dollar weights in their own Round-1 portfolio and their realized Round-2 group portfolio.

A member’s **coverage** is the share of the tickers in a pair’s realized Round-2 portfolio that the member also held in Round 1. It runs from 0 to 1 and is higher when a larger share of the final portfolio consists of stocks held by that member in Round 1. The **dominance gap** of a pair is the absolute difference between the coverage values of its two members. It is 0 when the two members are equally represented in the final portfolio and approaches 1 when the portfolio is drawn entirely from holdings unique to one member’s Round-1 portfolio.

Confidence is the subjective confidence (0–100) of a subject in the ability of their portfolio to generate a high Sharpe ratio, elicited separately after Round 1 and Round 2. **Attention to returns** is the self-reported level of attention and effort that a subject directed toward maximizing returns relative to minimizing risk in Round 2 (0–100). Finally, **ETF reallocation** is the share of a portfolio that a subject wishes to reallocate to an ETF that tracks the S&P 500. This measure is incentivized as described in Section 3 and elicited both for Round 1 and Round 2 portfolios.

Treatment variables. The **decision mode** variable records a pair’s assigned decision mode in Round 2, which is *Crowd*, *Panel*, or *Team*. The **thinking-style composition** variable captures whether a pair is assigned to the *Diversity* or *Homogeneity* condition.

We also define **belief diversity** as an indicator variable that records whether the two members of a pair are on the same or the opposite side of the median split of forecasted three-month changes in the S&P 500. Forecasts were elicited in index levels, but we convert them to changes before constructing this indicator.

Pre-treatment covariates. The **REI-R** and **REI-E** variables are the separate analytic and intuitive thinking-style scores measured by the *Rational-Experiential Inventory*. Higher values on each measure indicate more analytic and more intuitive thinking styles, respectively. The **CSI** variable records the thinking-style score measured by the *Cognitive Style Index*. Higher values indicate a more analytic thinking style.

The **numeracy** variable is a subject’s score on the Berlin Numeracy Test (0–4), while the **risk preference** variable captures a subject’s certainty equivalent elicited in the risk-preference module of the intake survey. The **EMH belief** variable captures the reported level of agreement with the statement that “stock prices already reflect what’s publicly known, so it is very hard to reliably pick stocks that will beat the market,” on a Likert scale from 0 (strongly disagree) to 4 (strongly agree). **Gender** is an indicator variable that takes a value of 1 if a subject reports their gender as female, and 0 otherwise. Finally, **age** is a numerical variable capturing the subject’s reported age.

4.3 Balance

Appendix Table A.2 reports the results of balance tests for pre-treatment covariates and baseline outcome variables, which we conduct both at the individual level and at the pair level.

At the individual level, 5 of the 57 tested differences in covariates and baseline outcomes across conditions are statistically significant at the 10% level. This result is close to the 5.7 imbalances expected to occur by chance under random assignment. The results are similar at the pair level, where 3 out of the 24 tested differences are statistically distinguishable from 0, compared with the 2.4 expected by chance.

4.4 Empirical strategy

We first compare joint and individual decision-making before describing our specification that isolates the contributions of composition, communication, and collaborative aggregation. Each step in the decomposition uses decisions made in the same round to account for confounders, such as learning and changes in effort. We complement this decomposition with a specification that examines within-pair changes between rounds.

We use the portfolio notation defined in Section 4.2. For a given outcome y and pair (i, j) , $\bar{y}_{ij}^{R1}$ denotes the pair’s Round-1 individual benchmark, y_{ij}^{POOL} the outcome of its pooled Round-1 portfolio, $\bar{y}_{ij}^{R2}$ its Round-2 individual benchmark, and y_{ij}^{R2} the outcome of its Round-2 group portfolio.

Team-individual comparison. We first estimate the difference in outcomes between *Teams* and individuals working alone in Round 2. This comparison captures the combined effects

of composition, communication and collaborative aggregation. We restrict the sample to pairs assigned to the *Team* and *Crowd* decision modes and define $\tilde{y}_{ij}^{R2}$ as the Round-2 group-portfolio outcome y_{ij}^{R2} for *Team* pairs and the Round-2 individual benchmark $\bar{y}_{ij}^{R2}$ for *Crowd* pairs.

We estimate

$$\tilde{y}_{ij}^{R2} = \alpha + \tau \text{Team}_{ij} + \gamma \bar{y}_{ij}^{R1} + \varepsilon_{ij}, \quad (15)$$

where Team_{ij} equals one for *Team* pairs and zero for *Crowd* pairs. The coefficient τ therefore estimates the difference between the Round-2 group-portfolio outcome of *Teams* and the Round-2 individual benchmark of *Crowd* pairs, controlling for the corresponding Round-1 individual benchmark. We refer to this as our primary ANCOVA specification.

Estimates of equation (15) are reported in Figure 7 and Appendix Table A.3.

Round-2 pooling effect. We next isolate the direct effect of composition by comparing pooled and unpooled portfolios within the *Crowd* condition in Round 2. We construct a stacked dataset where each *Crowd* pair appears twice, once with its Round-2 individual benchmark as the outcome variable and again with its Round-2 group portfolio as the outcome. We estimate

$$y_{ijk}^{R2} = \alpha + \delta \text{Pool}_k + \varepsilon_{ijk}, \quad k \in \{\text{individual, pooled}\}, \quad (16)$$

where $y_{ij,\text{individual}}^{R2} = \bar{y}_{ij}^{R2}$ and $y_{ij,\text{pooled}}^{R2} = y_{ij}^{R2}$, and Pool_k equals one for the pooled observation and zero for the individual-benchmark observation. The coefficient δ therefore identifies the Round-2 mechanical pooling effect for *Crowds*.

The coefficient estimate from equation (16) is reported as the coefficient of Composition in Panel (a) of Figures 9, 10, and 14, and in Appendix Table A.4.

Decision-mode comparisons. We then isolate the effects of communication and collaborative aggregation by comparing *Panels* with *Crowds* and *Teams* with *Panels*, respectively. We also compare *Teams* with *Crowds* in order to capture the combined effect of these two features. For each comparison, we construct a dataset restricted to Round-2 observations that belong to one of the contrasted decision modes. Let a denote the first decision mode in the contrast and b the second. We estimate

$$y_{ij}^{R2} = \alpha_{ab} + \theta_{ab} \mathbf{1}\{D_{ij} = a\} + \gamma_{ab} \bar{y}_{ij}^{R1} + \varepsilon_{ij}, \quad (17)$$

where D_{ij} denotes pair (i, j) 's assigned decision mode and $\bar{y}_{ij}^{R1}$ is their Round-1 individual benchmark. We calculate each member's outcome, including any logarithmic transformation, before averaging across members. The coefficient θ_{ab} therefore estimates the difference in the Round-2 group-portfolio outcome between decision modes a and b , controlling for the corresponding Round-1 individual benchmark. These regressions follow the same ANCOVA specification as the comparison between *Teams* and individuals. Analyses that instead use

the outcome of the pooled Round-1 portfolio as the pre-treatment baseline are flagged explicitly when they appear. For the improvement indicators, $\bar{y}_{ij}^{R1}$ in equations (15) and (17) is the continuous Round-1 individual benchmark of the underlying outcome, using realized volatility in levels for the volatility indicator.

Estimates from equation (17) are reported in Panel (a) of Figures 10 and 14, and in Appendix Table A.4.

Within-pair decomposition. We also estimate a within-pair decomposition that distinguishes the compositional gains available from pooling members' Round-1 portfolios from the gains or losses caused by subjects' choices in Round 2. For *Crowd* and *Panel* pairs, let $\bar{y}_{ij}^{R2}$ denote the average outcome of the two individual Round-2 portfolios. Define the initial pooling benefit as $B_{ij}^{R1} = y_{ij}^{POOL} - \bar{y}_{ij}^{R1}$ and the Round-2 pooling benefit as $B_{ij}^{R2} = y_{ij}^{R2} - \bar{y}_{ij}^{R2}$. With outcomes oriented so that higher values indicate better quality, the total change can be written as

$$y_{ij}^{R2} - \bar{y}_{ij}^{R1} = \underbrace{B_{ij}^{R1}}_{\text{Initial pooling benefit}} + \underbrace{(B_{ij}^{R2} - B_{ij}^{R1})}_{\text{Change in pooling benefit}} + \underbrace{(\bar{y}_{ij}^{R2} - \bar{y}_{ij}^{R1})}_{\text{Change in average decision quality}}. \quad (18)$$

The first term captures the gain from pooling the initial portfolios. The second captures the change in the pooling benefit between rounds and the third captures the change in average individual decision quality. For *Teams*, separate individual Round-2 portfolios are unobserved, so we retain the combined revision component $y_{ij}^{R2} - y_{ij}^{POOL}$ alongside the initial pooling benefit. Equation (7), with $v = -V$, explains how revisions can improve individual portfolios while changing the benefit of pooling. Here the empirical revision component also includes changes between rounds.

Table 1 reports estimates of the components of equation (18), while Figure 8 and Panel (b) of Figures 9, 10, and 14 plot the corresponding total changes separately by decision mode.

While this decomposition is useful for building intuition, these comparisons are confounded by round effects. Since subjects gain experience and may exert different levels of effort between Rounds 1 and 2, the overall change between individual and group decision-making and the revision component do not purely identify the intended quantities. The primary comparisons above address this identification challenge by comparing portfolios constructed in the same round.

Scaling. We standardize continuous outcomes using the standard deviation of the corresponding Round-1 individual benchmark across pairs with valid observations. We scale the improvement indicators instead by the sample standard deviation of the indicator among pooled *Crowd* portfolios, using the same reference across all contrasts. Coefficients in the figures are reported in SD units.

For each continuous outcome, the Round-1 control and the benchmark used for standardization follow the same transformation as the outcome. For selected portfolio risk outcomes, we reverse the sign of the coefficient such that positive values always indicate improvement. The text reports effects in SDs or natural units as indicated; table notes always specify the units and sign conventions used.

Standard errors. Between-arm regressions use HC1 standard errors. Specifications with repeated observations from the same pair cluster standard errors by pair.

5 Results

We report our main results in this section. Section 5.1 first describes the behavior and performance of individual subjects in Round 1. Section 5.2 then compares the performance of *Teams* with individuals working alone. Sections 5.3–5.5 decompose the performance gap between *Teams* and individuals by isolating the effects of composition, communication, and collaborative aggregation. Section 5.6 synthesizes these results and reports takeaways from our decomposition. Finally, Section 5.7 examines the relationship between within-pair differences in cognitive inputs and performance.

5.1 Individual portfolios in Round 1

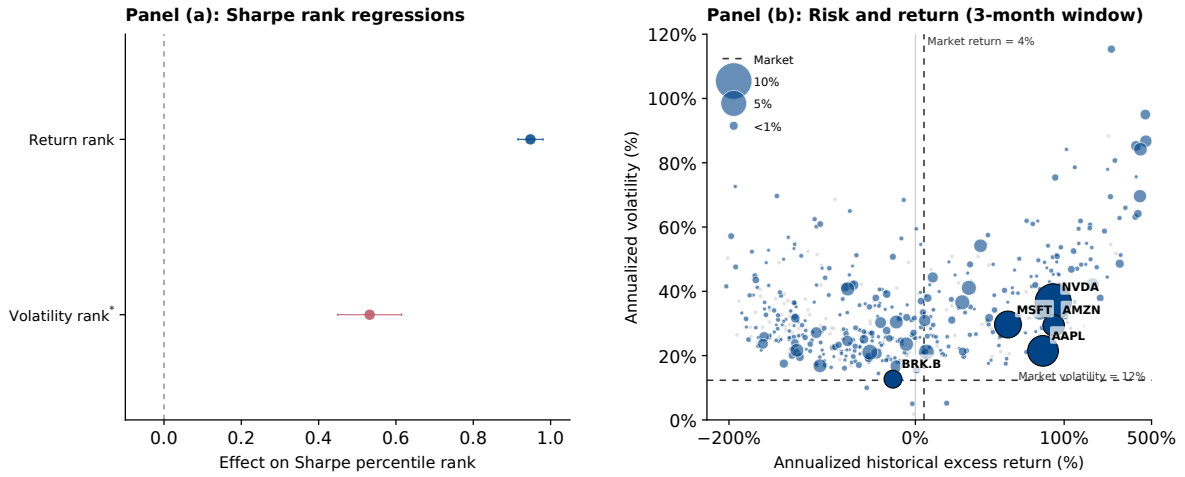
We start by examining subjects’ behavior and performance when they work alone and complete the portfolio allocation task for the first time in Round 1. Reassuringly, subjects’ editing activity suggests substantial engagement with the task. Subjects make a median of 22 portfolio edits in Round 1, with a median interval of 6.8 minutes between their first and last edits (Appendix Table A.6). The median subject makes their final edit 8.6 minutes after opening the portfolio allocation screen. As such, subjects use the majority of their allotted time to design their portfolios.

5.1.1 Investing strategies

Since subjects were paid based on their portfolio’s Sharpe ratio rank, they could improve their performance either by increasing excess returns or reducing portfolio volatility.⁷ In Panel (a) of Figure 5, we report coefficient estimates from regressions of Round-1 portfolio Sharpe-ratio rank on the respective return and volatility rank components of the same portfolios; volatility ranks are constructed such that higher ranks indicate lower volatility.

⁷As discussed in Section 2.2, reducing volatility improves the Sharpe ratio when excess returns are positive. Given that the equity premium in the S&P 500 has been positive over the long run and that 80% of our subjects report positive forecasts of market returns over the evaluation period, we treat ex ante risk reduction as a sensible objective. Excess portfolio returns were indeed positive for the vast majority of subjects during the evaluation window.

Figure 5: Portfolio performance and stock choices in Round 1



Notes: This figure reports (a) the relationship between the realized Sharpe-ratio rank of a portfolio and the ranks of its separate return and volatility components and (b) how Round-1 stock choices vary by historical asset returns and volatility. Panel (a) reports coefficients from separate OLS regressions of a portfolio's realized Sharpe-ratio rank on its realized return and realized volatility ranks, respectively. All ranks are percentiles ranging from 0 to 100 and are calculated using realized outcomes from the three-month evaluation window. Volatility rank is marked with * to indicate that it is sign-reversed, such that higher ranks indicate lower volatility. These regressions include all 388 subjects who submitted a valid Round-1 portfolio. The bars are 95% confidence intervals obtained using robust (HC1) standard errors. Panel (b) plots each S&P 500 ticker by its annualized mean daily excess return (x -axis) and annualized volatility (y -axis) over the three months preceding the experiment. The x -axis is linear between -50% and 50% . It compresses values outside that range for visibility. Marker size is proportional to the ticker's share of total dollars allocated across all portfolios in Round 1. Gray dots are tickers that received no allocations. The ticker labels report the five assets that received the largest dollar allocations. The vertical dashed line corresponds to the mean annualized excess return across tickers, while the horizontal dashed line indicates the annualized volatility of an equally weighted portfolio containing all S&P 500 stocks.

The results show a strong, positive association between a portfolio's Sharpe rank and its rank on each component. While realized return ranks are more strongly related to the incentivized outcome than volatility ranks, it is a separate question whether participants are able to identify assets with higher expected returns. As argued in Section 2.2, we expect subjects to have more control over portfolio risk than returns. We next examine whether subjects' choices are consistent with a strategy targeting portfolio returns, risk, or both.

5.1.2 Stock selection and returns

Descriptively, subjects appear to favor stocks that experienced high returns in the three months prior to the experiment (Panel (b) of Figure 5). To examine this relationship more carefully, we assemble a stock-subject-level dataset where the outcome variable in each row is an indicator for whether a particular stock is included in the subject's Round-1 portfolio. We regress this indicator on three-month historical returns, volatility, and further controls. We find that a one-standard-deviation increase in historical returns from the prior 3 months is associated with a 0.32 percentage point higher probability of a stock being selected ($p < 0.01$; Appendix Table A.8).

This preference for recent winner stocks is reflected in subjects’ asset and sector allocations. They disproportionately allocate funds to stocks in the information-technology sector, which delivered strong returns in the three months preceding the experiment. Their largest allocations are to Nvidia (NVDA, 9.8% of total allocations), Apple (AAPL, 7.1%), Microsoft (MSFT, 5.4%), and Amazon (AMZN, 3.4%) (Appendix Figure A.3). These four stocks experienced annualized historical excess returns ranging from 42% to 83% in the prior three months, compared with 3.9% for the equal-weighted average of all stocks in the S&P 500.

Taken together with the evidence from Greenwood and Shleifer (2014) that investors often extrapolate past returns when forming expectations about future returns, these findings suggest that subjects are attentive to the return component of Sharpe ratios. However, a strategy of favoring stocks with high historical returns was not effective at increasing realized returns over the evaluation window (Panel (a) of Appendix Table A.7). The coefficients relating historical returns and allocations to the information technology sector to realized returns are negative in all specifications and statistically significant in a subset, suggesting some reversal of prior return patterns. Although performance over one evaluation window cannot establish investing skill, this pattern is consistent with the argument in Section 2.2 that identifying assets with higher expected returns is difficult for subjects.

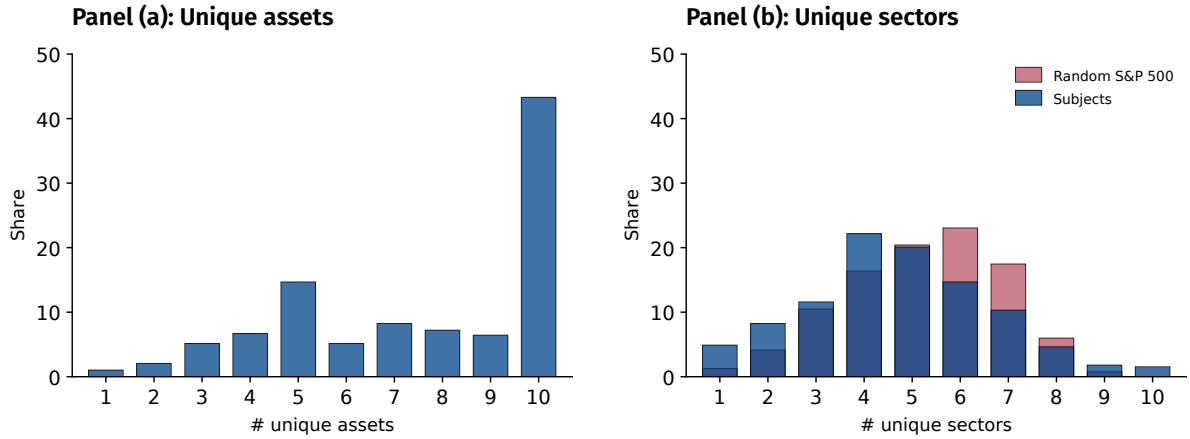
One concern is that subjects’ stock choices reflect adherence to the advice of LLMs or the copying of their suggested asset allocations. To investigate this, we compare human allocations with 100 portfolios elicited from GPT-5.2 using identical task instructions on the same day as the experiment. Human and LLM allocations differ substantially, despite a shared emphasis on the information technology sector (Appendix Figure A.4). For example, the LLMs allocate 12.1% of their total funds to Berkshire Hathaway Inc Class B (BRK.B), 7.4% to Johnson & Johnson (JNJ), and 6.3% to Procter & Gamble (PG). The corresponding shares in human portfolios are substantially smaller at 2.3%, 1.3%, and 1.3%, respectively. While we cannot rule out the selective use of LLM advice, these differences suggest that subjects were not systematically copying the portfolios produced by LLMs. This interpretation is further supported by the differences in portfolio breadth that we document in the next subsection.

5.1.3 Portfolio risk and diversification

Subjects’ choices and behavior suggest that they are also attentive to portfolio risk. In terms of individual asset choice, a one-standard-deviation increase in a stock’s volatility over the three months preceding the experiment is associated with a 0.71 percentage point lower probability of selection, controlling for historical returns and other stock features ($p < 0.01$; Appendix Table A.8).

Additionally, participant allocations tend to be distributed across several stocks. 43% of participants allocate positive weight to the maximum allowed number of 10 assets in Round

Figure 6: Diversification of individual Round-1 portfolios: unique assets and sectors



Notes: This figure displays the distributions of the numbers of unique assets and sectors held by subjects in Round 1. The height of each blue bar corresponds to the share of subjects who held that number of unique assets or sectors. Panel (a) is based on the number of unique S&P 500 tickers that receive a positive allocation in a portfolio, while Panel (b) is based on the number of unique GICS sectors represented in a portfolio. These distributions include all 388 subjects who submitted a valid Round-1 portfolio. In Panel (b), the red bars report a benchmark distribution of unique sector allocations obtained by drawing an asset count from the empirical distribution in Panel (a) and selecting that number of distinct S&P 500 tickers at random from the asset list.

1. The average subject holds 7.6 unique assets (Figure 6a). This is consistent with subjects seeking diversification by spreading their allocations across several assets.

However, their diversification across sectors is more limited. Their portfolios span only 4.7 unique sectors on average, slightly below the 5.2 sectors that random selection of assets produces while holding the number of assets constant (red benchmark in Figure 6b). The preference for information-technology stocks documented above contributes to this pattern.

Our measures of concentration and between-asset correlations offer further evidence that subjects take a simplistic approach to diversification. The HHI across assets is substantially above its theoretical minimum under the ten-stock limit, while the HHI across sectors exceeds a random-selection benchmark (Figure A.6). Meanwhile, the historical average pairwise correlation of subjects' holdings is 0.26, compared with 0.22 under a uniform draw from the S&P 500 and 0.21 under a draw weighted by how often subjects actually chose each asset (Figure A.7).

These outcomes are strongly associated with realized portfolio volatility. Panel (b) of Appendix Table A.7 shows that the weighted average historical asset risk, HHI, and the weighted historical average pairwise correlation of a portfolio are each positively and statistically significantly associated with its realized volatility over the evaluation window. When included jointly, they explain 59% of the variation in realized volatility across Round-1 portfolios. These results are consistent with the argument in Section 2.2 that investors can manage portfolio risk through asset selection and diversification.

Comparisons of portfolio breadth between human and LLM portfolios provide further evidence that subjects were not systematically copying the portfolios suggested by LLMs.

We find that human participants are 21 percentage points more likely to use all ten slots than LLM portfolios and are 45 percentage points less likely to hold six through nine stocks (Appendix Figure A.5 and Table A.9).

Taken together, these findings are consistent with subjects allocating attention to both the return and risk components of the incentivized Sharpe ratio. We now obtain our first result:

Result 1. *In Round 1, subjects’ strategies appear to target both portfolio returns and risk. Their approaches to doing so are naive in certain respects.*

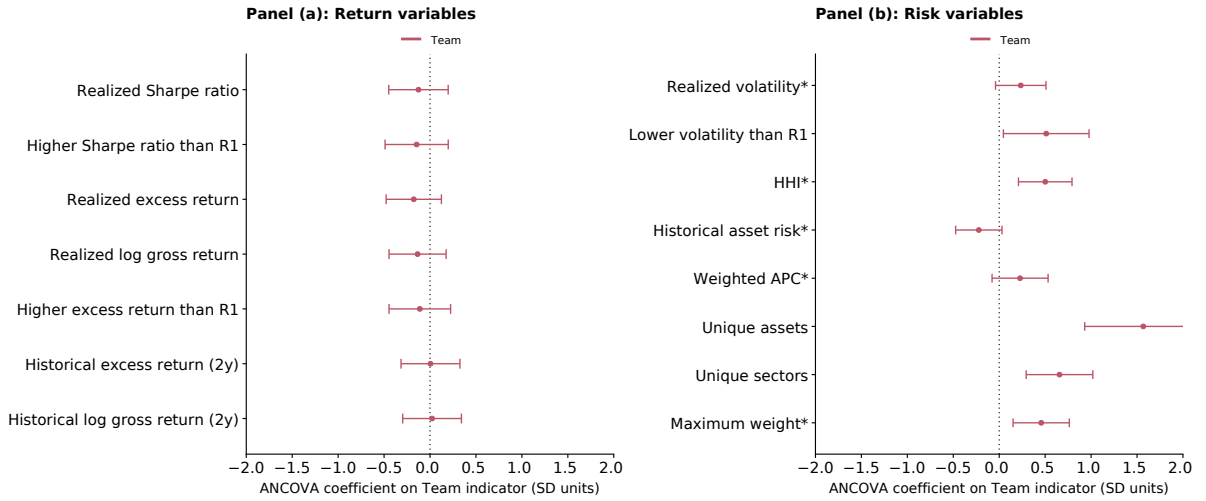
5.2 Teams versus individuals

We first compare the decision-making of individuals with joint decision-making as practiced by *Teams*. This comparison captures the combined effects of composition, communication, and collaborative aggregation and allows us to test whether beneficial or adverse effects dominate in our setting. All participants make decisions alone in Round 1, while *Crowd* subjects continue to do so in Round 2, albeit with awareness that they will be evaluated based on the performance of a pooled portfolio. As such, we are faced with two potential control groups. In our primary specification as written in equation (15), we use the average individual performance of *Crowd* pairs in Round 2; this specification avoids confounding from round effects. In complementary analyses, we also report results from within-pair decompositions (equation (18)), which compare each pair’s final portfolio with its own Round-1 individual benchmark. The main advantages of the within-pair comparison are that it sets up an intuitive comparison between individuals and *Teams* and that it enhances the comparability of our results with prior studies that use similar designs. As we report below, our findings do not differ much based on the chosen comparison group.

Result 2. *Team portfolios are more diversified and less volatile than those of individuals working alone, but they do not earn systematically higher returns.*

Figure 7 reports estimates from our primary ANCOVA specification. We find that *Team* portfolios are directionally lower in realized volatility by 0.23 SDs ($p = 0.09$) and 0.50 SDs less concentrated ($p < 0.01$) than the individual portfolios of *Crowd* members in Round 2. Assignment to the *Teams* condition also increases the probability that a pair realizes lower volatility than its own Round-1 individual benchmark by 19 percentage points ($p = 0.031$). In terms of alternative measures of portfolio risk that are independent of any particular evaluation window, *Team* portfolios also contain 2.79 more unique assets and 0.87 more sectors (both $p < 0.01$), and their largest allocation to any single asset in their portfolio is 5 percentage points smaller ($p < 0.01$). These structural measures provide complementary evidence consistent with greater portfolio diversification contributing to the reduction in realized volatility. Historical asset risk is higher in *Team* portfolios by 0.22 SDs ($p = 0.085$).

Figure 7: *Teams* versus individuals, return and risk outcomes



Notes: This figure reports ANCOVA estimates comparing Round-2 *Team* portfolios with the Round-2 individual benchmarks of *Crowd* members. Each point is a coefficient estimate for the outcome shown on the vertical axis. Panel (a) reports coefficient estimates for return outcomes, while Panel (b) reports them for risk outcomes. Outcomes marked with * are sign-reversed such that positive coefficients indicate improvement. All regressions control for the corresponding Round-1 individual benchmark. The sample contains 66 *Team* pairs and 64 *Crowd* pairs for most outcomes. Estimates are reported in SD units. The bars are 95% confidence intervals obtained using robust (HC1) standard errors.

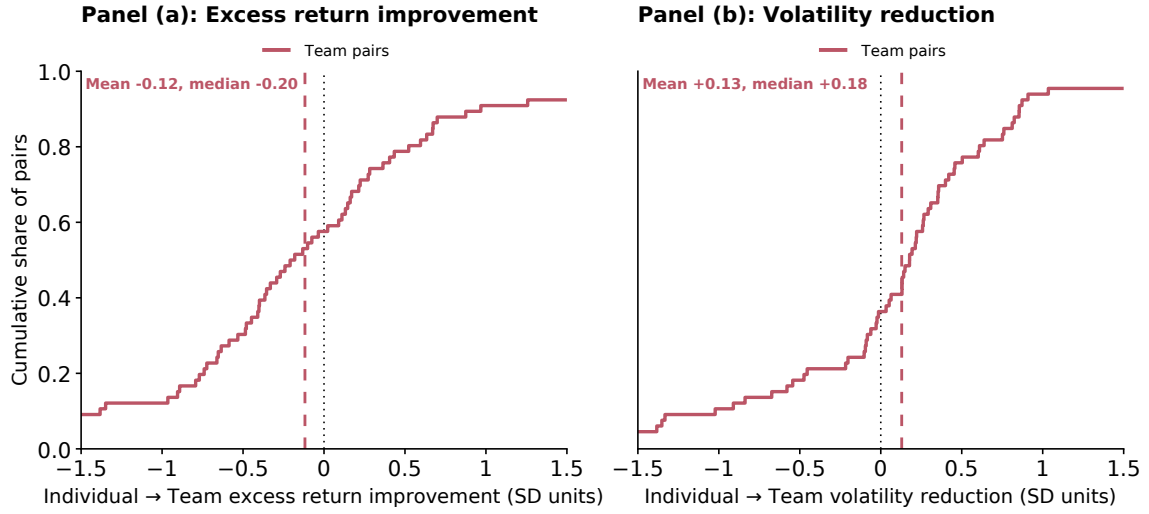
In contrast, we detect no significant effect on Sharpe ratios, realized returns, or any other primary outcome related to portfolio returns ($p > 0.10$). The corresponding coefficient estimates and standard errors are reported in Appendix Table A.3.

The main conclusions from this and the subsequent comparisons remain robust to extending the evaluation window by nearly two months and winsorizing outcomes at the 1st/99th and 5th/95th percentiles. As reported in Appendix Table A.5, the precision of some individual estimates varies across specifications.

Figure 8 complements our primary estimates by plotting the distribution of changes from each *Team*'s Round-1 individual benchmark to its realized Round-2 portfolio. These changes encompass the contributions of our three key features together with any round effects. This figure is particularly informative for examining dispersion and between-pair heterogeneity in performance. We observe that returns increase for 42% of *Team* pairs, while volatility falls for 64% of them. The mean (0.13 SD) and median (0.18 SD) reductions in realized volatility are similar in magnitude to the ANCOVA estimates. Thus, while a majority of *Teams* were able to achieve reductions in portfolio volatility, they were not successful at selecting stocks with high excess returns.

As a benchmark, we compare the performance of individuals working alone in Round 2 and *Teams* with 100 LLM portfolios generated on the day of the experiment (Appendix Figure A.8). The LLM portfolios achieve higher mean excess returns and lower mean realized volatility than both human comparison groups over the evaluation window. These findings suggest that there may be scope to improve upon portfolio choice among retail investors.

Figure 8: *Team* portfolios, changes from the Round-1 individual benchmark



Notes: This figure displays the cumulative distributions of changes in outcomes from the Round-1 individual benchmark to the Round-2 group portfolio among *Team* pairs. Panel (a) reports changes in realized excess returns, while Panel (b) reports changes in realized volatility. The sign of the change in volatility is reversed such that positive values indicate lower volatility. Changes are reported in SD units. The sample contains 66 *Team* pairs with valid Round-2 portfolios. The x -axis is truncated at ± 1.5 SD, although the cumulative distributions and summary statistics use all observations. The dashed lines correspond to the mean change among pairs.

Together with the differences in portfolio breadth, they also indicate that subjects' choices differ systematically from those generated by LLMs.

Sections 5.3–5.5 use the *Crowd* and *Panel* arms to decompose the differences between individual and joint decision-making into the separate contributions of composition, communication, and collaborative aggregation.

5.3 The composition effect: individuals versus Crowds

We first isolate the direct compositional effect of increasing the size of the decision unit from one to two and pooling their inputs, without adding any interactive elements to the group decision-making process. Our primary comparison is between the pooled Round-2 portfolio of a *Crowd* pair and the average outcome of its members' unpooled Round-2 portfolios, as specified in equation (16). The only difference between these types of portfolios is the application of the portfolio pooling procedure, which subjects were aware of but which did not otherwise affect their experience. The particular algorithm we describe in Section 3 is ideal for isolating direct composition effects as it is transparent, easy for subjects to understand, and treats each member's inputs symmetrically. Because each individual can select up to ten stocks, the pooled portfolio can contain up to twenty distinct assets. This expansion in access to inputs allows for broader diversification, which we consider to be part of the direct effect of composition. Moreover, it is analogous to other tasks from the WoC literature, such as forecasting and estimation, where benefits from aggregation are

also driven by an increase in the number of independent judgments available for pooling.

Result 3. *Mechanically pooling individual portfolios produces more diversified and less volatile group portfolios.*

Panel (a) of Figure 9 reports estimates from our primary specification. We find that pooled portfolios are 0.46 SDs lower in realized volatility and 0.89 SDs less concentrated than unpooled portfolios in Round 2 (both $p < 0.01$). Pooling also increases the probability that a pair realizes lower volatility than its own Round-1 individual benchmark by 39 percentage points ($p < 0.01$). The corresponding coefficient estimates and standard errors are reported in Appendix Table A.4.

Panel (b) of Figure 9 complements our primary estimates by plotting the distribution of changes from each *Crowd* pair's Round-1 individual benchmark to its pooled Round-2 portfolio. We observe that volatility falls for 84% of pairs, with mean and median reductions in realized volatility of 0.38 SDs and 0.30 SDs, respectively. Mean HHI is also 43% lower than the Round-1 individual benchmark. Appendix Figure A.9 shows similar distributions of gains in the effective number of holdings from pooling Round-1 portfolios across all pairs and from pooling Round-2 portfolios among *Crowds*.

Consistent with Prediction 1, all pooled *Crowd* portfolios realize lower concentration and volatility than their Round-2 individual benchmark. The volatility result holds even though these portfolios are invested without rebalancing, in contrast with the fixed-weight assumption used in the framework.

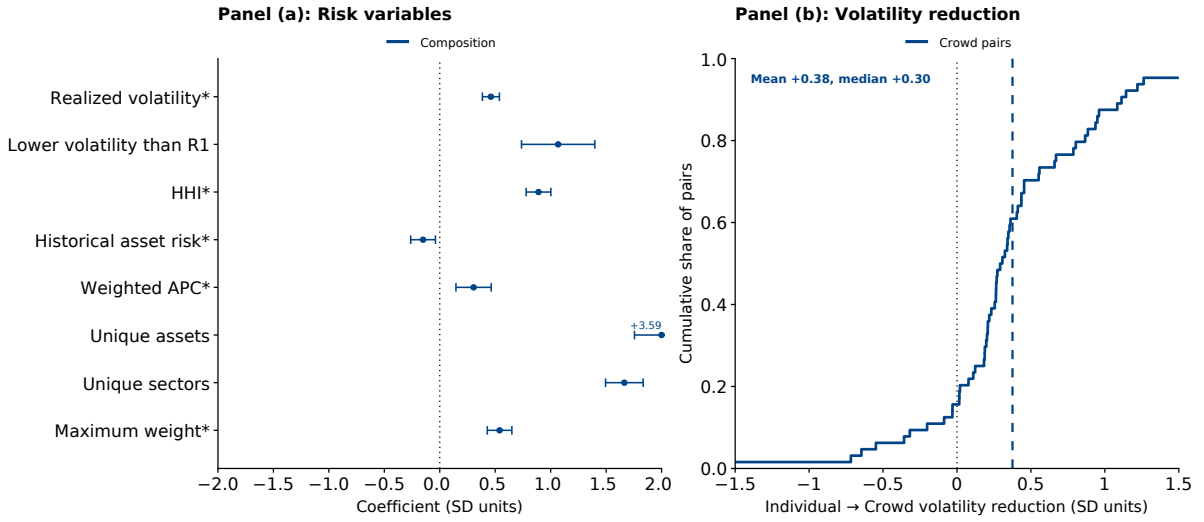
We also find that pooled portfolios hold less correlated assets, with APC falling by 0.32 SDs ($p < 0.01$). Historical asset risk increases by 0.15 SDs ($p < 0.01$). These findings are consistent with the reduction in volatility being driven by the adoption of broader and less correlated portfolios, rather than a shift toward individually safer stocks.

5.4 The communication effect: Crowds versus Panels

We next isolate the effect of communication by comparing *Crowds* with *Panels*. In both modes, each member of a pair submits an individual portfolio and the two portfolios are combined into a group portfolio using an identical pooling algorithm. The only difference is that *Panel* members can communicate with their partner and view each other's Round-1 portfolios. Our primary comparison is between the pooled Round-2 portfolios of these two decision modes, with controls for the corresponding Round-1 individual benchmark as specified in equation (17). This comparison therefore identifies the effect of allowing members to exchange and discuss their inputs while holding group size and the aggregation procedure fixed.

Result 4. *Communication undermines portfolio diversification and reduces the likelihood that pairs realize lower volatility than their individual benchmark.*

Figure 9: *Crowds* versus individuals, portfolio risk outcomes



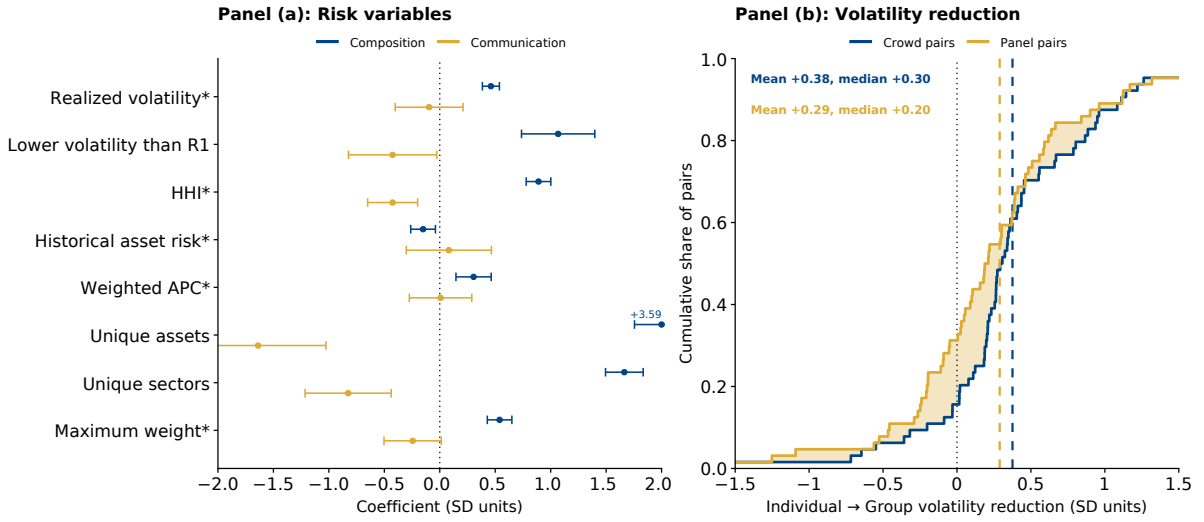
Notes: This figure reports the effect of composition on portfolio risk outcomes. Each point in Panel (a) is a coefficient estimate for the outcome shown on the vertical axis. Composition estimates compare pooled *Crowd* portfolios with their corresponding Round-2 individual benchmarks. Outcomes marked with * are sign-reversed such that positive coefficients indicate improvement. Composition regressions include only the pooling indicator and compare one pooled and one unpooled observation from each pair; all pre-treatment variables are held constant by construction. The sample contains 64 *Crowd* pairs for most outcomes. Estimates are reported in SD units. The bars are 95% confidence intervals obtained using clustered standard errors. Panel (b) displays the cumulative distribution of changes in realized volatility from the Round-1 individual benchmark to the Round-2 group portfolio among *Crowd* pairs. The sign of the change in volatility is reversed such that positive values indicate lower volatility. Changes are reported in SD units. The x -axis is truncated at ± 1.5 SD, although the cumulative distribution and summary statistics use all observations. The dashed line corresponds to the mean change among pairs.

Panel (a) of Figure 10 reports estimates from our primary ANCOVA specification. We find that *Panel* portfolios are directionally higher in realized volatility than *Crowd* portfolios, although this estimate is imprecise (0.10 SDs, $p = 0.54$), and 0.43 SDs more concentrated ($p < 0.01$). Communication reduces the probability that a pair realizes lower volatility than its own Round-1 individual benchmark by 16 percentage points ($p = 0.036$). *Panel* portfolios also contain 2.92 fewer unique assets and 1.10 fewer sectors (both $p < 0.01$). The corresponding coefficient estimates and standard errors are reported in Appendix Table A.4. Consistent with Prediction 3, we detect no corresponding effect of communication on realized excess returns (-0.12 SDs, $p = 0.37$).

Panel (b) complements these estimates by plotting changes from each pair's Round-1 individual benchmark to its Round-2 portfolio. We observe that volatility falls for 69% of *Panel* pairs, consistent with the 16-percentage-point drop identified in the main specification. The mean and median reductions in realized volatility and concentration are also similarly attenuated for *Panels* relative to *Crowds*.

Portfolio convergence. We next investigate the mechanisms behind the adverse effect of communication on portfolio risk. Consistent with Prediction 2, communication increases the similarity of partners' portfolios: the median change in portfolio cosine similarity between

Figure 10: *Panels* versus *Crowds*, portfolio risk outcomes



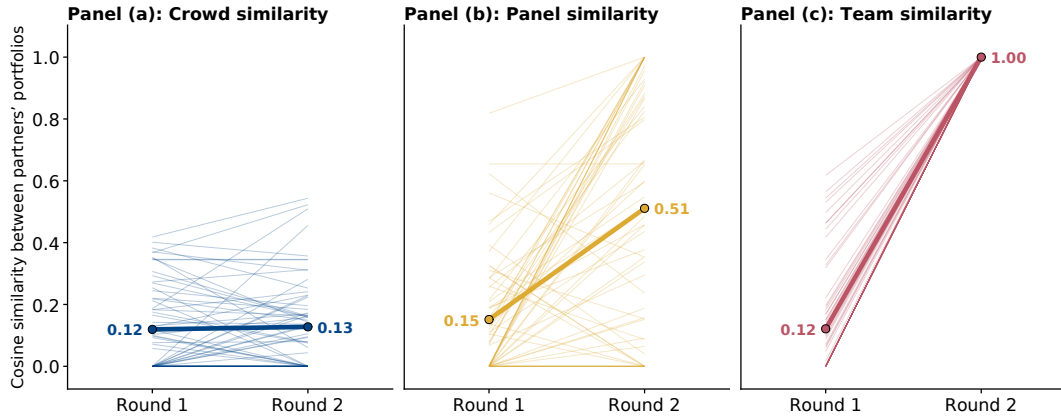
Notes: This figure reports the effects of composition and communication on portfolio risk outcomes. Each point in Panel (a) is a coefficient estimate for the outcome shown on the vertical axis. Composition estimates compare pooled *Crowd* portfolios with their corresponding Round-2 individual benchmarks. Communication estimates compare *Panel* portfolios with *Crowd* portfolios. Outcomes marked with * are sign-reversed such that positive coefficients indicate improvement. Composition regressions include only the pooling indicator and compare one pooled and one unpooled observation from each pair; all pre-treatment variables are held constant by construction. Communication regressions control for the corresponding Round-1 individual benchmark. The sample contains 64 *Crowd* pairs and 64 *Panel* pairs for most outcomes. Estimates are reported in SD units. The bars are 95% confidence intervals obtained using clustered standard errors for composition estimates and robust (HC1) standard errors for communication estimates. Panel (b) displays the cumulative distributions of changes in realized volatility from the Round-1 individual benchmark to the Round-2 group portfolio among *Crowd* and *Panel* pairs. The sign of the change in volatility is reversed such that positive values indicate lower volatility. Changes are reported in SD units. The *x*-axis is truncated at ± 1.5 SD, although the cumulative distributions and summary statistics use all observations. The dashed lines correspond to the mean change among pairs in each decision mode. The shaded area between the curves highlights the difference between the two distributions.

Rounds 1 and 2 is 0.34 for *Panel* pairs, compared to 0.00 for *Crowds* (Figure 11). Approximately 70% of *Panel* pairs submit more similar portfolios in Round 2 than in Round 1.

The behavior underlying this convergence is that *Panel* members replace some of their Round-1 holdings with assets and sectors held by their partner. *Panel* members drop an average of 3.76 assets and 1.30 sectors between rounds and adopt 1.60 assets and 0.75 sectors previously held by their partner. They also add 1.70 assets and 0.30 sectors that were held by neither member. These revisions reduce the number of holdings within individual portfolios and increase overlap between partners. The net effect is a decrease of 3.1 unique assets and 1.2 unique sectors in the average Round-2 *Panel* portfolio relative to the union of members' Round-1 portfolios (Figure 12). These effects strengthen with the intensity of communication (Appendix Table A.10).

The net negative effect of communication suggests that the change in average decision quality is insufficient to counteract the reduction in the pooling benefit (equation (7)). To investigate this further, we test whether portfolios converge more strongly in the direction of partners who hold better Round-1 portfolios. In separate OLS regressions, we find no

Figure 11: Portfolio convergence between Round 1 and Round 2



Notes: This figure plots within-pair changes in portfolio similarity between Rounds 1 and 2 separately for *Crowd* pairs in Panel (a), *Panel* pairs in Panel (b), and *Team* pairs in Panel (c). Each thin line represents the change in similarity for one pair, while the endpoints of each bold line are the mean levels of similarity within a given decision mode and round. Portfolio similarity is measured as the cosine similarity between partners' allocation weight vectors in a given round. It ranges from 0 when the partners' portfolios have no holdings in common to 1 when they are identical. *Team* portfolio similarity equals 1 in Round 2 by construction, as partners submit one joint portfolio in this decision mode. The sample contains 64 *Crowd* pairs, 63 *Panel* pairs, and 66 *Team* pairs.

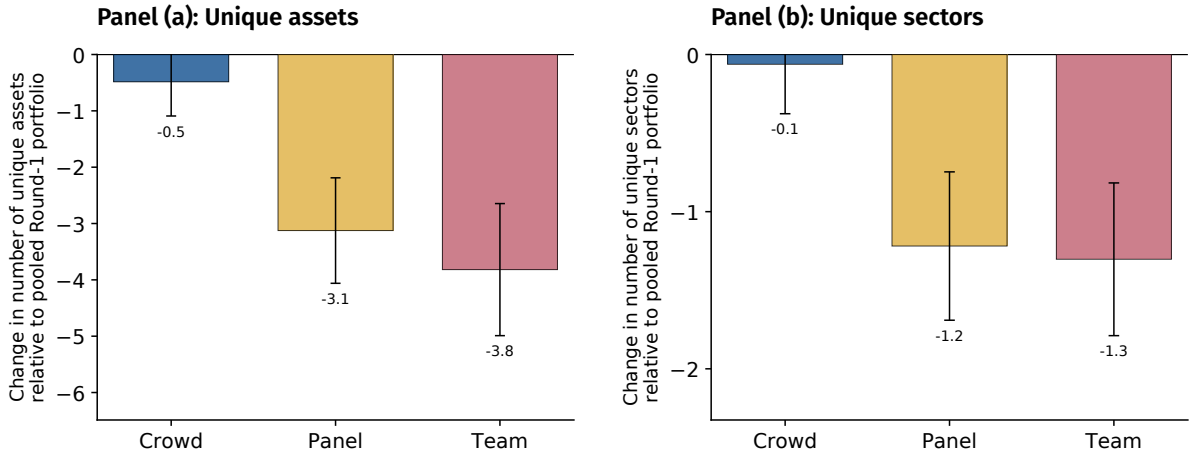
significant association between the extent of movement toward a partner's Round-1 portfolio and the realized Sharpe ratio ($p = 0.75$), excess returns ($p = 0.99$) and realized volatility ($p = 0.15$) of that portfolio.

Reallocation toward high past-return stocks. Portfolio revisions also change the weighting of stocks in subjects' portfolios. On average, pairs in all three decision modes increase their exposure to stocks with higher historical returns between Round 1 and Round 2. However, the mechanisms behind this increased exposure are somewhat different for *Panels* relative to *Crowds*. We decompose the change in the average historical returns of a pair's holdings between Rounds 1 and 2 into the separate contributions of newly added stocks, dropped stocks, and changes in the weight of stocks retained from Round 1 (Panel (a) of Figure 13).

In both modes, subjects retain stocks with higher two-year historical returns than those they drop; the size of this gap does not differ significantly between modes (Appendix Table A.11). However, *Panels* substantially increase the historical returns of their holdings by reweighting stocks retained from Round 1. This component contributes little among *Crowds*.

Attention to returns. Taken together, these findings are consistent with communication inducing a shift in subjects' strategies away from managing portfolio risk toward maximizing expected returns. Subjects' self-reports and chat logs provide further support for this interpretation. Relative to the *Crowd* condition, *Panel* subjects report 8 points higher levels of attention and effort directed toward maximizing returns versus minimizing risk in their portfolio ($p < 0.05$; Panel (b) of Figure 13). Furthermore, stocks that are mentioned

Figure 12: Changes in portfolio breadth by decision mode



Notes: This figure reports changes in the numbers of unique assets and sectors between a pair's Round-1 pooled portfolio and their Round 2 group portfolio. Panel (a) reports changes in the number of unique assets, while Panel (b) reports changes in the number of unique GICS sectors. The height of each bar corresponds to the mean change within a given decision mode. Negative values indicate a decline in the number of unique assets or sectors in Round 2. The sample contains 64 *Crowd* pairs, 64 *Panel* pairs, and 66 *Team* pairs. Bars are 95% confidence intervals around the sample mean.

in group discussions have higher past returns than the stocks in either partner's Round-1 portfolio that go unmentioned (Appendix Table A.12). This evidence parallels the concern stated in Section 2.2 that discussions focused on asset returns may draw subjects' attention away from opportunities to reduce portfolio risk. Relating this evidence to Prediction 3, we detect no significant effect of communication on realized volatility (-0.09 SDs, $p = 0.56$) or concentration (~ 0 SDs, $p = 0.99$) in *Panel* members' pre-pooling portfolios. These estimates provide no evidence of improvements in individual portfolio quality that could offset the reduction in pooling benefits.

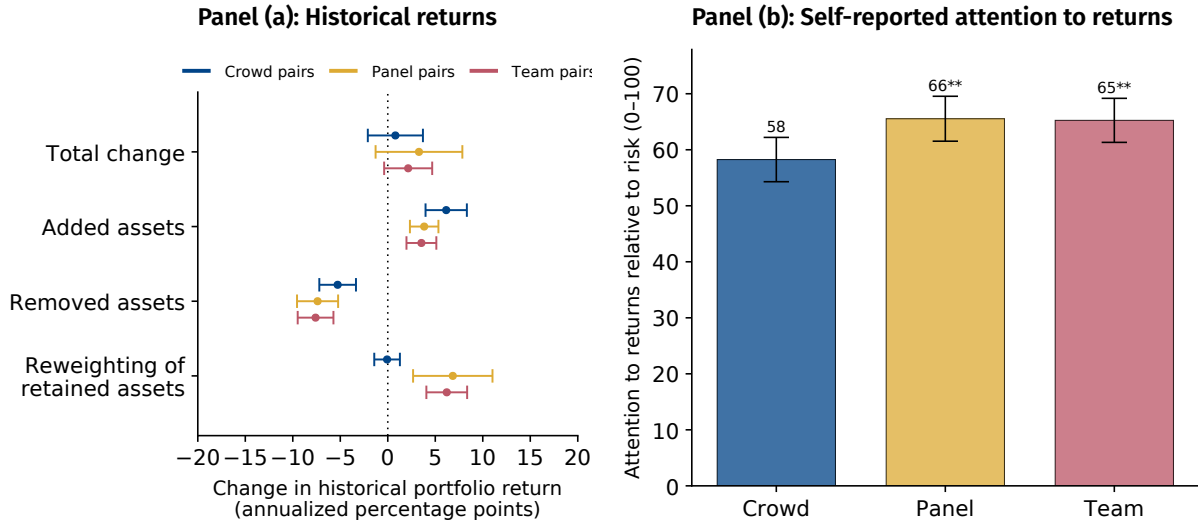
5.5 The aggregation effect: Panels versus Teams

Finally, we isolate the effect of collaborative aggregation by comparing *Panels* with *Teams*. In both modes, partners can communicate and view each other's Round-1 portfolios. However, the key difference is that *Team* pairs aggregate their individual judgments by coordinating on a single portfolio through a shared interface, in contrast with the pooling algorithm imposed upon *Panel* pairs. Our primary comparison is between the joint Round-2 portfolios of *Team* pairs and the pooled Round-2 portfolios of *Panel* pairs, with controls for the corresponding Round-1 individual benchmark as specified in equation (17).

Result 5. *We detect no significant additional effect of collaborative aggregation on any measure of portfolio risk.*

Panel (a) of Figure 14 reports estimates from our primary ANCOVA specification. We find that *Team* portfolios are directionally higher in realized volatility than *Panel* portfolios

Figure 13: Portfolio revisions and attention to returns

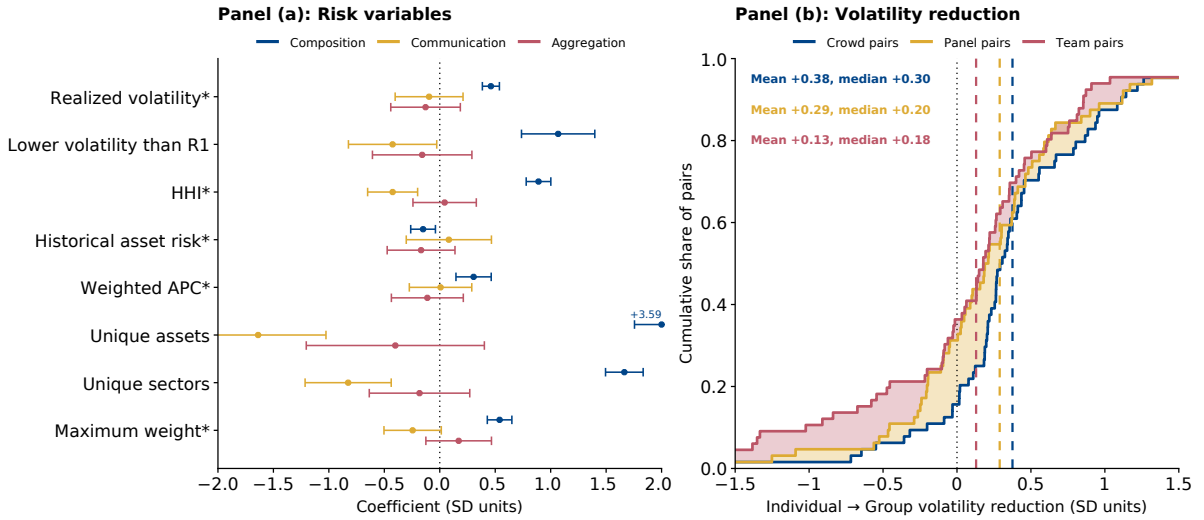


Notes: This figure reports differences in exposure to stocks with high historical returns and subjects' self-reported attention to returns between decision modes. Panel (a) decomposes the average change in annualized historical returns from pairs' pooled Round-1 portfolios to their Round-2 group portfolios into three separate components. "Added assets" captures the change attributable to allocations to stocks that are present in the Round-2 group portfolio but absent from the pooled Round-1 portfolio. "Removed assets" captures the change attributable to assets that are absent from the Round-2 group portfolio but present in the pooled Round-1 portfolio. "Reweighting of retained assets" captures the change attributable to changes in allocations to stocks held in both portfolios. Each point reports the mean contribution of the corresponding component within a given decision mode. Changes are reported in annualized percentage points. Historical returns are measured over the two years preceding the experiment. Panel (b) reports the mean Round-2 self-reported level of attention to returns relative to risk (0 to 100) in a given decision mode. The sample contains 64 *Crowd* pairs, 64 *Panel* pairs (63 in Panel (a)), and 66 *Team* pairs. The bars are 95% confidence intervals obtained using robust (HCl) standard errors in Panel (a) and clustered standard errors in Panel (b). Stars indicate differences from the *Crowd* condition. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

(0.13 SDs) and slightly less concentrated (0.04 SDs); these estimated effects point in competing directions for portfolio risk, and neither is statistically significant (both $p > 0.10$). The change in a pair's probability of realizing lower volatility than its Round-1 individual benchmark that is attributable to collaborative aggregation is also small and statistically insignificant (-6 percentage points, $p = 0.49$). We detect no significant differences across the remaining measures of portfolio risk and diversification. The reductions in realized volatility relative to the Round-1 individual benchmark that are plotted in Panel (b) of Figure 14 show a similar pattern. We likewise detect no significant effect on realized excess returns (0.00 SDs, $p = 1.00$).

These null results are consistent with influence in the *Team* aggregation procedure having no systematic association with decision quality (equation (10)) and with communication leaving little disagreement between partners about their preferred portfolios (Prediction 4). To test the first hypothesis, we measure the similarity of each pair's *Team* portfolio to the Round-1 portfolios of their members. Separate OLS regressions reveal no significant association between a portfolio's similarity and its realized Sharpe ratio ($p = 0.36$), excess returns ($p = 0.23$), and realized volatility ($p = 0.14$). Evidence consistent with the second hypothesis is reported in the *Panel* section of Figure 11.

Figure 14: *Teams* versus *Panels*, portfolio risk outcomes



Notes: This figure reports the effects of composition, communication, and collaborative aggregation on portfolio risk outcomes. Each point in Panel (a) is a coefficient estimate for the outcome shown on the vertical axis. Composition estimates compare pooled *Crowd* portfolios with their corresponding Round-2 individual benchmarks. Communication and aggregation estimates compare *Panel* portfolios with *Crowd* portfolios and *Team* portfolios with *Panel* portfolios, respectively. Outcomes marked with * are sign-reversed such that positive coefficients indicate improvement. Composition regressions include only the pooling indicator and compare one pooled and one unpooled observation from each pair; all pre-treatment variables are held constant by construction. Communication and aggregation regressions control for the corresponding Round-1 individual benchmark. The sample contains 64 *Crowd* pairs, 64 *Panel* pairs, and 66 *Team* pairs for most outcomes. Estimates are reported in SD units. The bars are 95% confidence intervals obtained using clustered standard errors for composition estimates and robust (HC1) standard errors for communication and aggregation estimates. Panel (b) displays the cumulative distributions of changes in realized volatility from the Round-1 individual benchmark to the Round-2 group portfolio among *Crowd*, *Panel* and *Team* pairs. The sign of the change in volatility is reversed such that positive values indicate lower volatility. Changes are reported in SD units. The *x*-axis is truncated at ± 1.5 SD, although the cumulative distributions and summary statistics use all observations. The dashed lines correspond to the mean change among pairs in each decision mode. The shaded areas between the curves highlight the differences between the *Crowd* and *Panel* distributions and the *Panel* and *Team* distributions.

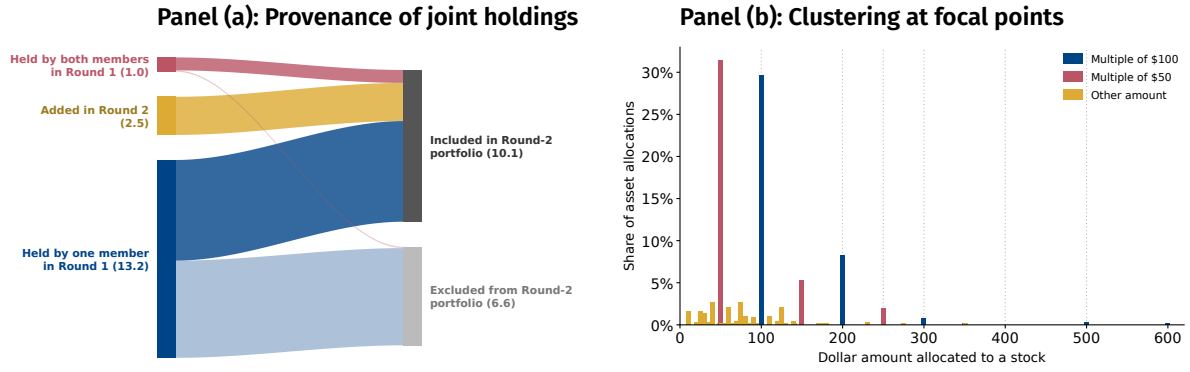
How Teams aggregate (descriptive). Nonetheless, descriptive evidence on *Team* aggregation practices could help us better understand subjects' behavior. For each *Team*, we classify every ticker by whether it was held by both members, held by one member, or held by neither member in Round 1 (Figure 15a).

We find that *Teams* almost always keep assets that both members held in Round 1, but drop roughly half of the stocks held by only one member. This pattern is consistent with the retention of salient stocks and the selective removal of each partner's idiosyncratic holdings documented in our discussion of communication.

Moreover, allocation amounts exhibit substantial clustering at salient focal points. 78% of *Team* holdings are entered at multiples of \$50 and 39% at multiples of \$100. 32% of *Teams* split their \$1,000 into equal shares (Figure 15b).

Finally, several descriptive measures of how *Teams* aggregate are strongly associated with final portfolio diversification. Every ten additional edits a *Team* makes to its portfolio are associated with 2.95 more unique assets ($p < 0.01$) and a lower HHI ($\beta = -533$ points,

Figure 15: Sources of Team holdings and patterns in allocation amounts



Notes: This figure reports aggregation and allocation patterns among *Teams* in Round 2. The left side of Panel (a) classifies each asset by whether it was held by one partner in Round 1, held by both partners, or held by neither partner and newly added to the Round-2 group portfolio. The right side classifies assets by whether they are included in or excluded from the Round-2 group portfolio. The numerical labels to the side report the average numbers of assets falling into each category. Panel (b) displays the distribution of asset allocations made by *Teams* in Round 2. The height of each bar corresponds to the share of all asset allocations equal to a given dollar amount.

$p < 0.01$). The edit log shows that the largest reductions in concentration occur during a portfolio's earliest edits. The first five edits reduce HHI by approximately 1,675 points on average, compared with 494 points for edits six through ten and 62 for edits beyond the twentieth (Appendix Figure A.10). This is partly because HHI is nonlinear in portfolio weights. The number of unique assets behaves differently: it rises by about 0.35–0.41 assets per edit throughout the whole sequence.

5.6 Synthesis: the pooling–revision decomposition

The preceding comparisons show that direct composition effects reduce portfolio risk, while communication exacerbates it and collaborative aggregation produces little additional change. We now bring these findings together to assess how much of the potential compositional benefit is retained by *Teams* engaging in joint decision-making. Our primary comparison is between the joint Round-2 portfolios of *Team* pairs and the pooled Round-2 portfolios of *Crowd* pairs, with controls for the corresponding Round-1 individual benchmark as specified in equation (17). We then use the within-pair decomposition in equation (18) to examine how groups' decisions in Round 2 affect the share of the potential compositional benefit from pooling that they capture. Together, these comparisons help to explain how *Teams* can outperform individuals while failing to make full use of the diversification opportunities provided by their members' inputs.

Result 6. *Team portfolios are substantially underdiversified relative to what they could have achieved with mechanical pooling. Communication accounts for most of the estimated difference in concentration.*

Consistent with the negative effect of communication and the small estimated effect

of collaborative aggregation, *Teams* underperform *Crowds* on portfolio risk. Our ANCOVA results show that *Team* portfolios have 0.22 SDs higher realized volatility ($p = 0.098$) and are 21 percentage points less likely to reduce volatility relative to their Round-1 individual benchmark ($p < 0.01$). They are also 0.40 SDs more concentrated and contain 3.61 fewer unique assets and 1.34 fewer sectors (all $p < 0.01$). The corresponding coefficient estimates and standard errors are reported in Appendix Table A.4.

The point estimate of the difference between *Panels* and *Crowds* is equivalent to 44% of the *Team*–*Crowd* gap in realized volatility and 107% of the gap in concentration. The estimated effects of collaborative aggregation remain small and statistically insignificant across all reported measures of portfolio risk (Appendix Table A.4). These findings are consistent with communication driving the greater concentration of *Team* relative to *Crowd* portfolios.

To complement these comparisons between decision modes, we now examine within-pair changes in portfolio outcomes between Round 1 and Round 2. For *Crowds* and *Panels*, Table 1 decomposes these changes into (i) the initial pooling benefit from combining individual portfolios in Round 1, (ii) the change in that pooling benefit when revised individual portfolios are used in Round 2, and (iii) the change in the average quality of individual portfolios between rounds (equation (18)). Since *Team* pairs submit only one joint portfolio, components (ii) and (iii) are presented under one combined revision component.

Estimated initial pooling benefits in Round 1 are similar across decision modes. Among *Crowds*, the magnitudes of components (ii) and (iii) are small and insignificant across all outcomes. This is consistent with our assumption that making subjects aware that their portfolio will be pooled with their partner’s does not meaningfully change their behavior (Section 2). In line with increased portfolio similarity in Round 2, the pooling benefit among *Panels* falls by 1.92 annualized percentage points for volatility and 292 HHI points for concentration (both $p < 0.01$). Changes in the quality of individual submissions in the *Panel* condition are small and statistically insignificant for these two outcomes. Consistent with Prediction 3, the decline in the pooling benefit for these outcomes is significantly larger for *Panels* than *Crowds* (both $p < 0.01$). As in Section 5.4, these changes in individual portfolio quality are insufficient to offset the reduction in the pooling benefit.

Among *Teams*, the combined revision component reverses 73% of the initial pooling benefit for volatility and 47% for concentration. As explained above, we cannot determine how much of this reversal reflects changes in the pooling benefit versus the average quality of *Team* members’ hypothetical individually preferred portfolios.

One caveat is that these within-pair changes are potentially confounded by learning and changes in effort between rounds. However, the small and insignificant *Crowd* components suggest that the results in Table 1 are unlikely to be driven by such round effects.

Table 1: Decomposition of within-pair changes in portfolio quality, by decision mode

Dependent variable:	Realized volatility*	HHI*	Hist. asset risk*	Weighted APC*	Unique assets	Unique sectors	Maximum weight*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a). Crowd							
Total change (R1 indiv. → R2 group)	3.51*** (0.67)	806.59*** (72.91)	-0.64 (0.83)	0.02* (0.01)	6.21*** (0.32)	2.08*** (0.16)	0.06*** (0.01)
Initial pooling benefit (R1 indiv. → pooled)	4.00*** (0.29)	837.18*** (40.36)	-1.23*** (0.33)	0.02*** (0.01)	6.71*** (0.21)	2.14*** (0.10)	0.06*** (0.01)
Change in pooling benefit (R1 → R2)	0.31 (0.28)	19.47 (53.69)	0.17 (0.36)	0.00 (0.00)	-0.28 (0.17)	0.07 (0.11)	0.00 (0.01)
Change in average quality (R1 → R2)	-0.80 (0.77)	-50.06 (101.19)	0.42 (0.66)	0.00 (0.01)	-0.21 (0.16)	-0.13 (0.11)	0.00 (0.01)
Panel (b). Panel							
Total change (R1 indiv. → R2 group)	1.90 (1.21)	353.43*** (96.79)	-0.53 (0.99)	0.01 (0.01)	3.42*** (0.48)	1.04*** (0.21)	0.03*** (0.01)
Initial pooling benefit (R1 indiv. → pooled)	3.80*** (0.35)	720.29*** (49.67)	-1.07** (0.47)	0.02*** (0.00)	6.44*** (0.26)	2.19*** (0.11)	0.05*** (0.01)
Change in pooling benefit (R1 → R2)	-1.92*** (0.44)	-292.22*** (59.46)	0.76 (0.49)	-0.01* (0.01)	-2.52*** (0.37)	-0.90*** (0.16)	-0.02*** (0.01)
Change in average quality (R1 → R2)	0.02 (1.15)	-74.64 (79.63)	-0.23 (0.96)	0.00 (0.01)	-0.50*** (0.19)	-0.24 (0.15)	0.00 (0.01)
Panel (c). Team							
Total change (R1 indiv. → R2 group)	1.21 (1.00)	447.38*** (152.09)	-1.31** (0.62)	0.01 (0.01)	2.46*** (0.56)	0.67*** (0.22)	0.05*** (0.02)
Initial pooling benefit (R1 indiv. → pooled)	4.45*** (0.41)	844.45*** (85.97)	-0.57* (0.30)	0.03*** (0.01)	6.28*** (0.32)	1.98*** (0.11)	0.06*** (0.01)
Revision (R1 pooled → R2 group)	-3.24*** (1.05)	-397.07*** (142.64)	-0.73 (0.67)	-0.02** (0.01)	-3.82*** (0.59)	-1.30*** (0.24)	-0.01 (0.02)
Pair fixed effects	✓	✓	✓	✓	✓	✓	✓
Observations (pair-stages)	386	382	386	376	382	382	382
Clusters (pairs)	193	191	193	188	191	191	191

Notes: This table decomposes changes in portfolio risk outcomes from the Round-1 individual benchmark to the Round-2 group portfolio into separate components. Each coefficient reports the mean within-pair change for the outcome and component shown. Panels (a), (b) and (c) report estimates for *Crowd*, *Panel* and *Team* pairs, respectively. For *Crowds* and *Panels*, the total change can be decomposed into the initial pooling benefit and the changes in the pooling benefit and in the average quality of individual portfolios between rounds. For *Teams*, the latter two components cannot be separately identified and are combined into a single revision component. Realized volatility and historical asset risk are expressed in annualized percentage points. Outcomes marked with * are sign-reversed such that positive coefficients indicate improvement. Sample sizes vary by outcome and are reported in each column. All regressions include pair fixed effects. Clustered standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

5.7 Heterogeneity by pair composition

The primary estimates reported in the previous sections capture the average effects of each feature of joint decision-making. However, our within-pair decompositions reveal substantial heterogeneity in how pairs are affected by these features. To examine this heterogeneity, we first relate the size of a pair's pooling and revision components to their cognitive diversity. We measure cognitive diversity on three dimensions of inputs that vary in their domain-specificity: subjects' Round-1 asset allocations, their beliefs about the future path of the S&P 500, and their thinking style on the *Rational-Experiential Inventory*. We next examine

whether the effect of cognitive diversity differs across decision modes.

Result 7. *Pooling benefits are larger among pairs with less overlap in their Round-1 portfolios. Pairs with diverse beliefs about the future path of the S&P 500 achieve lower volatility and broader diversification than those with similar beliefs.*

Initial portfolio composition. The first component of cognitive diversity that we examine is the most domain-specific one: differences between partners' Round-1 portfolio allocations. In particular, we regress each pair's pooling and revision components on the cosine similarity between partners' Round-1 portfolios, controlling for belief diversity, thinking-style diversity, the decision mode, and the Round-1 individual benchmark (Table 2).

We find that a one-standard-deviation increase in initial portfolio similarity predicts a 0.11 SD smaller reduction in realized volatility and a 0.18 SD smaller reduction in concentration in the initial pooling benefit (both $p < 0.01$). These findings are consistent with the expressions in equation (13) relating the magnitude of pooling gains to portfolio similarity.

The corresponding associations of initial portfolio similarity with the revision component have the opposite sign and similar magnitudes (0.14 and 0.18 SDs, respectively; both $p < 0.01$) to those with the initial pooling benefit. As such, pairs with more distinct initial portfolios may benefit more from pooling, but their subsequent revisions largely offset the gains from this diversity.

Belief and thinking-style diversity. We next examine two dimensions of pair composition that are less mechanically related to the group portfolio: market beliefs and thinking styles.

As discussed in Section 3, we elicited forecasts from subjects about the level of the S&P 500 in three months' time. Appendix Figure A.11 plots the distribution of these forecasts net of the index level prevailing when each subject was asked. As such, the x -axis can be interpreted as each subject's prediction of the change in the level of the S&P 500 over the forecast period.

The median subject expects the index to rise by 136 points over the following three months, with 80% reporting positive expected market returns. However, forecasts exhibit considerable dispersion. We classify a pair as belief-diverse when one member's predicted change is above the sample median and the other's is at or below it, and as belief-similar otherwise. While belief diversity is not directly randomized, we argue that partner assignment generates conditionally random variation in belief diversity across pairs. This interpretation is supported by the balance in average S&P 500 forecasts across thinking-style composition arms in Table A.2 and by the similarity of our estimated effects of belief diversity between unconditional specifications and those that include controls for thinking-style scores.

We examine belief and thinking-style diversity using the same pooling and revision regressions as those estimated for initial portfolio composition (Table 2).

Table 2: Pair composition and the pooling and revision components

Dependent variable:	Realized volatility*	HHI*	Historical asset risk*	Weighted APC*	Unique assets	Unique sectors	Maximum weight*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a). Initial portfolio diversity							
Initial pooling benefit	0.11*** (0.01)	0.18*** (0.02)	-0.05* (0.03)	0.18*** (0.04)	0.51*** (0.04)	0.14*** (0.04)	0.25*** (0.02)
Revision component	-0.14*** (0.05)	-0.18*** (0.06)	0.03 (0.05)	-0.07 (0.07)	-0.39** (0.15)	-0.11 (0.09)	-0.20*** (0.06)
Panel (b). Belief diversity							
Initial pooling benefit	0.02 (0.03)	0.01 (0.02)	-0.09* (0.05)	0.13 (0.09)	0.03 (0.05)	0.00 (0.08)	0.06 (0.04)
Revision component	0.26** (0.12)	0.16 (0.12)	0.29** (0.12)	0.10 (0.14)	0.53* (0.29)	0.37** (0.17)	0.17 (0.12)
Panel (c). Thinking-style diversity							
Initial pooling benefit	0.02 (0.03)	0.00 (0.02)	0.03 (0.05)	-0.11 (0.09)	-0.06 (0.05)	-0.14* (0.08)	0.03 (0.04)
Revision component	0.08 (0.12)	0.05 (0.11)	-0.08 (0.13)	0.25* (0.14)	-0.04 (0.29)	0.21 (0.18)	-0.01 (0.12)
Decision mode indicators	✓	✓	✓	✓	✓	✓	✓
Round-1 individual benchmark	✓	✓	✓	✓	✓	✓	✓
Observations (pairs)	187	187	187	186	187	187	187

Notes: This table reports associations between pair composition and the size of the initial pooling benefit and revision components. Panels (a), (b) and (c) report coefficients for estimated associations with initial portfolio diversity, belief diversity, and thinking-style diversity, respectively. Separate regressions are run for each outcome and component. All regressions jointly include all three measures of compositional diversity, the decision mode, and the corresponding Round-1 individual benchmark. The initial pooling benefit is the pooled Round-1 portfolio outcome minus the Round-1 individual benchmark. The revision component is the Round-2 group portfolio outcome minus the pooled Round-1 portfolio outcome. Estimates are reported in SD units. Outcomes marked with * are sign-reversed such that positive coefficients indicate improvement. Sample sizes vary by outcome and are reported in each column. Observation counts apply to both components. Robust (HC1) standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

We detect no significant associations between a pair's belief diversity and the size of its initial pooling benefit at the 5% level. However, belief-diverse pairs realize more favorable revision components than belief-similar pairs across several measures of portfolio risk. In particular, belief diversity is favorably associated with realized volatility (0.26 SDs, $p = 0.031$), the number of unique sectors held in the group portfolio (0.37 SDs, $p = 0.033$), and historical asset risk (0.29 SDs, $p = 0.020$). Its association with the number of unique assets is significant only at the 10% level (0.53 SDs, $p = 0.071$).

Our most domain-general measure of cognitive diversity is not associated with comparable improvements. We detect no statistically significant effect of thinking-style diversity at the 5% level on either the pooling or revision component across the reported measures of portfolio risk (Table 2). Estimates are small and statistically insignificant for both realized volatility ($p = 0.45$ for pooling and $p = 0.50$ for revision) and concentration ($p = 0.78$ for pooling and $p = 0.65$ for revision).

Thinking-style scores are nonetheless associated with several individual outcomes. Sub-

jects with higher analytic thinking scores report greater confidence in their Round-1 portfolios and choose to reallocate smaller shares of them to an S&P 500 ETF (Figure A.12). Their initial portfolios also more closely resemble their pair's final Round-2 portfolio (Appendix Table A.13), which is consistent with them exerting greater influence over group decision-making. These associations suggest that thinking styles are relevant to individual behavior even though we detect no significant effects of thinking-style diversity on the pooling or revision components.

5.7.1 Composition as a moderator of communication and aggregation

In addition to its effects on the magnitude of the initial pooling benefit, group composition may also affect how partners communicate and aggregate their judgments. We next examine whether the relationship between belief diversity and portfolio outcomes differs across decision modes. In our main analysis, we report estimates of ANCOVA regressions that include treatment indicators for each decision mode and interact each of them with belief diversity (Appendix Table A.14). We control for the corresponding Round-1 individual benchmark, as well as our two other measures of cognitive diversity.

Result 8. *Belief diversity is associated with the most favorable changes in portfolio risk among communicating pairs and Teams in particular.*

In the combined sample of *Panel* and *Team* pairs, belief diversity is associated with lower volatility and greater diversification in the revision component. We detect no corresponding associations among non-communicating *Crowd* pairs (Appendix Tables A.15 and A.16).

These associations are strongest among *Teams*. Belief-diverse *Team* portfolios are 0.56 SDs lower in realized volatility ($p = 0.023$) and 0.79 SDs less concentrated ($p = 0.005$) than belief-similar *Team* portfolios. They also contain 4.68 more unique assets and 1.65 more sectors, while their largest allocation to any single asset is 9 percentage points smaller (all $p < 0.01$).

The estimated relationship between belief diversity and diversification is significantly more favorable among *Teams* than *Panels* for concentration, the number of unique assets and sectors, and the maximum weight assigned to any asset in the portfolio (all $p < 0.01$). *Team* portfolios are 0.45 SDs less concentrated than *Panel* portfolios among belief-diverse pairs ($p = 0.028$). The corresponding differences are not significant for realized volatility.

Taken together, these findings are consistent with the intensity of interaction operating as a moderator of the potential effect of belief diversity. Differences in beliefs may help partners preserve more of each other's independent inputs during interaction, especially when they must agree upon one joint portfolio despite differing beliefs. One caveat is that the relatively small sample sizes within each decision mode limit the precision of these estimates.

6 Conclusion

The primary contribution of the present study is to estimate the separate effects of composition, communication, and collaborative aggregation within a single randomized design. At a high level, we find that joint decision-making results in more diversified and less volatile portfolios than individual decision-making. This advantage is primarily driven by the compositional effect of pooling independent portfolios together. Communication attenuates these gains, while collaborative aggregation has little effect on average. Consequently, joint decision-making underperforms simple pooling on portfolio risk despite access to comparably diverse initial inputs and the ability to communicate and collaboratively aggregate.

We interpret these results through our conceptual framework and empirical decomposition. Two key predictions from the framework are that communication makes partners' portfolios more similar and that this convergence reduces the pooling benefit for certain portfolio risk outcomes. Our findings support both of these predictions. Whether communication is net beneficial for portfolio risk depends on whether it generates sufficient improvements in the average quality of individual portfolios to offset the loss in the pooling benefit. In our experiment, we find that it does not. Subjects' choices and self-reports are instead consistent with communication inducing them to prioritize maximizing expected returns over managing portfolio risk, despite evidence that identifying assets with higher expected returns is difficult for subjects.

The magnitude and significance of our main effects depend partly on pair composition. In particular, pairs with more distinct Round-1 portfolios see larger benefits from mechanical pooling. However, their subsequent choices largely offset these gains. Furthermore, pairs with diverse beliefs about the stock market are more successful at reducing portfolio risk. We also find evidence of heterogeneity in the effect of collaborative aggregation, as it is associated with significant improvements in some portfolio risk outcomes among belief-diverse pairs. These findings suggest that the value of interaction depends partly on the diversity of inputs brought by group members.

Several limitations qualify these conclusions. First, the realized return and volatility outcomes are measured over one three-month evaluation period. Although this evaluation window was pre-registered, our conclusions could change if outcomes were measured over alternative time periods. This limitation is partly addressed by our reporting of treatment effects on structural proxies for portfolio risk, such as HHI and average pairwise correlations, and our robustness check extending the evaluation window by nearly two months. A second limitation is that, due to the relatively small incentives we offer and the subject pool of retail investors, our conclusions may not generalize to investing decisions made by professional investors. Similar experiments in natural field settings with professional investors would therefore be a useful extension.

Nevertheless, these findings have implications for existing literatures and for the design

of groups in organizations. For settings where the diversity of opinions is valuable, including for investing, forecasting, and estimation tasks, eliciting individuals' initial judgments may be insufficient for preserving independent inputs, as members can simply choose to discard them during discussion. One possibility for future research is to allow members to communicate while retaining explicit weight on their independently elicited judgments in the final decision. This proposal follows from a central insight of our findings: forming a group with diverse ideas and inputs does not guarantee that its members will make full use of them.

References

- Adams, John C., Takeshi Nishikawa, and Ramesh P. Rao (2018).** "Mutual Fund Performance, Management Teams, and Boards." *Journal of Banking & Finance* 92: 358–68.
- Allinson, Christopher W., and John Hayes (1996).** "The Cognitive Style Index: A Measure of Intuition-Analysis for Organizational Research." *Journal of Management Studies* 33 (1): 119–35.
- Alternative Investment Management Association (2019).** "The Alternatives." Industry report. Alternative Investment Management Association (AIMA).
- Anderson, Cameron, and Gavin J. Kilduff (2009).** "Why Do Dominant Personalities Attain Influence in Face-to-Face Groups? The Competence-Signaling Effects of Trait Dominance." *Journal of Personality and Social Psychology* 96 (2): 491–503.
- Atanasov, Pavel, Phillip Rescober, Eric Stone, Samuel A. Swift, Emile Servan-Schreiber, Philip Tetlock, Lyle Ungar, and Barbara Mellers (2017).** "Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls." *Management Science* 63 (3): 691–706.
- Bär, Michaela, Alexander Kempf, and Stefan Ruenzi (2011).** "Is a Team Different From the Sum of Its Parts? Evidence From Mutual Fund Managers." *Review of Finance* 15 (2): 359–96.
- Barahona, Ricardo, Stefano Cassella, Kristy A.E. Jansen, and Vincenzo Pezone (2026).** "Do Teams Alleviate or Exacerbate Overreaction in Beliefs?" *Journal of Financial Economics* 176: 104219.
- Becker, Joshua, Devon Brackbill, and Damon Centola (2017).** "Network Dynamics of Social Influence in the Wisdom of Crowds." *Proceedings of the National Academy of Sciences* 114 (26): E5070–E5076.
- Bénabou, Roland (2013).** "Groupthink: Collective Delusions in Organizations and Markets." *Review of Economic Studies* 80 (2): 429–62.
- Blackwell, David (1953).** "Equivalent Comparisons of Experiments." *Annals of Mathematical Statistics* 24 (2): 265–72.
- Blinder, Alan S., and John Morgan (2005).** "Are Two Heads Better than One? Monetary Policy by Committee." *Journal of Money, Credit and Banking* 37 (5): 789–811.
- Calder-Wang, Sophie, and Paul A. Gompers (2021).** "And the Children Shall Lead: Gender Diversity and Performance in Venture Capital." *Journal of Financial Economics* 142 (1): 1–22.
- Campbell, John Y. (2006).** "Household Finance." *Journal of Finance* 61 (4): 1553–604.
- Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay (1997).** *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.
- Chan, David C. (2021).** "Influence and Information in Team Decisions: Evidence from Medical Residency." *American Economic Journal: Economic Policy* 13 (1): 106–37.
- Charness, Gary, and Matthias Sutter (2012).** "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives* 26 (3): 157–76.

Cheung, Stephen L., and Stefan Palan (2012). "Two Heads Are Less Bubbly Than One: Team Decision-Making in an Experimental Asset Market." *Experimental Economics* 15 (3): 373–97.

Cokely, Edward T., Mirta Galesic, Eric Schulz, Saima Ghazal, and Rocio Garcia-Retamero (2012). "Measuring Risk Literacy: The Berlin Numeracy Test." *Judgment and Decision Making* 7 (1): 25–47.

Conlon, John J., Malavika Mani, Gautam Rao, Matthew Ridley, and Frank Schilbach (2026). "Not Learning from Others." *Journal of Political Economy*, forthcoming.

Cookson, J. Anthony, William Mullins, and Marina Niessner (2024). "Social Media and Finance." In *Oxford Research Encyclopedia of Economics and Finance*. Anindya Banerjee, ed. Oxford University Press.

Cooper, David J., and John H. Kagel (2005). "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games." *American Economic Review* 95 (3): 477–509.

Costa-Gomes, Miguel A., and Georg Weizsäcker (2008). "Stated Beliefs and Play in Normal-Form Games." *Review of Economic Studies* 75 (3): 729–62.

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri (2013). "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications." *Journal of Economic Literature* 51 (1): 5–62.

Eling, Martin, and Frank Schuhmacher (2007). "Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds?" *Journal of Banking & Finance* 31 (9): 2632–47.

Enke, Benjamin, and Thomas Graeber (2023). "Cognitive Uncertainty." *Quarterly Journal of Economics* 138 (4): 2021–67.

Enke, Benjamin, Thomas Graeber, and Ryan Oprea (2023). "Confidence, Self-Selection, and Bias in the Aggregate." *American Economic Review* 113 (7): 1933–66.

Enke, Benjamin, and Florian Zimmermann (2019). "Correlation Neglect in Belief Formation." *Review of Economic Studies* 86 (1): 313–32.

Evans, Richard B., Melissa Porras Prado, A. Emanuele Rizzo, and Rafael Zambrana (2025). "Identity, Diversity, and Team Performance: Evidence from U.S. Mutual Funds." *Management Science* 71 (4): 3026–51.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus (2021). "Five Facts about Beliefs and Portfolios." *American Economic Review* 111 (5): 1481–522.

Goeree, Jacob K., and Leeat Yariv (2011). "An Experimental Study of Collective Deliberation." *Econometrica* 79 (3): 893–921.

Goetzmann, William N., and Alok Kumar (2008). "Equity Portfolio Diversification." *Review of Finance* 12 (3): 433–63.

Greenwood, Robin, and Andrei Shleifer (2014). "Expectations of Returns and Expected Returns." *Review of Financial Studies* 27 (3): 714–46.

Guiso, Luigi, and Luana Zaccaria (2023). "From Patriarchy to Partnership: Gender Equality and Household Finance." *Journal of Financial Economics* 147 (3): 573–95.

Hong, Lu, and Scott E. Page (2004). "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* 101 (46): 16385–89.

Hoogendoorn, Sander, Hessel Oosterbeek, and Mirjam van Praag (2013). "The Impact of Gender Diversity on the Performance of Business Teams: Evidence From a Field Experiment." *Management Science* 59 (7): 1514–28.

Imas, Alex (2016). "The Realization Effect: Risk-Taking after Realized versus Paper Losses." *American Economic Review* 106 (8): 2086–109.

Kang, Jun-Koo, Seil Kim, and Seungjoon Oh (2022). "Does Board Demographic Diversity Enhance Cognitive Diversity and Monitoring?" *Accounting Review* 97 (6): 385–415.

Ke, Da (2021). "Who Wears the Pants? Gender Identity Norms and Intrahousehold Financial Decision-Making." *Journal of Finance* 76 (3): 1389–425.

Kilduff, Martin, Reinhard Angelmar, and Ajay Mehra (2000). "Top Management-Team Diversity and Firm Performance: Examining the Role of Cognitions." *Organization Science* 11 (1): 21–34.

Levine, Sheen S., Evan P. Apfelbaum, Mark Bernard, Valerie L. Bartelt, Edward J. Zajac, and David Stark (2014). "Ethnic Diversity Deflates Price Bubbles." *Proceedings of the National Academy of Sciences* 111 (52): 18524–29.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing (2011). "How Social Influence Can Undermine the Wisdom of Crowd Effect." *Proceedings of the National Academy of Sciences* 108 (22): 9020–25.

Lyons, Elizabeth (2017). "Team Production in International Labor Markets: Experimental Evidence from the Field." *American Economic Journal: Applied Economics* 9 (3): 70–104.

Markowitz, Harry (1952). "Portfolio Selection." *Journal of Finance* 7 (1): 77–91.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa (2019). "Failures in Contingent Reasoning: The Role of Uncertainty." *American Economic Review* 109 (10): 3437–74.

Mellers, Barbara, Lyle Ungar, Jonathan Baron, Jaime Ramos, Burcu Gurcay, Katrina Fincher, Sydney E. Scott, Don Moore, Pavel Atanasov, Samuel A. Swift, Terry Murray, Eric Stone, and Philip E. Tetlock (2014). "Psychological Strategies for Winning a Geopolitical Forecasting Tournament." *Psychological Science* 25 (5): 1106–15.

Merton, Robert C. (1980). "On Estimating the Expected Return on the Market: An Exploratory Investigation." *Journal of Financial Economics* 8 (4): 323–61.

Nagel, Rosemarie (1995). "Unraveling in Guessing Games: An Experimental Study." *American Economic Review* 85 (5): 1313–26.

Norris, Paul, and Seymour Epstein (2011). "An Experiential Thinking Style: Its Facets and Relations with Objective and Subjective Criterion Measures." *Journal of Personality* 79 (5): 1043–80.

Odean, Terrance (1999). "Do Investors Trade Too Much?" *American Economic Review* 89 (5): 1279–98.

Palan, Stefan (2013). "A Review of Bubbles and Crashes in Experimental Asset Markets." *Journal of Economic Surveys* 27 (3): 570–88.

Radbruch, Jonas, and Amelie Schiprowski (2023). "Committee Deliberation and Gender Differences in Influence." ECONtribute Discussion Paper No. 234, University of Bonn and University of Cologne.

Rau, Holger A. (2015). "The Disposition Effect in Team Investment Decisions: Experimental Evidence." *Journal of Banking & Finance* 61: 272–82.

Surowiecki, James (2005). *The Wisdom of Crowds*. Anchor.

Tetlock, Philip E., and Dan Gardner (2016). *Superforecasting: The Art and Science of Prediction*. Random House.

A Appendix

A.1 Additional tables

Table A.1: Construct validity of the thinking-style scores

Dependent variable: Model:	CSI (1)	Numeracy (2)	Risk CE (3)
REI-R (analytic)	0.39*** (0.04)	0.03** (0.01)	1.29*** (0.42)
REI-E (intuitive)	-0.29*** (0.04)	-0.01 (0.01)	0.57 (0.50)
Observations (people)	388	388	388
R ²	0.40	0.02	0.03

Notes: This table reports individual-level OLS estimates of the relationship between thinking styles and related cognitive and preference measures. The dependent variable in column (1) is the Cognitive Style Index (CSI). Higher values on the CSI indicate more analytic thinking. In column (2), it is the Berlin Numeracy Test score (0–4). In column (3), it is the certainty equivalent from our risk preference elicitation. The independent variables are REI-R and REI-E in all specifications. HC1 standard errors are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.2: Balance of Round-1 performance and pre-treatment covariates

	Composition arm			Decision-mode arm				
	Homog.	Divers.	$p(\text{Div})$	Crowd	Panel	Team	$p(\text{Panel})$	$p(\text{Team})$
Panel (a). Individual level								
<i>Round-1 performance</i>								
R1 # assets	7.80	7.36	0.081*	7.53	7.66	7.55	0.652	0.946
R1 # sectors	4.82	4.66	0.402	4.69	4.83	4.72	0.523	0.897
R1 HHI	1835	1997	0.246	1893	1809	2042	0.507	0.409
R1 max weight	0.24	0.26	0.395	0.25	0.24	0.26	0.588	0.797
R1 APC (3m)	0.15	0.17	0.038**	0.16	0.16	0.16	0.714	0.989
R1 APC (2y)	0.25	0.26	0.142	0.26	0.25	0.25	0.713	0.679
R1 hist. ret. (3m, %)	51.27	52.11	0.917	61.05	47.76	46.42	0.212	0.136
R1 hist. ret. (2y, %)	21.83	22.92	0.518	22.81	21.95	22.35	0.703	0.826
R1 confidence	63.64	64.07	0.851	65.86	62.30	63.41	0.189	0.352
<i>Cognitive inputs</i>								
REI-R	35.42	35.53	0.893	35.91	35.09	35.43	0.400	0.633
REI-E	21.71	21.69	0.980	21.88	21.55	21.67	0.678	0.785
CSI	26.79	26.48	0.612	26.80	26.75	26.36	0.947	0.544
Numeracy	2.28	2.26	0.845	2.08	2.38	2.36	0.047**	0.072*
<i>Beliefs & preferences</i>								
EMH belief	2.22	2.15	0.542	2.09	2.30	2.17	0.126	0.562
Risk CE	89.47	88.85	0.903	85.66	89.28	92.44	0.573	0.309
S&P 500 forecast	7627	7703	0.260	7636	7717	7644	0.293	0.912
Reliability	4.40	4.49	0.257	4.50	4.42	4.42	0.420	0.447
<i>Demographics</i>								
Share female	0.28	0.40	< 0.01***	0.31	0.37	0.34	0.271	0.506
Age	39.22	38.78	0.717	38.88	40.48	37.67	0.290	0.392
<i>N</i> (individuals)	194	194		128	128	132		
Panel (b). Pair level: pooled Round-1 portfolios								
Pooled R1 # assets	14.55	13.54	0.056*	14.11	14.08	13.94	0.959	0.798
Pooled R1 # sectors	7.02	6.64	0.144	6.77	6.98	6.74	0.480	0.946
Pooled R1 HHI	1059	1144	0.334	1087	1070	1147	0.858	0.576
Pooled R1 max weight	0.19	0.19	0.647	0.19	0.18	0.19	0.549	0.881
Pooled R1 APC (3m)	0.13	0.15	0.041**	0.15	0.15	0.14	0.666	0.178
Pooled R1 APC (2y)	0.23	0.25	0.101	0.25	0.24	0.23	0.386	0.107
Pooled R1 hist. ret. (3m, %)	57.67	54.84	0.754	68.77	51.01	49.20	0.135	0.074*
Pooled R1 hist. ret. (2y, %)	26.97	28.43	0.652	30.38	27.43	25.37	0.499	0.178
<i>N</i> (pairs)	97	97		64	64	66		

Notes: This table reports balance in Round-1 portfolio outcomes and pre-treatment covariates across the composition and decision-mode treatments. Panel (a) reports individual-level means, while Panel (b) reports the pair-level outcomes for the pooled Round-1 portfolio. $p(\text{Div})$ is the p -value on the Diversity treatment indicator. The columns labeled $p(\text{Panel})$ and $p(\text{Team})$ report p -values on Panel and Team indicators from regressions with Crowd as the omitted decision mode. Standard errors for the individual-level regressions are clustered by pair. Asset and sector counts are proxies for diversification, while higher HHI and maximum weight are indicative of higher concentration. APC is the weighted average pairwise return correlation. Historical return is the annualized mean daily excess return of the buy-and-hold portfolio over the stated window, expressed as a percentage. The S&P 500 forecast is winsorized to [5,000, 15,000]. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.3: *Team* portfolios versus individuals working alone: ANCOVA estimates

Model	Dependent variable:	Estimate (SD units)	Std. error	Estimate (raw units)	Units	Obs.
Panel (a): Return variables						
(1)	Realized Sharpe ratio	-0.13	(0.16)	-0.09	ratio	130
(2)	Higher Sharpe ratio than R1	-0.15	(0.18)	-7.28	pct. pts	130
(3)	Realized excess return	-0.18	(0.15)	-2.98	annual pp	130
(4)	Realized log gross return	-0.14	(0.16)	-0.57	log pts	130
(5)	Higher excess return than R1	-0.11	(0.17)	-5.59	pct. pts	130
(6)	Historical excess return (2y)	+0.01	(0.16)	+0.06	annual pp	125
(7)	Historical log gross return (2y)	+0.02	(0.16)	+0.47	log pts	125
Panel (b): Risk variables						
(8)	Realized volatility	+0.23*	(0.14)	+2.19	annual pp	130
(9)	Lower volatility than R1	+0.51**	(0.24)	+18.73	pct. pts	130
(10)	HHI	+0.50***	(0.15)	+486.83	HHI pts	130
(11)	Historical asset risk	-0.22*	(0.13)	-1.56	annual pp	130
(12)	Weighted APC	+0.23	(0.16)	+1.61	corr. pts	124
(13)	Unique assets	+1.57***	(0.32)	+2.79	assets	130
(14)	Unique sectors	+0.66***	(0.18)	+0.87	sectors	130
(15)	Maximum weight	+0.46***	(0.16)	+5.04	pct. pts	130

Notes: Each row reports the coefficient on assignment to *Teams* from a separate regression comparing Round-2 *Team* portfolios with the mean outcome across the two unpooled portfolios of each *Crowd* pair. The specifications correspond to Figure 7 and equation (15). All models control for the corresponding continuous Round-1 individual benchmark. The improvement outcomes equal one when Round-2 Sharpe or return exceeds that benchmark, or volatility falls below it, and zero otherwise. These linear probability estimates are scaled by the sample standard deviation of the corresponding indicator among pooled *Crowd* portfolios. Other estimates are scaled by the standard deviation of the Round-1 individual benchmark across pairs with valid observations. The raw estimates for improvement indicators are in percentage points. Signs for volatility, HHI, historical asset risk, APC and maximum weight are reversed so positive values indicate improvement. Standard errors in SD units use HC1. Outcome definitions and evaluation windows follow Section 4.2. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.4: Composition, communication, and aggregation: stepwise regression estimates

Comparison: Dependent variable:	<i>Team</i> – individual (1) (combined)	Composition (2) (pooling)	Communication (3) (<i>Panel</i> – <i>Crowd</i>)	Aggregation (4) (<i>Team</i> – <i>Panel</i>)	<i>Team</i> – <i>Crowd</i> (5) (combined)
Realized volatility	+0.23* (0.14) [+2.19]	+0.46*** (0.04) [+4.31]	-0.10 (0.16) [-0.90]	-0.13 (0.16) [-1.20]	-0.22* (0.13) [-2.05]
Lower volatility than R1	+0.51** (0.24) [+18.73]	+1.07*** (0.17) [+39.06]	-0.43** (0.20) [-15.57]	-0.16 (0.23) [-5.81]	-0.57*** (0.21) [-20.90]
HHI	+0.50*** (0.15) [+486.83]	+0.89*** (0.06) [+865.13]	-0.43*** (0.11) [-413.11]	+0.04 (0.15) [+42.60]	-0.40*** (0.13) [-385.02]
Historical asset risk	-0.22* (0.13) [-1.56]	-0.15*** (0.06) [-1.06]	+0.08 (0.20) [+0.58]	-0.17 (0.16) [-1.19]	-0.08 (0.15) [-0.55]
Weighted APC	+0.23 (0.16) [+1.61]	+0.30*** (0.08) [+2.16]	+0.01 (0.14) [+0.05]	-0.11 (0.17) [-0.79]	-0.11 (0.16) [-0.76]
Unique assets	+1.57*** (0.32) [+2.79]	+3.59*** (0.12) [+6.40]	-1.64*** (0.31) [-2.92]	-0.40 (0.41) [-0.71]	-2.03*** (0.35) [-3.61]
Unique sectors	+0.66*** (0.18) [+0.87]	+1.66*** (0.09) [+2.21]	-0.83*** (0.20) [-1.10]	-0.18 (0.23) [-0.24]	-1.01*** (0.21) [-1.34]
Maximum weight	+0.46*** (0.16) [+5.04]	+0.54*** (0.06) [+5.95]	-0.24* (0.13) [-2.70]	+0.17 (0.15) [+1.88]	-0.09 (0.15) [-0.95]
Observations (except APC)	130	128	128	130	130
Observations (APC)	124	112	117	120	117
Clusters (pairs, except APC)		64			
Clusters (pairs, APC)		56			

Notes: Each outcome is estimated separately for each contrast. Column (1) compares Round-2 *Team* outcomes with the average outcome across each *Crowd* pair’s two unpooled Round-2 portfolios. Column (2) compares pooled and unpooled Round-2 *Crowd* outcomes and includes only the pooling indicator, with standard errors clustered by pair. Columns (3)–(5) compare *Panels* with *Crowds*, *Teams* with *Panels* and *Teams* with *Crowds*. Columns (1) and (3)–(5) control for the corresponding Round-1 individual benchmark and use HC1 standard errors. The indicator variable “Lower volatility than R1” is set equal to one if the pair’s Round-2 volatility is below their Round-1 individual benchmark, and zero otherwise. Improvement coefficients are divided by the indicator’s sample standard deviation among pooled *Crowd* portfolios. Coefficients with continuous outcomes are divided by the standard deviation of the corresponding Round-1 benchmark across pairs with valid observations. Estimates and standard errors in parentheses are reported in SD units. Brackets give the coefficients in natural units. These are percentage points for the improvement indicator and maximum weight, annualized percentage points for realized volatility and historical asset risk, HHI points for concentration, correlation points for APC and counts for assets and sectors. Signs for risk outcomes are reversed so positive estimates indicate improvement. Section 4.2 defines the outcomes and evaluation windows. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.5: Robustness of the main comparisons to the evaluation window and winsorization

Comparison: Dependent variable:	<i>Team</i> — individual (1)	Composition (2)	Communication (3)	Aggregation (4)	<i>Team</i> — <i>Crowd</i> (5)
Panel (a) Extended evaluation window					
Realized volatility	+0.21* (0.13)	+0.42*** (0.03)	-0.08 (0.14)	-0.13 (0.14)	-0.20* (0.12)
Lower volatility than R1	+0.60** (0.24)	+1.15*** (0.17)	-0.47** (0.20)	-0.12 (0.23)	-0.57*** (0.20)
HHI	+0.50*** (0.15)	+0.89*** (0.06)	-0.43*** (0.11)	+0.04 (0.15)	-0.40*** (0.13)
Historical asset risk	-0.22* (0.13)	-0.15*** (0.06)	+0.08 (0.20)	-0.17 (0.16)	-0.08 (0.15)
Weighted APC	+0.23 (0.16)	+0.30*** (0.08)	+0.01 (0.14)	-0.11 (0.17)	-0.11 (0.16)
Unique assets	+1.57*** (0.32)	+3.59*** (0.12)	-1.64*** (0.31)	-0.40 (0.41)	-2.03*** (0.35)
Unique sectors	+0.66*** (0.18)	+1.66*** (0.09)	-0.83*** (0.20)	-0.18 (0.23)	-1.01*** (0.21)
Maximum weight	+0.46*** (0.16)	+0.54*** (0.06)	-0.24* (0.13)	+0.17 (0.15)	-0.09 (0.15)
Panel (b) Original window, 1st/99th percentile winsorization					
Realized volatility	+0.23* (0.14)	+0.46*** (0.04)	-0.01 (0.12)	-0.19 (0.14)	-0.22* (0.13)
Lower volatility than R1	+0.51** (0.24)	+1.07*** (0.17)	-0.43** (0.20)	-0.16 (0.23)	-0.57*** (0.21)
HHI	+0.55*** (0.12)	+0.88*** (0.05)	-0.43*** (0.11)	+0.09 (0.13)	-0.34*** (0.10)
Historical asset risk	-0.22* (0.13)	-0.14*** (0.05)	+0.16 (0.17)	-0.22 (0.15)	-0.08 (0.14)
Weighted APC	+0.23 (0.15)	+0.31*** (0.08)	+0.01 (0.14)	-0.11 (0.17)	-0.10 (0.16)
Unique assets	+1.59*** (0.32)	+3.59*** (0.12)	-1.63*** (0.31)	-0.39 (0.41)	-2.01*** (0.35)
Unique sectors	+0.65*** (0.18)	+1.66*** (0.09)	-0.83*** (0.20)	-0.18 (0.23)	-1.02*** (0.20)
Maximum weight	+0.51*** (0.14)	+0.54*** (0.06)	-0.24* (0.13)	+0.22 (0.14)	-0.03 (0.13)
Panel (c) Original window, 5th/95th percentile winsorization					
Realized volatility	+0.24** (0.12)	+0.40*** (0.03)	+0.04 (0.11)	-0.17 (0.12)	-0.15 (0.12)
Lower volatility than R1	+0.51** (0.24)	+1.07*** (0.17)	-0.43** (0.20)	-0.16 (0.23)	-0.57*** (0.21)
HHI	+0.53*** (0.09)	+0.79*** (0.04)	-0.32*** (0.08)	+0.04 (0.09)	-0.28*** (0.08)
Historical asset risk	-0.18 (0.12)	-0.11** (0.05)	+0.18 (0.15)	-0.22* (0.13)	-0.06 (0.14)
Weighted APC	+0.21 (0.15)	+0.29*** (0.08)	+0.02 (0.14)	-0.13 (0.16)	-0.11 (0.16)
Unique assets	+1.54*** (0.30)	+3.54*** (0.13)	-1.58*** (0.30)	-0.45 (0.38)	-2.01*** (0.33)
Unique sectors	+0.60*** (0.17)	+1.61*** (0.09)	-0.77*** (0.18)	-0.24 (0.21)	-1.01*** (0.19)
Maximum weight	+0.47*** (0.11)	+0.49*** (0.05)	-0.17 (0.10)	+0.14 (0.11)	-0.02 (0.11)
Observations (except APC)	130	128	128	130	130
Observations (APC)	124	112	117	120	117
Clusters (pairs, except APC)		64			
Clusters (pairs, APC)		56			

Notes: Specifications and column comparisons follow Table A.4. Panel (a) extends the evaluation window from June 2–September 2 to June 2–October 23, 2026, increasing the number of daily returns from 64 to 100. Both initial and final portfolios are evaluated over the extended window. Panels (b) and (c) retain the original window and winsorize continuous dependent variables at the indicated percentiles. Cutoffs are common across decision modes and are calculated from the available final group outcomes and unpooled *Crowd* pair means. Baseline controls are not winsorized. The binary improvement outcome is unchanged by winsorization. Historical asset risk and diversification measures are unchanged by extending the evaluation window. Estimates (standard errors in parentheses) are standardized using the original uncapped reference standard deviations from Table A.4 and held fixed across panels. Positive estimates indicate improvement. Standard errors are clustered by pair in column (2) and use HC1 in the other columns. Observation and cluster counts apply to every panel. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.6: Round-1 portfolio construction and editing time

Measure	Observed window	N	Mean	Median	P25	P75
Total construction time	Screen entry to last edit	386	8.18	8.60	5.81	10.66
Active editing time	First edit to last edit	386	6.47	6.78	4.13	8.93
Time before first edit	Screen entry to first edit	386	1.71	1.36	0.85	2.19
Recorded edit events	Ticker and amount changes	386	22.69	22.00	18.00	26.00

Notes: This table reports portfolio edits and the time subjects spend designing their portfolios in Round 1. Total construction time is the interval from opening the Round-1 portfolio allocation screen to the last recorded edit. Active editing time is the interval between the first and last recorded edit. The table also reports the time before the first edit. All times are in minutes. We count an edit as any change made to a ticker or a dollar amount. We have recorded edits from 386 participants. Due to a technical failure, edit histories were not recorded for two participants.

Table A.7: Portfolio composition and realized performance in Round 1

Panel (a). Realized excess return				
Dependent variable:	Realized excess return			
Model:	(1)	(2)	(3)	(4)
Historical excess return (3m)	-9.22*** (1.54)			-2.43 (2.21)
Historical excess return (2y)		-9.79*** (1.05)		-6.04*** (1.69)
Information technology share			-8.58*** (1.21)	-3.30** (1.50)
Constant	16.12*** (1.09)	16.12*** (1.08)	16.12*** (1.11)	16.12*** (1.07)
Observations	373	373	373	373
R ²	0.16	0.18	0.14	0.21
Panel (b). Realized volatility				
Dependent variable:	Realized volatility			
Model:	(1)	(2)	(3)	(4)
HHI	3.92*** (0.61)			2.37*** (0.43)
Weighted APC		4.80*** (0.85)		2.41*** (0.53)
Historical asset risk			6.80*** (0.60)	5.55*** (0.49)
Constant	20.84*** (0.48)	20.84*** (0.45)	20.84*** (0.38)	20.84*** (0.33)
Observations	369	369	369	369
R ²	0.15	0.23	0.47	0.59

Notes: This table reports OLS regressions using individual Round-1 portfolios. The outcome in Panel (a) is annualized realized excess return in percentage points. Panel (b) uses annualized realized volatility in percentage points. Columns (1)–(3) include each characteristic separately and column (4) includes all three. Historical excess returns use the preceding two-year and three-month windows and enter in levels. Information technology share is the fraction of the portfolio allocated to stocks in that sector. Other measures follow Section 4.2. Each panel uses the same complete observations across columns, with 373 individuals in Panel (a) and 369 in Panel (b). Predictors are centered and divided by their standard deviation within the corresponding panel's sample. HC1 standard errors are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.8: Stock return and volatility as joint predictors of being chosen

Dependent variable:	Positive allocation to stock in Round 1			
Return and volatility window: Model:	3 months (1)	6 months (2)	1 year (3)	2 years (4)
Historical excess return	0.32*** (0.05)	0.08 (0.05)	0.14** (0.06)	-0.17*** (0.06)
Historical volatility	-0.71*** (0.05)	-0.60*** (0.05)	-0.63*** (0.05)	-0.34*** (0.05)
Log market capitalization	-0.09 (0.10)	0.09 (0.10)	-0.08 (0.12)	0.51*** (0.12)
Log median daily dollar volume	2.76*** (0.13)	2.64*** (0.14)	2.77*** (0.14)	2.28*** (0.13)
S&P common stock quality rank	0.06** (0.03)	0.01 (0.03)	-0.00 (0.03)	-0.01 (0.03)
Observations	180,213	180,213	180,213	180,213
Clusters (participants)	381	381	381	381
R ²	0.05	0.05	0.05	0.05

Notes: We examine whether subjects allocate positive weight to each stock available in the investment menu. The dependent variable is an indicator for selection that is equal to one when the stock is selected by a subject in Round 1, and zero otherwise. The independent variable differs by row and is measured over the historical window indicated in each column title. Coefficients are estimated using a linear probability model and reported in percentage points, with historical return and volatility measured in standard deviation units. The controls are log market capitalization, log median daily dollar volume, and the S&P common stock quality rank. Standard errors are clustered by participant. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.9: Participant and LLM portfolio breadth

Dependent variable:	Number of assets	Six to nine assets	All ten slots
Model:	(1)	(2)	(3)
LLM benchmark mean	7.62*** (0.16)	0.72*** (0.04)	0.22*** (0.04)
Participant portfolio	-0.04 (0.21)	-0.45*** (0.05)	0.21*** (0.05)
Observations	488	488	488
R ²	0.00	0.14	0.03

Notes: We compare 388 participant portfolios from Round 1 with 100 LLM portfolios that we generated using the same instructions. The dependent variable in column (1) is the number of unique assets held in a portfolio. The outcomes in columns (2) and (3) are indicators for the portfolio holding between six and nine assets and exactly ten assets, respectively. Each coefficient in the “LLM benchmark mean” row captures the average value of that outcome among LLM portfolios. Each coefficient in the Participant portfolio row captures the average difference in that outcome between human and LLM portfolios. HC1 standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.10: Communication intensity and portfolio breadth loss

Dependent variable:	Log chat words			Assets lost			Sectors lost				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Model:											
Team relative to Panel	0.34** (0.16)	0.87 (0.78)	0.40 (0.82)	0.31 (0.81)	0.37 (0.78)	0.49 (0.83)	0.25 (0.33)	0.07 (0.35)	0.05 (0.36)	0.07 (0.34)	0.11 (0.35)
Log chat words			1.09*** (0.41)					0.41** (0.16)			
Log messages				1.71*** (0.53)					0.60*** (0.21)		
Speaker switches					0.21*** (0.06)					0.07*** (0.02)	
Centered log chat words						1.38*** (0.50)					0.53** (0.21)
Team relative to Panel $\times$ Centered log chat words						-0.88 (0.89)					-0.35 (0.36)
Constant	4.38*** (0.15)	-0.27 (1.61)	-4.53** (2.09)	-3.86** (1.83)	-1.60 (1.55)	0.37 (1.58)	-0.99 (0.74)	-2.52*** (0.89)	-2.16*** (0.80)	-1.34* (0.71)	-0.69 (0.75)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	126	120	120	120	120	120	120	120	120	120	120
R ²	0.04	0.05	0.09	0.11	0.13	0.10	0.09	0.12	0.13	0.14	0.12

Notes: The unit of observation is one of the 126 *Panel* or *Team* pairs with a nonempty Round-2 transcript. Assets lost and sectors lost are calculated relative to the union of members' Round-1 portfolios, with positive values indicating fewer holdings in the Round-2 group portfolio. Column (1) compares *Teams* with *Panels* using log chat words as the outcome. The outcomes in columns (2)–(6) and (7)–(11) are assets lost and sectors lost, respectively. All regressions control for thinking-style treatment. For the loss outcomes, we also control for the corresponding number of assets or sectors in the Round-1 union. Log words and log messages are calculated as $\log(1 + \text{count})$. Columns (3) and (8) include log words, (4) and (9) include log messages, and (5) and (10) include speaker switches. Columns (6) and (11) interact assignment to *Teams* with log words centered at the sample mean. Assignment to *Teams* is randomized. Parentheses contain HC1 standard errors. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.11: Historical returns of retained, adopted, and dropped stocks, by decision mode

	Crowd (no communication)				Panel subjects				Teams			
	2-year		3-month		2-year		3-month		2-year		3-month	
Dependent variable:	Historical annual excess return (pp)											
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Retained (own / pooled R1)	12.62*** (2.89)	6.36** (3.11)	34.70*** (9.20)	25.49** (11.15)	7.12*** (2.74)	5.65* (2.92)	8.21 (8.19)	7.23 (8.94)	8.07*** (2.70)	5.69** (2.83)	28.14*** (8.72)	24.82*** (9.26)
Adopted (teammate)					8.95** (3.94)	6.97 (4.56)	14.40 (11.25)	6.68 (13.38)				
Adopted (new)	0.81 (2.62)	0.14 (2.89)	-2.97 (8.06)	-5.73 (8.57)	2.27 (3.58)	3.01 (3.78)	-8.17 (10.77)	-3.08 (11.90)	6.12 (5.54)	3.59 (4.93)	7.88 (10.77)	-1.51 (11.96)
Retained × Communication					-5.51 (3.97)	-2.53 (3.98)	-26.49** (12.27)	-26.16* (15.18)	-4.65 (4.07)	-3.56 (4.56)	-6.86 (12.83)	-9.96 (16.55)
Constant	15.95*** (1.77)		20.35*** (5.36)		17.23*** (2.00)		26.58*** (5.94)		17.33*** (1.76)		18.95*** (5.53)	
Subject / team FE		✓		✓		✓		✓		✓		✓
R ²	0.03	0.21	0.03	0.28	0.01	0.19	0.01	0.18	0.01	0.12	0.02	0.14
Observations (stocks)	1,345	1,345	1,345	1,345	1,400	1,400	1,400	1,400	1,102	1,102	1,102	1,102
Clusters (pairs)	64	64	64	64	64	64	64	64	66	66	66	66

Notes: This table reports stock-level OLS estimates of the relationship between a stock's portfolio status and its historical return. Each observation is a stock, and the dependent variable is its annualized mean daily excess return in percentage points over the stated two-year or three-month window. Columns (1)–(4) report *Crowd* subjects, columns (5)–(8) report *Panel* subjects, and columns (9)–(12) report *Teams*. For *Crowd* and *Panel* subjects, we distinguish stocks they retain from Round 1 from those they add in Round 2. “Adopted (teammate)” stocks are those that were held by the subject's partner in Round 1, while “Adopted (new)” stocks are those that are new to both members. For each *Team* pair, we compare their joint Round-2 portfolio with the union of their members' Round-1 holdings. Stocks from this union are counted as retained if they remain in the Round 2 portfolio, and are counted as dropped otherwise. The “Adopted (teammate)” category does not apply to *Teams*, where we cannot observe the decisions of individuals to adopt their teammates' holdings. Within each sample and return window, the first specification is unconditional. The second includes subject fixed effects for *Crowd* and *Panel* subjects and Team fixed effects for *Teams*. The *Retained* $\times$ *Communication* coefficient is the interaction between *Retained* and an indicator for the communicating arm in a sample restricted to retained and dropped stocks. It tests whether communication changes the historical-return difference between retained and dropped stocks. *Panel* subjects are matched to the subject-level *Crowd* sample used in columns (1)–(4). *Teams* are matched to a pair-level *Crowd* counterpart, constructed exactly like the *Team* sample with the pooled Round-1 union as the reference and the pool of the two members' separate Round-2 portfolios as the Round-2 portfolio (not shown separately). Standard errors clustered by pair are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.12: Historical returns of stocks mentioned in chat

Dependent variable:	Historical annual excess return (pp)							
	<i>Panel subjects</i>				<i>Teams</i>			
	2-year		3-month		2-year		3-month	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mentioned in chat	10.28*** (3.37)	11.12*** (4.01)	27.29** (10.64)	25.71** (12.16)	3.02 (2.50)	3.17 (2.98)	18.86** (8.79)	23.50** (9.61)
Team FE		✓		✓		✓		✓
R^2	0.01	0.11	0.01	0.11	0.00	0.12	0.01	0.15
Observations (stocks)	991	991	991	991	1,017	1,017	1,017	1,017
Clusters (pairs)	64	64	64	64	66	66	66	66
Omitted-category mean	18.73	18.73	22.96	22.96	20.38	20.38	28.27	28.27

Notes: This table reports stock-level OLS estimates of the relationship between whether a stock is mentioned in Round-2 chat and its historical return. Each observation is a stock, and the dependent variable is its annualized mean daily excess return in percentage points over the stated two-year or three-month window. We include all stocks held by either partner in Round 1, as well as any additional stock mentioned in the Round-2 chat. The omitted category is the set of Round-1 holdings never mentioned in the chat. Chat mentions are counted using ticker symbols and common company names (e.g., both “Nvidia” and “NVDA”). Columns (1)–(4) report *Panel* pairs, while columns (5)–(8) report *Team* pairs. Within each decision mode and return window, the first specification is unconditional and the second includes pair fixed effects. Standard errors clustered by pair are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.13: Analytic thinking style, confidence, weight influence, and ETF reallocation

Dependent variable:	Confidence (R1)			ETF reallocation (R1)			Weight influence ($\times 100$)		
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Variables</i>									
REI-R (analytic)	0.61*** (0.18)	0.58*** (0.18)	0.57*** (0.18)	-0.63*** (0.24)	-0.56** (0.24)	-0.55** (0.25)	0.38* (0.20)	0.41** (0.21)	0.43** (0.21)
REI-E (intuitive)	0.55** (0.23)	0.55** (0.23)	0.52** (0.23)	-0.35 (0.32)	-0.37 (0.32)	-0.36 (0.33)	-0.45** (0.22)	-0.54** (0.22)	-0.54** (0.22)
<i>Controls</i>									
Demographics		✓	✓		✓	✓		✓	✓
Beliefs & prefs.			✓			✓			✓
Observations (people)	388	386	386	387	385	385	388	386	386
Clusters (pairs)	194	194	194	194	194	194	194	194	194
R ²	0.05	0.07	0.07	0.02	0.03	0.03	0.02	0.05	0.05

Notes: This table reports individual-level OLS estimates of the relationship between thinking styles and confidence, ETF reallocation, and portfolio influence. *REI-R* and *REI-E* are the analytic and intuitive thinking-style scores, entered jointly. *Confidence* is subjects' Round-1 confidence rating on a 0–100 scale. *ETF reallocation* is the percentage of the Round-1 portfolio that subjects choose to move into an S&P 500 ETF. *Weight influence* is 100 times the cosine similarity between the realized Round-2 group portfolio and the subject's Round-1 portfolio weights. For each outcome, the first specification includes only the two thinking-style scores, the second adds gender, age and numeracy, and the third further adds EMH belief and the risk certainty equivalent. Standard errors clustered by pair are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.14: Belief diversity and portfolio diversification, by decision mode

Dependent variable: Model:	Realized volatility*	Lower volatility than R1	HHI*	Historical asset risk*	Weighted APC*	Unique assets	Unique sectors	Maximum weight*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Belief diversity within each decision mode</i>								
Crowds	0.08 (0.15)	-0.06 (0.26)	-0.06 (0.11)	-0.03 (0.26)	0.28 (0.18)	-0.06 (0.33)	0.31 (0.24)	-0.01 (0.17)
Panels	0.19 (0.21)	0.40 (0.32)	-0.20 (0.18)	0.47* (0.25)	0.09 (0.22)	-0.81 (0.52)	-0.37 (0.31)	-0.12 (0.19)
Teams	0.56** (0.24)	0.48 (0.35)	0.79*** (0.28)	0.17 (0.20)	0.41 (0.29)	2.63*** (0.56)	1.24*** (0.32)	0.82*** (0.25)
<i>Differences between belief-diversity coefficients</i>								
Panel minus Crowd	0.12 (0.26)	0.46 (0.41)	-0.14 (0.22)	0.50 (0.37)	-0.19 (0.28)	-0.75 (0.62)	-0.68* (0.40)	-0.11 (0.25)
Team minus Panel	0.37 (0.31)	0.08 (0.48)	1.00*** (0.34)	-0.30 (0.32)	0.32 (0.37)	3.44*** (0.77)	1.61*** (0.47)	0.94*** (0.32)
Team minus Crowd	0.49* (0.29)	0.54 (0.43)	0.86*** (0.31)	0.20 (0.33)	0.13 (0.34)	2.69*** (0.64)	0.93** (0.40)	0.83*** (0.30)
Equal slopes, p-value	0.25	0.35	0.01	0.39	0.65	< 0.01	< 0.01	< 0.01
<i>Controls</i>								
Decision-mode indicators	✓	✓	✓	✓	✓	✓	✓	✓
Round-1 individual benchmark	✓	✓	✓	✓	✓	✓	✓	✓
Initial portfolio diversity	✓	✓	✓	✓	✓	✓	✓	✓
Thinking-style diversity	✓	✓	✓	✓	✓	✓	✓	✓
Observations	187	187	187	187	170	187	187	187
R ²	0.54	0.07	0.40	0.53	0.38	0.51	0.50	0.38
Crowd pairs	63	63	63	63	56	63	63	63
Panel pairs	62	62	62	62	58	62	62	62
Team pairs	62	62	62	62	56	62	62	62

Notes: Each column estimates one pooled ANCOVA regression of the Round-2 group outcome on belief diversity interacted with decision mode, controlling for decision-mode indicators, the corresponding Round-1 individual benchmark, initial portfolio diversity and thinking-style treatment. Group outcomes refer to pooled portfolios for *Crowds* and *Panels* and joint portfolios for *Teams*. Belief diversity equals one when one member's forecast of the change in the S&P 500 exceeds the sample median and the other's is at or below it. Initial portfolio diversity is a one-standard-deviation decrease in the cosine similarity of partners' Round-1 allocations. The first block reports the implied belief-diversity coefficient within each mode. The second reports differences between these coefficients and the p-value from a joint test of equality across all three modes. Lower volatility than R1 is an indicator for beating the pair's Round-1 individual benchmark, with benchmark volatility in levels as its control. Its coefficients are scaled by the indicator's standard deviation among pooled *Crowd* portfolios. Other coefficients use the standard deviation of the corresponding Round-1 individual benchmark across all pairs with a valid benchmark. Volatility and historical asset risk enter in annualized percentage points. Historical asset risk and average pairwise asset correlation (APC) use the two-year historical window. Outcomes marked with * have their signs reversed so positive coefficients indicate improvement. Parentheses contain HC1 standard errors. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.15: Pair composition and the pooling and revision components, communicating pairs

Dependent variable:	Realized volatility*	HHI*	Historical asset risk*	Weighted APC*	Unique assets	Unique sectors	Maximum weight*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a). Initial portfolio diversity							
Initial pooling benefit	0.09*** (0.01)	0.19*** (0.02)	-0.05* (0.03)	0.17*** (0.05)	0.52*** (0.05)	0.15*** (0.05)	0.24*** (0.02)
Revision component	-0.14** (0.05)	-0.23*** (0.08)	0.04 (0.06)	-0.14* (0.08)	-0.46** (0.20)	-0.18 (0.11)	-0.24*** (0.07)
Panel (b). Belief diversity							
Initial pooling benefit	-0.01 (0.04)	0.00 (0.02)	-0.11 (0.07)	-0.02 (0.11)	0.05 (0.07)	-0.08 (0.10)	0.03 (0.05)
Revision component	0.39** (0.17)	0.29* (0.17)	0.44*** (0.15)	0.26 (0.19)	0.83** (0.40)	0.48** (0.24)	0.33** (0.16)
Panel (c). Thinking-style diversity							
Initial pooling benefit	0.01 (0.04)	0.00 (0.02)	0.05 (0.07)	-0.13 (0.11)	-0.04 (0.07)	-0.14 (0.10)	0.02 (0.05)
Revision component	0.05 (0.16)	0.13 (0.16)	-0.06 (0.15)	0.37* (0.20)	-0.03 (0.41)	0.10 (0.24)	0.10 (0.15)
Decision mode indicator	✓	✓	✓	✓	✓	✓	✓
Round-1 individual benchmark	✓	✓	✓	✓	✓	✓	✓
Observations (pairs)	124	124	124	123	124	124	124

Notes: This table reports the specifications in Table 2 restricted to *Panel* and *Team* pairs. Panels (a), (b) and (c) report coefficients on initial portfolio diversity, belief diversity and thinking-style diversity, respectively. Each outcome and component has a separate regression jointly including all three composition measures and the corresponding Round-1 individual benchmark. The regressions also include a *Team* indicator. Initial portfolio diversity is a one-standard-deviation decrease in cosine similarity. Belief diversity indicates whether partners' forecasted three-month changes in the S&P 500 fall on opposite sides of the sample median. Thinking-style diversity is the randomized *Diversity* treatment indicator. The initial pooling benefit is the pooled Round-1 outcome minus the Round-1 individual benchmark. The revision component is the Round-2 outcome minus the pooled Round-1 outcome. Coefficients use the same full-sample reference standard deviations as Table 2. Outcomes marked with * are sign-reversed so positive coefficients indicate improvement. Observation counts apply to both components. HC1 standard errors are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

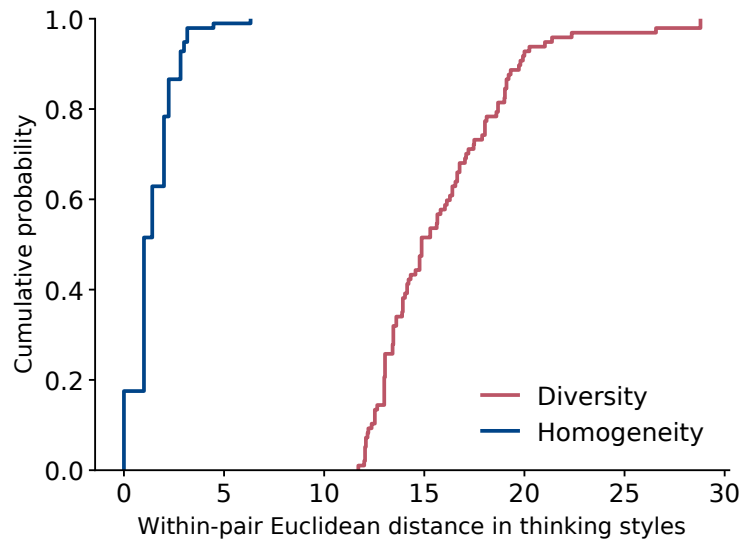
Table A.16: Pair composition and the pooling and revision components, non-communicating pairs

Dependent variable:	Realized volatility*	HHI*	Historical asset risk*	Weighted APC*	Unique assets	Unique sectors	Maximum weight*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a). Initial portfolio diversity							
Initial pooling benefit	0.15*** (0.03)	0.16*** (0.01)	-0.04 (0.06)	0.25*** (0.09)	0.50*** (0.05)	0.10 (0.08)	0.28*** (0.04)
Revision component	-0.07 (0.07)	-0.07 (0.05)	0.06 (0.10)	0.16* (0.09)	-0.18 (0.19)	0.17 (0.14)	-0.13 (0.08)
Panel (b). Belief diversity							
Initial pooling benefit	0.07 (0.04)	0.01 (0.01)	-0.06 (0.10)	0.40*** (0.12)	-0.01 (0.07)	0.16 (0.14)	0.10* (0.06)
Revision component	-0.01 (0.15)	-0.04 (0.09)	-0.03 (0.21)	-0.21 (0.16)	-0.07 (0.34)	0.15 (0.22)	-0.09 (0.14)
Panel (c). Thinking-style diversity							
Initial pooling benefit	0.08 (0.05)	-0.01 (0.01)	-0.01 (0.09)	-0.06 (0.13)	-0.06 (0.07)	-0.17 (0.14)	0.04 (0.06)
Revision component	0.10 (0.13)	-0.10 (0.12)	-0.11 (0.21)	0.08 (0.16)	0.05 (0.38)	0.54** (0.24)	-0.24 (0.18)
Round-1 individual benchmark	✓	✓	✓	✓	✓	✓	✓
Observations (pairs)	63	63	63	63	63	63	63

Notes: This table reports the specifications in Table 2 restricted to *Crowd* pairs. Panels (a), (b) and (c) report coefficients on initial portfolio diversity, belief diversity and thinking-style diversity, respectively. Each outcome and component has a separate regression jointly including all three composition measures and the corresponding Round-1 individual benchmark. Initial portfolio diversity is a one-standard-deviation decrease in cosine similarity. Belief diversity indicates whether partners' forecasted three-month changes in the S&P 500 fall on opposite sides of the sample median. Thinking-style diversity is the randomized *Diversity* treatment indicator. The initial pooling benefit is the pooled Round-1 outcome minus the Round-1 individual benchmark. The revision component is the Round-2 outcome minus the pooled Round-1 outcome. Coefficients use the same full-sample reference standard deviations as Table 2. Outcomes marked with * are sign-reversed so positive coefficients indicate improvement. Observation counts apply to both components. HC1 standard errors are in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

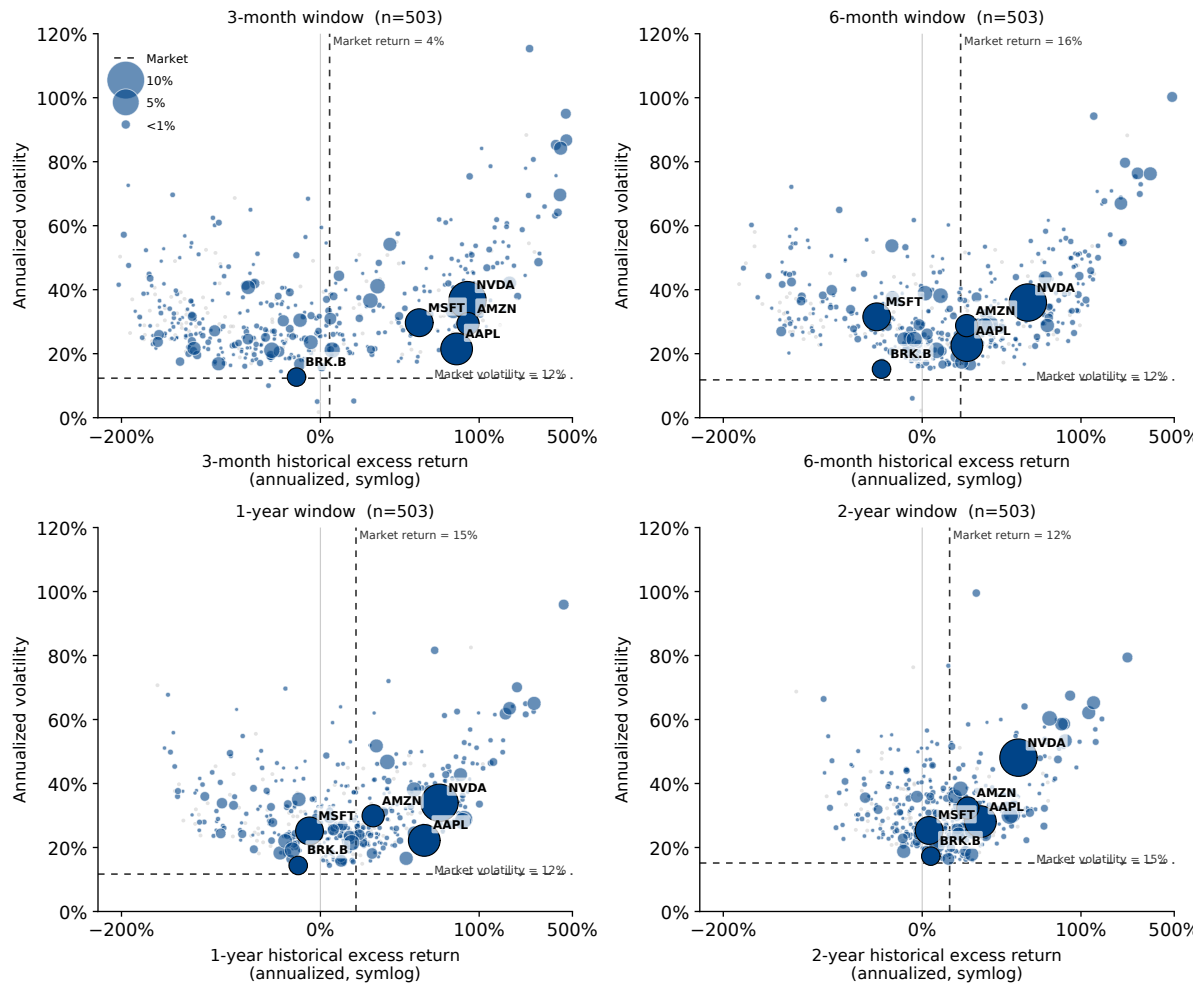
A.2 Additional figures

Figure A.1: Within-pair thinking-style distance by cognitive-composition condition



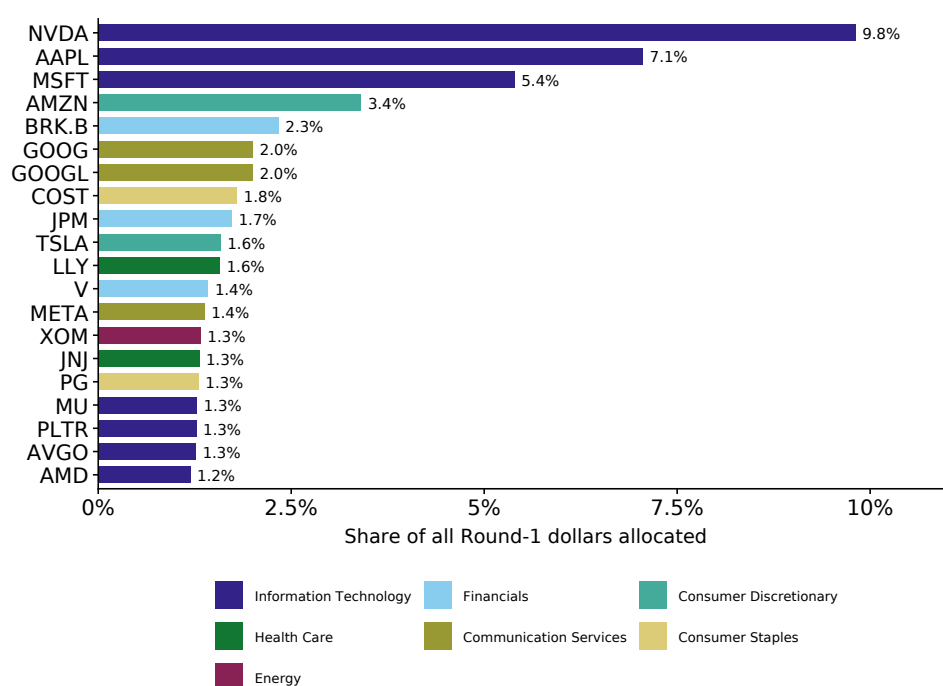
Notes: Cumulative distribution of the within-pair Euclidean distance in thinking styles (the rationality and experientiality scales of the Rational-Experiential Inventory, REI-R and REI-E) by cognitive-composition condition. The assignment algorithm separates the *Diversity* and *Homogeneity* arms across the full distribution.

Figure A.2: Risk, return, and dollars allocated across the S&P 500, by evaluation window



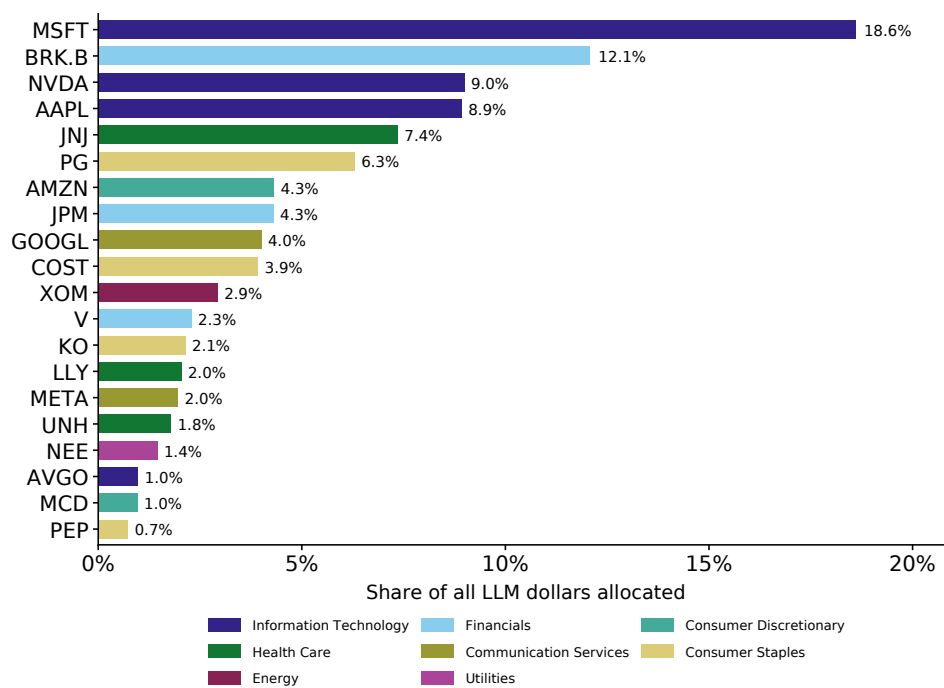
Notes: Each point is an S&P 500 ticker, positioned by its annualized mean daily excess return (x-axis, symlog) and annualized volatility (y-axis) over the window ending on the experiment date; marker area is proportional to the ticker's share of all dollar holdings across subjects' Round-1 portfolios. Dashed lines report the mean annualized excess return across tickers and the annualized volatility of an equally weighted portfolio that holds all tickers. The labels identify the five stocks that receive the most weight in Round 1; these are NVDA, AAPL, MSFT, AMZN, and BRK.B. The 3-month panel appears in the main text (Figure 5, Panel (b)); the 2-year panel is the bottom-right panel here.

Figure A.3: Most-selected tickers by Round-1 dollar allocation



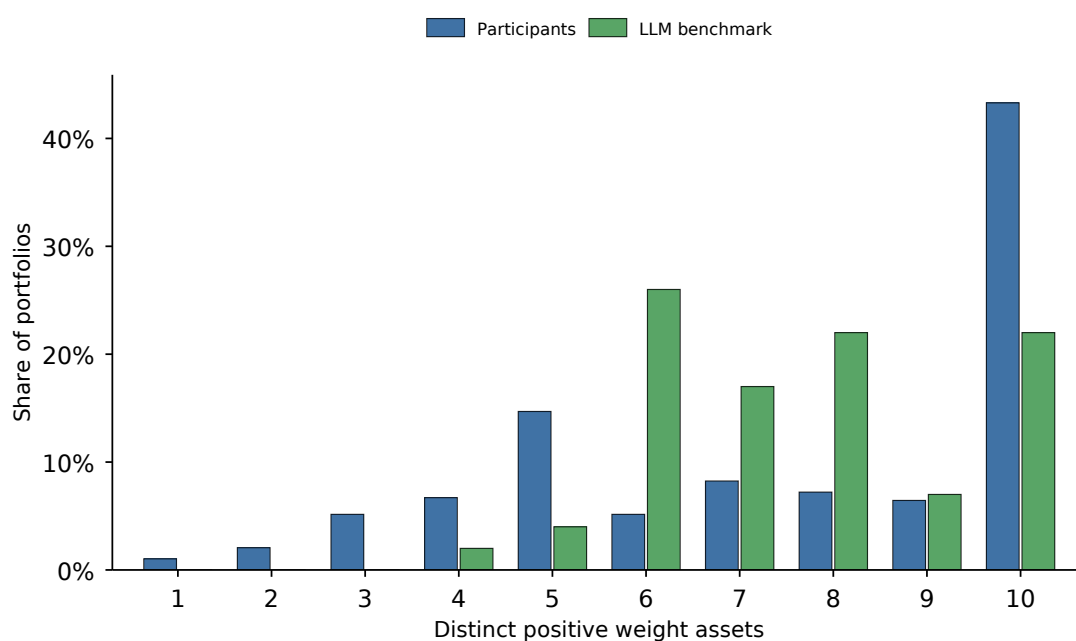
Notes: The figure ranks the 20 tickers receiving the largest shares of total dollars allocated across subjects' Round-1 portfolios. Bar colors denote GICS sectors.

Figure A.4: Most-selected tickers by LLM dollar allocation



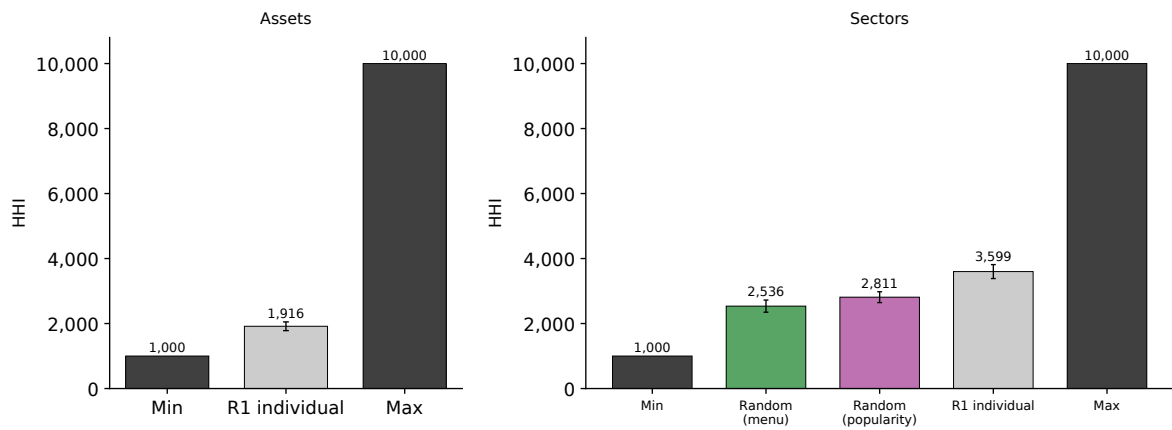
Notes: The figure ranks the 20 tickers receiving the largest shares of total dollars allocated across the 100 LLM portfolios. Each portfolio allocates \$1,000 across up to ten stocks. Bar colors denote GICS sectors and match those in Figure A.3.

Figure A.5: Participant portfolio breadth relative to the LLM benchmark



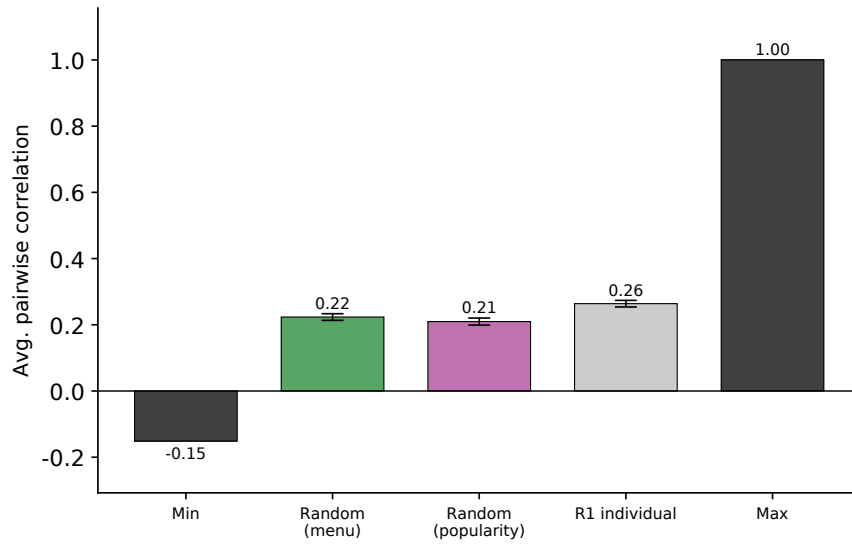
Notes: The figure compares the number of distinct positive weight assets in 388 participant Round-1 portfolios with 100 independently generated LLM benchmark portfolios. Bars show the percentage of portfolios holding each number of distinct assets. Appendix Table A.9 reports the differences between participant and LLM portfolios.

Figure A.6: Round-1 concentration (HHI) against theoretical and random benchmarks



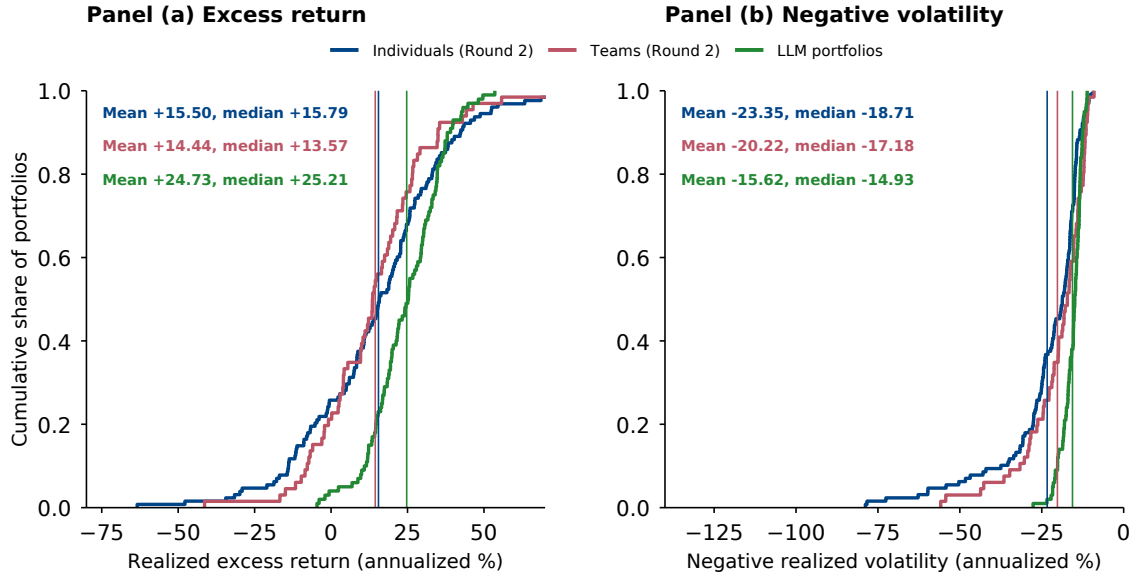
Notes: Mean Round-1 HHI for allocations across assets (left) and sectors (right). Asset-level HHI is plotted against its theoretical minimum (equal weights over the held stocks) and the theoretical maximum (a single asset). Sector-level HHI is plotted against the theoretical minimum and random-selection benchmarks that fix the number of assets to the participant distribution and draw tickers from the S&P 500 menu (uniformly and in proportion to Round-1 popularity). Bars are 95% confidence intervals around the sample mean. Intervals for the random benchmarks reflect Monte Carlo uncertainty conditional on the observed inputs.

Figure A.7: Round-1 average pairwise correlation against theoretical and random benchmarks



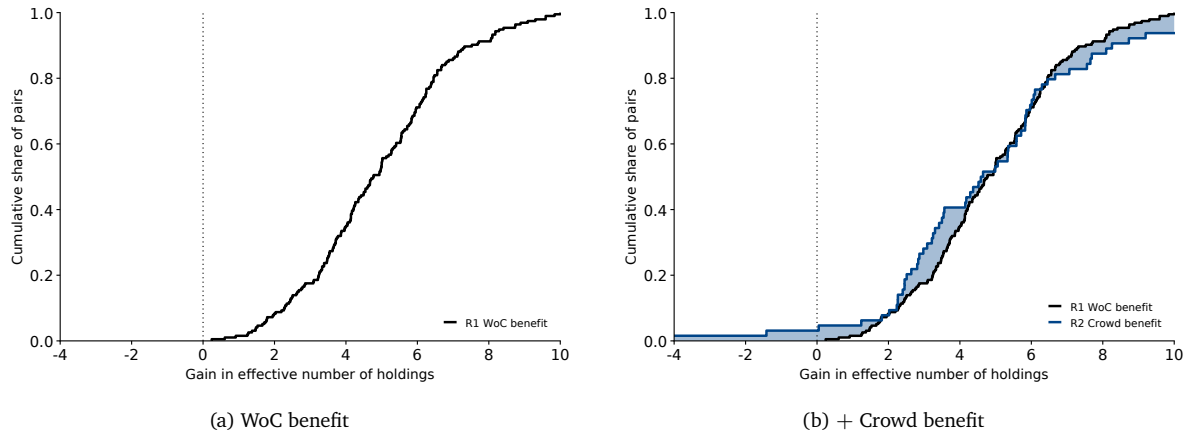
Notes: Mean Round-1 average pairwise correlation (APC) of subjects' holdings, against the theoretical minimum for a portfolio of the average observed size ($-1/(\bar{n} - 1)$), the theoretical maximum (perfectly correlated holdings), and random-selection benchmarks that fix the number of assets to the participant distribution and draw tickers from the S&P 500 menu (uniformly and in proportion to Round-1 popularity, 200 draws each). Correlations are computed on daily returns over the two years to the Round-1 date. Subjects' APC is weighted by the dollars allocated to each holding; the random benchmarks are equal-weighted, which is the same estimator applied to a uniform weight vector. Whiskers show 95% confidence intervals for the means. Intervals for the random benchmarks reflect Monte Carlo uncertainty conditional on the observed inputs. Based on the 384 subjects holding at least two priced menu names.

Figure A.8: Individual, Team, and LLM portfolio performance



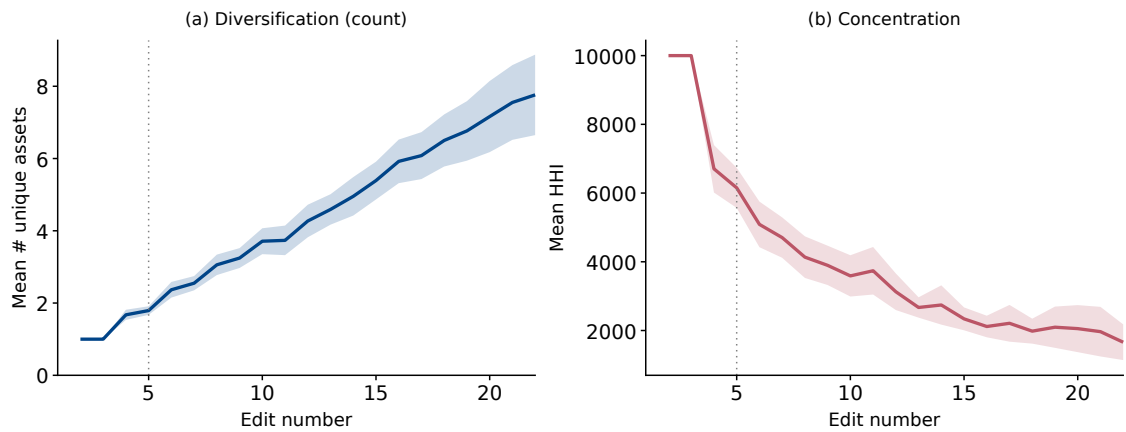
Notes: This figure reports cumulative distributions of realized excess return in Panel (a) and negative realized volatility in Panel (b) for 128 unpooled individual Round-2 portfolios from the *Crowd* condition, 66 *Team* Round-2 portfolios and 100 LLM portfolios. Individuals and LLMs can select up to ten stocks, while *Teams* can select up to twenty. Outcomes use the same evaluation window and definitions as in Section 4.2. The sign of volatility is reversed so higher values indicate lower risk. Vertical lines mark means. The return axis is truncated at 70%, while the distributions, means and medians use all observations.

Figure A.9: Mechanical WoC benchmark and realized Crowd benefit (pair level)



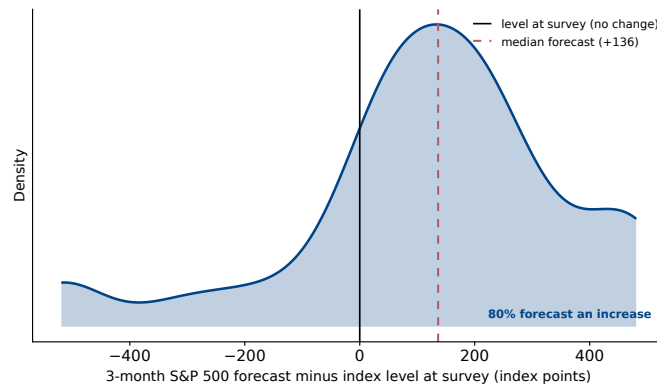
Notes: Cumulative distribution of pair-level gains in the effective number of holdings. Effective holdings are calculated as $10,000/\text{HHI}$. Panel (a) plots the gain from pooling the Round-1 portfolios of all 194 pairs, using $10,000/\text{HHI}_{ij}^{R1}$ as each pair's individual benchmark. Panel (b) adds the corresponding gains for the pooled Round-2 portfolios of the 64 *Crowd* pairs. Positive values indicate greater diversification.

Figure A.10: Portfolio structure over the sequence of Round-2 edits



Notes: Mean number of unique assets (Panel (a)) and mean HHI (Panel (b)) in *Team* pairs' joint Round-2 portfolios as a function of the edit index, reconstructed by replaying the timestamped edit log for the shared portfolio and recomputing holdings after every edit. Shaded bands are pointwise 95% confidence intervals around the sample mean across pairs; edit indices are shown where at least 20 pairs are still editing. The dotted line marks the fifth edit.

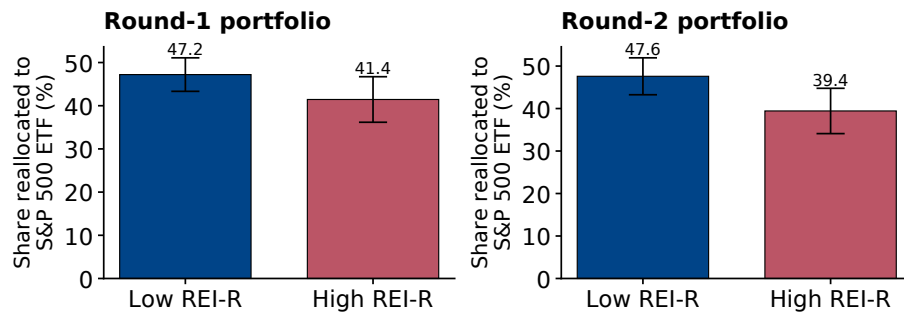
Figure A.11: Distribution of subjects' 3-month S&P 500 price forecasts



Notes: This figure displays the distribution of subjects' three-month forecasts of changes in the S&P 500. Each observation is a subject's forecast minus the level of the S&P 500 from the day that they took the intake survey. Zero therefore represents no expected change, while positive values represent an expected increase in the index. The curve is a Gaussian kernel density estimate of the forecasts, winsorized at the 5th and 95th percentiles to limit the influence of apparent data-entry errors. The solid vertical line marks zero, while the dashed vertical line marks the sample median of 136 index points above the reference level. We use the most recent completed daily close available when each subject began the intake survey as their reference level. The sample consists of the 388 subjects with a valid forecast.

Figure A.12: ETF reallocation by analytic thinking style (REI-R median split)

Index (ETF) reallocation by analytic thinking (REI-R)



Notes: Mean ETF reallocation after Round 1 and Round 2, for participants below and above the sample median of the analytic thinking-style score (REI-R). ETF reallocation is the share of the portfolio a participant chooses to move into an S&P 500 ETF, as defined in Section 4.2, so higher values indicate more reallocation. The unit of observation is the participant, and the sample is all participants with a valid Round-2 submission. Bars are 95% confidence intervals around the sample mean. Standard errors are clustered by pair.

A.3 Experimental instructions

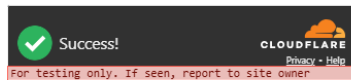
A.3.1 Intake survey

Welcome!

This survey is a signup form for an online **investing tournament** taking place on **Tuesday, June 2nd**, at **11am EDT* / 4pm BST*** and lasting approximately **30 minutes**. Your payment for signing up and completing this form is **\$5**. Your tournament show-up fee will be **\$6**, and you can win **bonus payments**.

You will only receive tournament-related payments if you **sign up, show up, and participate at exactly the specified time**. To participate in the tournament, you will need to **stay the whole time**.

More information and the **official sign-up** will follow on the next pages. But first, **check your calendar**. If you cannot attend, please **exit the survey** and **return the task** on Prolific.



I am available at 11am EDT* / 4pm BST* on Tuesday, June 2nd

I am not available at that time

This is just to check your availability. You will decide later whether you wish to participate.

* 8am Pacific (PDT, LA/SF) = 10am Central (CDT, Chicago/Houston) = 11am Eastern (EDT, NY/DC) = 4pm London (BST)

Figure A.13: Welcome (part 1)

Welcome!

Now: you will be paid **\$5** for completing this survey, which involves answering some questions and a brief comprehension check.

Tuesday (June 2nd): you will **design a portfolio** alone and/or jointly with another participant (assigned to you). Arriving on time ensures teammates can work together; being late might prevent you from participating and from receiving any payments.

The **sign-up form** is on the next page. Further instructions will follow via a **Prolific message** shortly before the tournament starts.

Next

Figure A.14: Welcome (part 2)

Consent and conditions of participation

- ☐ I will participate in the **online investing tournament**, taking place at **11am EDT / 4pm BST** on Tuesday (June 2nd), in which I will **design a stock portfolio** alone and/or with an assigned teammate.
- ☐ I understand that **I will only be paid** if I **sign up, show up on time**, and **participate** for the **full tournament duration (~30 minutes)**. My payments include the \$6 show-up fee and bonus payments of up to \$10.

I understand that arriving on time ensures teammates can work together; being late might mean I cannot participate and might not get paid.
- ☐ I confirm that I will answer all questions and complete all tasks to the best of my ability, with appropriate attention and without assistance from AI tools. I understand that if the researchers detect **AI use or insufficient attention**, I may be **excluded from the tournament** and forfeit tournament-related payments.
- ☐ I have read the privacy policy below and **consent to taking part in this study**. By consenting, I **agree to the processing of my data for this research project**.

► *Click here to display the full privacy policy*

Next

Figure A.15: Informed consent

Prolific ID

Please enter your unique Prolific ID below.

Next

Figure A.16: Prolific identifier

Your responses really matter

The responses you provide in this study are very valuable to the researchers conducting this work.

Please take a moment to read the instructions on each page carefully, consider the problems thoroughly, and respond honestly and to the best of your ability.

When you're ready, click Next to continue.

Next

Figure A.17: Study importance

Questions about you

Please choose the option that best expresses your opinion of whether each statement **applies to you**.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I try to avoid situations that require thinking in depth about something.	☐	☐	☐	☐	☐
I enjoy intellectual challenges.	☐	☐	☐	☐	☐
I am not very good in solving problems that require careful logical analysis.	☐	☐	☐	☐	☐
I don't like to have to do a lot of thinking.	☐	☐	☐	☐	☐
I enjoy problems that require hard thinking.	☐	☐	☐	☐	☐
I am not a very analytical thinker.	☐	☐	☐	☐	☐
Reasoning things out carefully is not one of my strong points.	☐	☐	☐	☐	☐
I prefer complex problems to simple problems.	☐	☐	☐	☐	☐
I am much better at figuring things out logically than most people.	☐	☐	☐	☐	☐
I have a logical mind.	☐	☐	☐	☐	☐
Using logic usually works well for me in figuring out problems in my life.	☐	☐	☐	☐	☐
Knowing the answer without having to understand the reasoning behind it is good enough for me.	☐	☐	☐	☐	☐

Next

Figure A.18: Rational-Experiential Inventory: rational (analytic) items

Questions about you

Please choose the option that best expresses your opinion of whether each statement **applies to you**.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I enjoy learning by doing something, instead of figuring it out first.	☐	☐	☐	☐	☐
I tend to use my heart as a guide for my actions.	☐	☐	☐	☐	☐
I don't think it is a good idea to rely on one's intuition for important decisions.	☐	☐	☐	☐	☐
I like to rely on my intuitive impressions.	☐	☐	☐	☐	☐
I trust my initial feelings about people.	☐	☐	☐	☐	☐
For me, descriptions of actual people's experiences are more convincing than discussions about "facts."	☐	☐	☐	☐	☐
I'm not a very spontaneous person.	☐	☐	☐	☐	☐
I can often tell how people feel without them having to say anything.	☐	☐	☐	☐	☐
I often go by my instincts when deciding on a course of action.	☐	☐	☐	☐	☐
I generally don't depend on my feelings to help me make decisions.	☐	☐	☐	☐	☐

Next

Figure A.19: Rational-Experiential Inventory: experiential (intuitive) items

Questions about you

Please choose the option that best expresses your opinion of whether each statement **applies to you**.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I prefer chaotic action to orderly inaction.	☐	☐	☐	☐	☐
I find that to adopt a careful, analytical approach to making decisions takes too long.	☐	☐	☐	☐	☐
I am more at home with ideas rather than facts and figures.	☐	☐	☐	☐	☐
When making a decision, I take my time and thoroughly consider all relevant factors.	☐	☐	☐	☐	☐
I am inclined to scan through reports rather than read them in detail.	☐	☐	☐	☐	☐
I find detailed, methodical work satisfying.	☐	☐	☐	☐	☐
The kind of work I like best is that which requires a logical, step-by-step approach.	☐	☐	☐	☐	☐
Formal plans are more of a hindrance than a help in my work.	☐	☐	☐	☐	☐
My understanding of a problem tends to come more from thorough analysis than flashes of insight.	☐	☐	☐	☐	☐
I make decisions and get on with things rather than analyze every last detail.	☐	☐	☐	☐	☐

Next

Figure A.20: Cognitive Style Index (unidimensional) items

Probability and logic questions

Please answer the following questions to the best of your ability.

Five-sided die question

Imagine we are throwing a five-sided die 25 times. On average, how many times would this die show an odd number (1, 3, or 5)?

Two guards question

You are a prisoner in a room with 2 doors and 2 guards. One of the doors will bring you to freedom, the other to death. One guard is a truth-teller, the other is a never-liar. You can only ask one question to one guard. What do you do to find the door leading to freedom?

Poisonous plant question

In a forest, 20% of plants are red, 50% are brown, and 30% are white. A red plant is poisonous with a probability of 40%. A plant that is not red is poisonous with a probability of 10%.

What is the probability that a poisonous plant in the forest is red? (0–100)

Loaded die question

Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 2 is twice as high as the probability of each of the other numbers. On average, out of 35 throws, how many times would the die show the number 2?

Doctor riddle

A woman and her son are in a car accident. The woman is sadly killed. The boy is rushed to hospital. When the doctor sees the boy, he says "I can't operate on this child, he is my son." Who is the doctor, in relation to the boy? (one word)

Bowling question

Out of 2,000 people in a town, 1,000 are members of a bowling team. Out of these 1,000 members, 200 are men. Out of the 1,000 inhabitants that are not on the bowling team, 600 are men.

What is the probability that a randomly drawn man is a member of the bowling team? (0–100)

Next

Figure A.21: Berlin Numeracy Test items

Lottery choices

In this task, you will make **6 decisions in total**. Each decision is a hypothetical choice between a **sure payment** and a **lottery**. The lottery is always the same, but the sure payment will change based on your choices.

Please read carefully and choose the option you truly prefer in each step. Your choices do not directly affect your payment in this study. There are no right or wrong answers.

In this decision, please choose between:

- **Option A (Sure amount):** receive **\$150** for sure
- **Option B (Lottery):** **50% chance** of \$300 and **50% chance** of \$0

This is decision **1** of **6**. After you make a choice on this screen, the sure amount will change on the next screen.

Choose Option A (Sure amount)

Choose Option B (Lottery)

Figure A.22: Risk preferences

Understanding the Sharpe ratio

Below is some information that will help you make decisions in the investing tournament.
You will **only be invited** to participate if you **answer all questions below correctly within 2 attempts**.

1) Risk and return

Investing usually involves a **trade-off between risk and return**. Choosing a few well-performing companies can generate high returns but carries more risk than spreading investments widely. In this study, your objective is to **build portfolios with a high Sharpe ratio**, which rewards strong returns relative to risk.

Figure A.23: Sharpe ratio explainer (part 1)

2) Core definition and formula

The Sharpe ratio measures how much **extra return** you earn for each unit of **risk**. Higher values indicate better risk-adjusted performance.

$$\text{Sharpe ratio} = \frac{\text{Portfolio return} - \text{Risk-free return}}{\text{Standard deviation of portfolio returns}}$$

Portfolio return: How much your portfolio earned (e.g., 10% growth in your portfolio's value).

Risk-free return: The return you could earn with near-zero risk, e.g., by investing in a U.S. government bond. We subtract this to capture your portfolio's extra returns above the risk-free rate.

Standard deviation: Volatility, or how much your portfolio's returns fluctuated over the investment period. A higher value means bigger swings up and down in your portfolio's returns, i.e., more unpredictability and risk.

Figure A.24: Sharpe ratio explainer (part 2)

3) Example and takeaway

If a portfolio returns 12%, the risk-free return is 4%, and standard deviation is 16%, then the Sharpe ratio = $(12\% - 4\%) / 16\% = 0.50$. That means 0.50 units of extra return per unit of risk.

That example shows the **main takeaway**: you can raise the Sharpe ratio by **increasing returns, reducing volatility, or both**. Simply picking stocks with the highest return is not enough — you also need to manage risk. One practical way to do that is to invest in multiple stocks, especially stocks that are uncorrelated with each other.

Figure A.25: Sharpe ratio explainer (part 3)

4) Comprehension questions

Please answer all questions below. Correctly answering all three within 2 attempts is a prerequisite for tournament invitation.

Important: if your first attempt is incorrect, you have one more attempt to meet this prerequisite.

1. What does the Sharpe ratio measure?

- ☐ The total dollar value of a portfolio
- ☐ The extra return earned per unit of risk taken
- ☐ The number of stocks in a portfolio
- ☐ The total return of a portfolio, regardless of risk

2. Consider two portfolios (assume a risk-free rate of 4%):

Portfolio A earns a 10% return with a standard deviation (risk) of 20%.

Portfolio B earns an 8% return with a standard deviation (risk) of 10%.

Which portfolio has the higher Sharpe ratio?

Figure A.26: Sharpe ratio explainer (part 4)

2. Consider two portfolios (assume a risk-free rate of 4%):

Portfolio A earns a 10% return with a standard deviation (risk) of 20%.

Portfolio B earns an 8% return with a standard deviation (risk) of 10%.

Which portfolio has the higher Sharpe ratio?

- ☐ Portfolio A
- ☐ Portfolio B
- ☐ They have the same Sharpe ratio

3. Which of the following approaches is most likely to generate a high Sharpe ratio?

- ☐ Investing all of your money in a single high-return stock
- ☐ Picking multiple stocks, especially those that move together (correlated)
- ☐ Picking multiple stocks, especially those that do not move together (uncorrelated)
- ☐ Investing all of your money in a single low-risk stock

Figure A.27: Sharpe ratio explainer (part 5)

Great job, you are almost done with the survey!

This is the final section. Please answer the questions below.

Which of the following best describes your gender identity?	☒ Man ☐ Woman ☐ Other / non-binary
What is your age?	22
What is your country of residence?	Select your country
Which time zone are you in?	-----
What is the highest level of education you have achieved?	-----
Which of the following best describes your race or ethnicity? (Select all that apply)	☐ White ☐ Black or African American ☐ Hispanic or Latino/a ☐ Asian (East Asian) ☐ Asian (South Asian) ☐ Asian (Southeast Asian) ☐ Middle Eastern or North African ☐ Native American or Alaska Native ☐ Native Hawaiian or Other Pacific Islander ☐ Multiracial ☐ Another race or ethnicity
Please describe your race or ethnicity (if 'Another race or ethnicity' selected)	
What industry do you primarily work in?	-----
Please describe your industry (if 'Other' selected)	

Figure A.28: Demographics (part 1)

Which industries are you most informed about? (Select all that apply)	☐ Energy ☐ Materials ☐ Industrials ☐ Consumer discretionary ☐ Consumer staples ☐ Health care ☐ Financials ☐ Information technology ☐ Communication services ☐ Utilities ☐ Real estate ☐ Other				
Please describe the industry you are most informed about (if 'Other' selected)					
How much experience do you have with trading stocks and/or other securities?					
Please rate your agreement with the following statement: <i>"Stock prices already reflect what's publicly known, so it is very hard to reliably pick stocks that will beat the market."</i>	Strongly disagree ☐	Disagree ☐	Neither agree nor disagree ☐	Agree ☐	Strongly agree ☐
The S&P 500 index is currently at 7383.74 . Please enter your best estimate for the value of the S&P 500 in three months' time:					
How reliable are your answers to this survey?	Not at all reliable ☐	Slightly reliable ☐	Moderately reliable ☐	Very reliable ☐	Extremely reliable ☐
Do you have any thoughts or comments about this study that you would like to share with us?					

[Next](#)

Figure A.29: Demographics (part 2)

Notes: The S&P 500 level shown in the forecast question was fetched live when the page loaded. This screenshot is from an earlier test session, so the level subjects saw differs slightly depending on when they completed the survey.

Tournament reminder

The tournament is taking place at **11am EDT*** / **4pm BST*** on **Tuesday, June 2nd**. Please add it to your calendar.

In your selected time zone (**Mountain Time (e.g., Denver, Calgary)**), that is **9:00 AM** on **Tuesday, June 2nd**.

Add to calendar (.ics file)

By showing up on time and participating for the full session, you will receive a show-up fee of **\$6**, and you may be selected to receive a **bonus payment of up to \$10** based on the performance of your portfolio.

☐ I still plan to attend the tournament on Tuesday, June 2nd at 11am EDT / 4pm BST.

You may now click the button below to return to Prolific. Further instructions will follow via Prolific.

* 8am Pacific (PDT, LA/SF) = 10am Central (CDT, Chicago/Houston) = 11am Eastern (EDT, NY/DC) = 4pm London (BST)

Next

Figure A.30: Reminder for main session

A.3.2 Main experiment

Welcome!

Thank you for joining. You will be paid **\$6** for participating and can earn bonus payments up to **\$10**.

This tournament consists of **two rounds** with fixed durations. In each round, you will design a portfolio. Round 1 will start after a few introductory pages and runs until **11:12 AM EDT**. Round 2 will begin shortly thereafter and run until **11:25 AM EDT**. After the two rounds, you will make additional decisions that affect your portfolios.

Taking into account the time needed to read instructions, each round will last approximately 10 minutes. If you finish a round early, please stay on the survey page and wait for the next round to start.

Leaving the survey early and returning later might **lead to your exclusion from the tournament**. This is a necessary precaution to ensure the integrity of the study.

You have the chance to win bonus payments. These payments will depend on the **Sharpe ratio** of one **randomly selected** portfolio of yours, so you need to take both rounds seriously. More information on incentives and the Sharpe ratio will follow on the next pages.

Continue to the next page to give your consent.

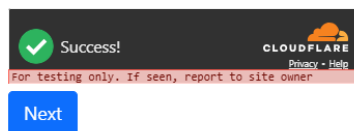


Figure A.31: Welcome

Consent and conditions of participation

☒ I have read the privacy policy below and **consent to take part in this study**. By consenting, I **agree to the processing of my data for this research project**.

► *Click here to display the full privacy policy*

Next

Powered by [oTree](#)

Figure A.32: Informed consent

Your responses really matter

The responses you provide in this study are very valuable to the researchers conducting this work.

Please take a moment to read the instructions on each page carefully, consider the problems thoroughly, and respond honestly and to the best of your ability.

When you're ready, click Next to continue.

Next

Figure A.33: Study importance

Prolific ID

Please enter your unique Prolific ID below.

Next

Powered by [oTree](#)

Figure A.34: Prolific identifier

Time left for this task: **5:15**

Investment task: Round 1

Please complete this task as if you have \$1,000 to allocate to stocks in the S&P 500. Choose your investment(s) by selecting a ticker name and the amount invested for **at least one**, and **at most ten** stocks.

We will calculate your portfolio's Sharpe ratio using **real returns data** starting from today through 3 months from today. You can earn up to \$10 by maximizing your Sharpe ratio. As such, your best strategy is to design a portfolio with high returns and low volatility. More info is available below.

► *More information on incentives*

To qualify for bonus payments, your portfolio must have **exactly \$1,000** invested. You cannot continue until the timer has ended. When time is up, the page will **automatically advance**, regardless of whether you have invested \$1,000.

Total invested: \$0 / \$1,000

Stock 1 Ticker:	Amount (\$):
	0
Stock 2 Ticker:	Amount (\$):
	0
Stock 3 Ticker:	Amount (\$):
	0

Figure A.35: Round-1 portfolio interface

▼ *More information on incentives*

You are incentivized to maximize the Sharpe ratio of your portfolio, which can be thought of as the ratio of extra returns earned to risk taken. The ex-post Sharpe ratio of your portfolio, and the portfolios of other participants, will be calculated after 3 months. We will randomly select every 10th participant to receive a prize payment. If you are selected and your Sharpe ratio is in the top 10% of participant portfolios, then you will be paid \$10. If it is only in the top 20%, you will be paid \$9, and so on. If it is in the bottom 10%, you will be paid \$1. As such, your best strategy is to design a portfolio with a high Sharpe ratio, i.e., high returns and low risk.

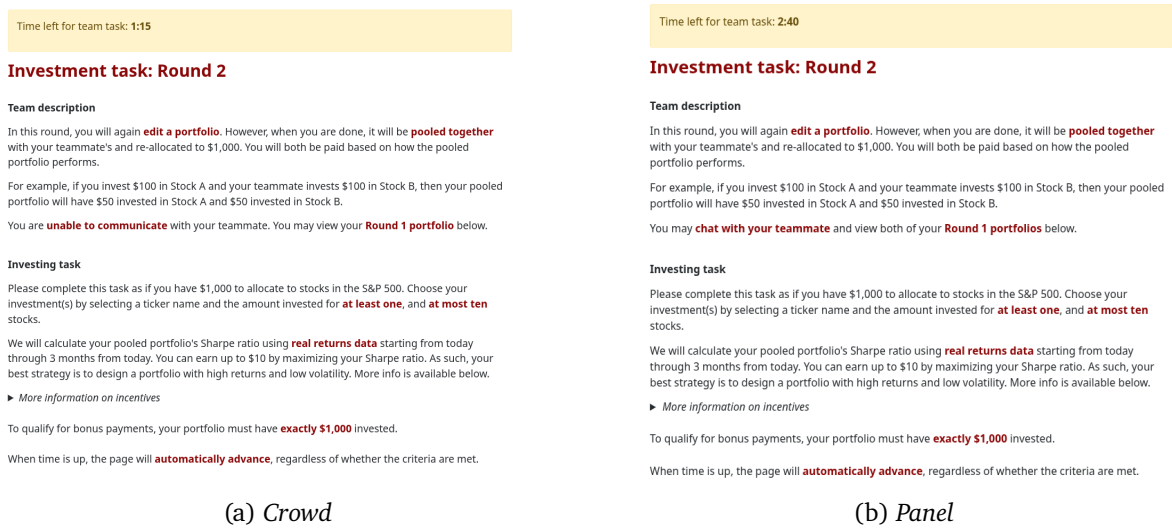
Figure A.36: More information on incentives

Portfolio confidence

How **certain** are you that the portfolio you just designed will generate a high Sharpe ratio?



Figure A.37: Round-1 confidence elicitation



Time left for team task: 2:40

Investment task: Round 2

Team description

In this round, you will again **edit a portfolio**. However, when you are done, it will be **pooled together** with your teammate's and re-allocated to \$1,000. You will both be paid based on how the pooled portfolio performs.

For example, if you invest \$100 in Stock A and your teammate invests \$100 in Stock B, then your pooled portfolio will have \$50 invested in Stock A and \$50 invested in Stock B.

You may **chat with your teammate** and view both of your **Round 1 portfolios** below.

Investing task

Please complete this task as if you have \$1,000 to allocate to stocks in the S&P 500. Choose your investment(s) by selecting a ticker name and the amount invested for **at least one**, and **at most ten** stocks.

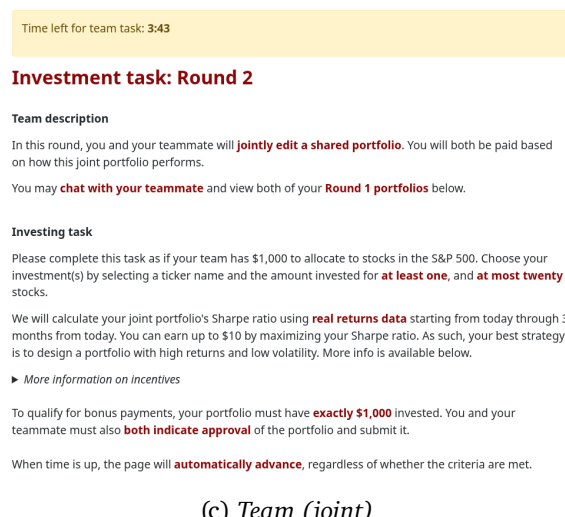
We will calculate your pooled portfolio's Sharpe ratio using **real returns data** starting from today through 3 months from today. You can earn up to \$10 by maximizing your Sharpe ratio. As such, your best strategy is to design a portfolio with high returns and low volatility. More info is available below.

► *More information on incentives*

To qualify for bonus payments, your portfolio must have **exactly \$1,000** invested.

When time is up, the page will **automatically advance**, regardless of whether the criteria are met.

(b) Panel



(c) Team (joint)

Figure A.38: Round-2 portfolio interface (top of page) by decision mode

Show my Round 1 portfolio

You have designed a portfolio of \$1,000. As such, you have met the eligibility requirement for this round. Feel free to keep editing your portfolio until the timer runs out, but only your latest valid \$1,000 portfolio will be used.

Total invested: \$1,000 / \$1,000

Stock 1 Ticker:	Amount (\$):
PPG — PPG Industries Inc. — Materials	1000
Stock 2 Ticker:	Amount (\$):
	0

(a) Crowd

Team chat

You: [11:37:42] [bot55 seq=11 t=15:37:41] load test

Teammate: [11:37:43] [bot70 seq=11 t=15:37:43] load test

You: [11:38:13] [bot55 seq=12 t=15:38:13] load test

Teammate: [11:38:14] [bot70 seq=12 t=15:38:14] load test

Type your message... Send

Show my Round 1 portfolio

Show teammate's Round 1 portfolio

You have designed a portfolio of \$1,000. As such, you have met the eligibility requirement for this round. Feel free to keep editing your portfolio until the timer runs out, but only your latest valid \$1,000 portfolio will be used.

Total invested: \$1,000 / \$1,000

Stock 1 Ticker:	Amount (\$):
SO — Southern Co (The) — Utilities	1000
Stock 2 Ticker:	Amount (\$):
	0

(b) Panel

Team chat

You: [11:32:34] agielnt

Type your message... Send

Show my Round 1 portfolio

Show teammate's Round 1 portfolio

Team agreement status

Teammate: ✓ Agreed

You: ✓ Agreed

I agree to these investments

You have both agreed to a portfolio of \$1,000. As such, you have met the eligibility requirement for this round. Feel free to keep editing your portfolio until the timer runs out, but just make sure to both press "I agree to these investments" again if you want to save your edits.

Total invested: \$1,000 / \$1,000

Stock 1 Ticker:	Amount (\$):
KMB — Kimberly-Clark Corporation — Consumer Staples	1000
Stock 2 Ticker:	Amount (\$):
	0

(c) Team (joint)

Figure A.39: Round-2 portfolio interface (bottom of page) by decision mode

Portfolio confidence

How **certain** are you that your Round 2 portfolio will generate a high Sharpe ratio?

▼ View your Round 2 portfolio

Ticker	Amount (\$)
SYY	\$500
PFG	\$500

0 – Not at all certain

100 – Extremely certain

Next

Powered by [oTree](#)

Figure A.40: Round-2 confidence elicitation

Explain your decision

Below we remind you of the confidence levels you stated for your **Round 1** and **Round 2** portfolios. Please briefly explain why you chose each level.

Your previously stated confidence in your portfolio in Round 1 was **74/100**. Please briefly explain why you chose this confidence level.

▼ View your Round 1 portfolio

Ticker	Amount (\$)
DASH	\$1000

Your previously stated confidence in your portfolio in Round 2 was **6/100**. Please briefly explain why you chose this confidence level.

▼ View your Round 2 portfolio

Ticker	Amount (\$)
SYN	\$500
PFG	\$500

Next

Figure A.41: Open-text elicitation about confidence ratings

Round 2 portfolio description

Please describe, in your own words, how your team (i) communicated, (ii) decided which stocks to include, and (iii) allocated the \$1,000.

▼ View your Round 2 portfolio

Ticker	Amount (\$)
SYY	\$500
PFG	\$500

How willing would you be to work with this person again on a similar task?

0 – not willing at all

100 – very willing

How much of your attention and effort did you allocate to **maximizing returns** versus **minimizing risk**?

0 – All on minimizing risk

100 – All on maximizing returns

Next

Figure A.42: Open-text elicitation about Round 2 experience and risk-return trade-off

Portfolio re-allocation

You now have the option to **re-allocate** some of your portfolio holdings to an **ETF** (exchange-traded fund) that tracks the S&P 500. Your reallocation choice has a small chance of being implemented before bonus payments are calculated.

An ETF is a type of investment fund that holds a **collection of assets**, such as stocks. They can be traded on stock exchanges like individual stocks. ETFs are usually designed to return identical performance to the index that they track. It combines the benefits of owning a diversified portfolio with the flexibility of trading like a stock.

The ETF we use is the **SPDR S&P 500 ETF Trust (SPY)**, which tracks the performance of the S&P 500.

For each portfolio, choose what percentage of your **\$1,000** investment should go to your selected stocks, with the remainder automatically allocated to the SPY ETF.

For example, if your Round 2 portfolio allocated \$1000 to Stock A and you choose 70%, then \$700 will be invested in Stock A and \$300 will be re-allocated to the ETF.

If implemented, your bonus payment will be based on the Sharpe ratio of this compound portfolio (your stocks + SPY).

► *More Information on incentives*

Benefit: The ETF captures overall market returns and provides diversification across many stocks.

Drawback:

- Allocating more to the ETF reduces your opportunity to outperform the market through active stock selection.
- Allocating more to the ETF reduces your opportunity to design a portfolio based on your own beliefs and judgment.

Figure A.43: Description of choice of whether to reallocate portfolios to an ETF

Round 1 ETF allocation

0% – All SPY ETF

100% – All your stocks

Please briefly explain the reasoning for your chosen Round 1 ETF allocation:

▼ View your Round 1 portfolio

Ticker	Amount
DASH	\$1000

Round 2 ETF allocation

0% – All SPY ETF

100% – All your stocks

Please briefly explain the reasoning for your chosen Round 2 ETF allocation:

► View your Round 2 portfolio

Next

Figure A.44: Sliders for reallocation decision

Final comments

Thank you for taking part in this study. You will receive your show-up fee within a few days if you successfully completed all stages of the tournament. Any investment-based winnings will come in about three months.

If you have any comments about the tasks, the instructions, or anything else related to the study, you can write them in the box below. This is optional.

Next

Powered by oTree

Figure A.45: Final comments