Delegated Decision-Making in Finance

Felix Holzmeister[†] Martin Holmén^{‡,*} Michael Kirchler^{†,‡}
Matthias Stefan[†] Erik Wengström^{§,¶}

[†] University of Innsbruck, Department of Banking and Finance

[‡] University of Gothenburg, Department of Economics, Centre for Finance

[§] Lund University, Department of Economics

[¶] Hanken School of Economics, Department of Finance and Economics

* Corresponding author: martin.holmen@gu.se

Abstract

We run an online experiment with 408 finance professionals (money managers) and 550 subjects from the general population in Sweden (clients). We examine drivers of clients' delegation decisions, differences in decision-making quality between both groups, and professionals' ability to implement investment portfolios that suit the clients' risk attitudes. We find that clients' trust in money managers increases the likelihood of delegating their investment decisions, whereas decision-making quality is associated with a decrease. We further show that decision-making quality of finance professionals is not significantly higher compared to their clients' when controlling for risk taking. Finally, we observe high variability among professionals' perception of delegated risk levels and overlaps in portfolio risk across self-reported risk-levels of clients. This finding indicates that communicating risk between clients and professionals constitutes a potential pitfall in delegated investment decisions.

JEL: C93, G11, G41.

Keywords: Experimental finance, finance professionals, delegated decision-making.

We thank Pol Campos-Mercade, Alain Cohn, Dawei Fang, Christian König-Kersting, Michael Razen, Markus Walzl, seminar participants at the Berlin Behavioral Economics Colloquium and Seminar, the Helsinki GSE Colloquium, as well as conference participants of the Organizations and Society Workshop 2019 at the University of Innsbruck, the NOeG 2019 (Austrian Economic Association) in Graz, the SFB Workshop 2019 in Innsbruck, the Experimental Finance Conference 2019 in Copenhagen, the Decision-Making for Others Workshop in Nijmegen 2018, and the Workshop on Trust and Cooperation in Markets and Organizations in Stavanger 2019 for very valuable comments. We particularly thank Fredrik Bergdahl at *Statistiska centralbyrån* for the excellent collaboration on the project. Financial support from the Austrian Science Fund FWF (START-grant Y617-G11 and SFB F63), and the Swedish Research Council (grant 2015-01713) is gratefully acknowledged. This study was ethically approved by the review boards at *Statistiska centralbyrån* (SCB; Statistics Sweden) and in Gothenburg (Sweden).

1. Introduction

Given the complexity of financial products and markets, private investors often opt for delegating decisions to finance professionals. This involves decisions about portfolio investments, insurance and pension plans, and seeking advice on various other financial aspects, all of which have a potentially strong impact on a client's wealth. The economic importance of delegated decision-making in finance is indicated by the large and growing market for financial advice and decision-making on behalf of clients. For instance, in 2017, the net asset value of US mutual funds equaled 18.8 trillion USD¹, and over 271,000 professionals were employed as personal financial advisors in the United States². Thus, it is surprising that research in economics and finance has predominantly focused on individual decision-making (e.g., Holt and Laury, 2002; Abdellaoui et al., 2011; Dohmen et al., 2011; Falk et al., 2018) without giving much consideration to how delegation decisions by clients and delegated investment decisions by finance professionals are actually taken (as emphasized by, e.g., Foerster et al., 2017; Kirchler et al., 2018a; Linnainmaa et al., 2019).

We study delegation decisions by laypeople and how finance professionals behave on their behalf by running a controlled lab-in-the field experiment implemented online with finance professionals (agents) and subjects from the general population (principals) in Sweden. We examine (i) the motivations and characteristics of principals delegating investment decisions, (ii) differences in decision-making quality between professionals and the general population, and (iii) the agents' ability to construct portfolios that suit the risk preferences of principals.

The reasons why private (and, partly, also institutional) investors delegate investment decisions to professional money managers (e.g., financial advisers or fund managers) are not well understood. Straight-forward potential explanations include that investors lack—or believe to lack—sufficient knowledge or information, or are time-constrained. An alternative motive for delegating investments may be the possibility to blame the agent if the investment does not turn out as expected (Shefrin, 2007; Chang et al., 2016).³ Moreover, Gennaioli et al. (2015) argue that investors delegate because they do not know much about finance and are too anxious to make investment decisions. Just like doctors, money managers are trusted, and they give investors confidence to take risks (see also Guiso et al., 2004, 2008, for the “trust channel” as a major motive of delegated investment decisions). In their theoretical framework Gennaioli et al. (2015) show that, under rational expectations, money managers enable investors to take more risk and, consequently, being better off.

However, since the seminal work of Jensen (1968) there is clear and persistent evidence that fund managers underperform passive benchmark indices after costs. The annual underperformance varies and mainly falls within the range of 0.6 and 2.0 percent (see, for instance, Gruber, 1996; Carhart, 1997; French, 2008). This insight renders the question of why people delegate their financial decisions more puzzling. Moreover, money managers have a strong incentive not to correct investors' biased beliefs, because they enable charging higher fees (Mullainathan et al., 2012). This finding is also related to the credence goods characteristics of financial advice (Dulleck and Kerschbamer, 2006; Inderst and Ottaviani, 2012a,b,c), outlining the prevailing information asymmetry between advisers and clients.

¹ <https://perma.cc/5VUB-U98U>

² <https://perma.cc/5RYT-CP6H>

³ For a general account of shifting blame see also Bartling and Fischbacher (2012).

In addition, not only asymmetric information, but also monetary incentives of financial professionals and their role of intensifying conflicts of interest between clients and money managers are relevant in delegating financial investment decisions. Payment schemes have become a hotly debated topic in finance, as misaligned incentives have been portrayed as major contributors to the last financial crisis (Financial Crisis Inquiry Commission, 2011; Dewatripont and Freixas, 2012). In particular, high-powered payment schemes that align professionals' incentives with clients' returns (e.g., bonus schemes or tournament incentives) have been identified among the main drivers of excessive risk taking in financial markets (Jensen and Meckling, 1976; Rajan, 2006; Diamond and Rajan, 2009; Bebchuk and Spamann, 2010). Since this debate—in particular on bonus incentives—has spilled over to the public, incentives might also play a role for decisions whether to delegate.

A growing strand of literature uses experiments with student or general population samples to examine risk taking in delegated decision-making. Several studies report a “risky shift” in risk-taking, indicating that decision-makers take more risks or show less loss-averse behavior for others than for themselves (e.g., Sutter, 2009; Chakravarty et al., 2011; Andersson et al., 2016; Vieider et al., 2016). However, a substantial number of studies also find a “cautious shift” when the money of third parties is invested (Bolton and Ockenfels, 2010; Eriksen and Kvaløy, 2010)—see Füllbrunn and Luhan (2015) and Eriksen et al. (2017) for overviews. Andersson et al. (2019) run a large-scale study with a random population sample from Denmark. The agents face high-powered incentives to increase risk-taking on behalf of others through hedged compensation contracts or tournament incentives. The authors report that the decision-makers respond to these incentives, resulting in an increased risk exposure of the principals. Yet another strand of experimental studies suggests that even strong financial incentives hardly interfere with agents' attempt to adhere to their clients' preferences (see, e.g., Rud et al., 2018; Ifcher and Zarghamee, 2019; Kling et al., 2019).

In recent years, robo advice and algorithm-based investments have emerged as an alternative to traditional financial services. Although they promise to offer affordable advice in investment matters, tailored to the clients' needs, several pitfalls remain (D'Acunto et al., 2019). Similar to the case of delegating to financial professionals, the decision to opt for robo advice is likely shaped by trust. Previous research from other decision domains provides evidence of algorithm aversion, with people distrusting advice and predictions based on algorithms more than those based on human judgement (Dietvorst et al., 2014; Harvey et al., 2017; Longoni et al., 2019). However, the evidence is mixed, and other studies report the opposite pattern of algorithm appreciation (Logg et al., 2017). To date, there is a lack of evidence on this issue in the realm of financial decision-making.

Given that finance professionals regularly make decisions on behalf of their clients and that there is a lack of knowledge regarding the motives for delegation among laypeople, it is surprising that no evidence on professionals' fiduciary and clients' delegation choices exists. In this paper, we report the results of an online experiment with participants from a sample of Swedish finance professionals and a representative sample of the Swedish general population. Via *Statistiska centralbyrån* (SCB; Statistics Sweden) invitations were sent to financial analysts, investment advisors, traders, fund managers, and financial brokers and to an equally large randomly selected sample of Swedish employees, excluding finance professionals. In particular, we address the following research questions:

RQ1: What drives clients' decision to delegate? Which economic preferences and personal characteristics influence delegation decisions, and do clients delegate to increase risk-taking? Does knowledge about the agents' financial incentives affect clients' decision whether or not to delegate, and do clients prefer delegating to an investment algorithm over a finance professional?

RQ2: Do finance professionals make better decisions? Do finance professionals systematically outperform laypeople in terms of decision-making quality, and is the agents' decision-making quality impacted by financial incentives and whether investment decisions are made on one's own account or on behalf of others?

RQ3: Can investment preferences be communicated? Can professionals construct portfolios that match the riskiness requested by their principals, i.e., can risk be communicated between principals and agents such that risky decisions can be effectively delegated?

We set up six treatments, differing in (i) the subjects enrolled (finance professionals or general population subjects), (ii) the agent, participants from the general population could delegate to (investment algorithm, linearly incentivized or flat paid finance professional), and (iii) whether finance professionals decided on their own account or on behalf of a client. In 25 investment decisions, subjects had to allocate an endowment across two or five investment alternatives that differed in their expected payout, riskiness, and diversification potential. Subjects from the general population were thereafter given the opportunity to delegate their decisions, by replacing their own investments with those of a financial professional/investment algorithm. In total, 408 finance professionals and 550 people from the general population completed the experiment. A set of predefined variables of the subjects' register data for those who completed the experiment were provided by SCB after the experiment.

Our study provides the following insights. First, we show that clients in our setting are most likely to delegate to an investment algorithm, followed by professionals with aligned incentives and professionals with fixed incentives. We further observe that clients' propensity to delegate their decisions decreases with their own decision-making quality, but increases with trust in the agent and their propensity to shift blame on others. Moreover, clients delegating their decisions, on average, request the agents to take more risk than they perceive they took for themselves. Second, we find that the overall decision-making quality of finance professionals is not significantly higher compared to clients in our sample, leaving little room for delegation being superior for principals. Finally, we report that communication of investment preferences, based on the four risk-levels clients choose from when delegating their investment decisions, constitutes a potential pitfall. In particular, we find considerable overlaps in the risk of portfolios implemented by finance professionals on behalf of clients with varying investment preferences, which implies that clients requesting different levels of risk eventually may end up with very similar portfolios.

Our study adds to several emerging areas in the literature. First, we contribute to the expanding literature on delegated decision-making for third parties in financial frameworks. When it comes to the decision whether to delegate, our findings are in line with the literature on algorithm appreciation, trust, and blame-shifting. In particular, we provide evidence on the relevance of trust in financial professionals for delegation decisions (Lachance and Tang, 2012). This finding is also closely related to the rationale discussed by Gennaioli et al. (2015). Just as doctors are trusted by patients, "money doctors" (Gennaioli

et al., 2015) are trusted when investing money on behalf of their clients, even if the outcome is not significantly better. Our finding that many laypeople request professionals to take more risk, than they take themselves, is related to the notion that investors “are too nervous or anxious to make risky investments on their own” (Gennaioli et al., 2015, p.92). For these clients, delegation serves as a way to increase risk taking, and thereby expected returns.

Moreover, following our results, the communication of risk between money managers and clients appears to be difficult, as the same portfolio risk can be considered very differently by either party. Several studies show that clients’ portfolio risk depends on professionals’ risk attitudes (see, e.g. Foerster et al., 2017; Kirchler et al., 2018a; Linnainmaa et al., 2019). In an experimental study with finance professionals, Kirchler et al. (2018a) show that professionals’ beliefs about clients’ willingness to take risks do not explain risk taking, but professionals’ self-assessed risk attitude in financial matters does. Foerster et al. (2017) report results from Canadian households and financial advisers and show that advisor fixed effects explain considerably more variation in household portfolio risk than a broad set of investor attributes. Linnainmaa et al. (2019) provide evidence from a large sample of Canadian financial advisers and their clients. The authors show that most advisers invest their personal portfolios just as they advise their clients. Hence, we contribute by showing professionals’ difficulties in implementing a suitable level of risk when constructing portfolios for their clients. Furthermore, we add to the literature on incentives of money managers. Interestingly, we find that decision-making quality is not affected by the incentives professionals face, even though this seems to be expected by clients, who delegate more frequently to professionals with aligned incentives rather than with a flat compensation.

Second, we contribute to the small but growing corpus analyzing the behavior of finance professionals. Across studies, one major result is that professionals’ behavior tends to be closer to neoclassical benchmarks compared to student subjects and representative general population samples. For instance, professionals are less prone to anchoring than students (Kaustia et al., 2008), can better discern the quality of public signals in information cascades (Alevy et al., 2007), and produce price bubbles less likely in experimental asset markets (Weitzel et al., 2019). However, other studies point towards opposite results and show that professionals exhibit a higher degree of myopic loss aversion (Haigh and List, 2005), react more strongly to rank incentives (Kirchler et al., 2018b), show herd behavior (Cipriani and Guarino, 2009) and framing heuristics (Schwaiger et al., 2019), and behave in line with prospect theory (Abdellaoui et al., 2013). We contribute to the literature by showing that decision-making quality of finance professionals in portfolio decisions is not superior compared to our general population sample who selected themselves into the experiment.

2. Experimental Design

Recruitment and data collection. We conducted an online experiment in Sweden in cooperation with *Statistiska centralbyrån* (SCB; Statistics Sweden), who invited subjects for the experiment and provided a set of predefined variables of the subjects’ register data for those who completed the experiment. SCB sent out invitations (including a hyperlink to the online experiment and a personalized alphanumeric identifier serving as login credentials) to 8,215 finance professionals and a randomly selected representative sample

of 8,215 subjects from Sweden's working general population, excluding finance professionals. The sample of finance professionals includes financial analysts and investment advisors, traders and fund managers, and financial brokers. For the general population, following Edin and Fredriksson (2000) and Böhm et al. (2018), we only include people with a declared labor income exceeding the minimum amount that qualifies for the earnings related part of the public pension system. Invitations were sent out in two waves. 20% of the sample were invited in the first week of 2019. Since no technical issues had arisen, the remaining 80% of the sample were invited in the third week of 2019.

Once participants logged in to the online software, programmed in *oTree* (Chen et al., 2016), using their personal identifier, they were presented with a detailed outline of the experiment. In particular, on the first screen, participants were informed that register data provided by SCB will be matched with the data collected in the experiment. Moreover, participants were informed that the study has been approved by the ethical review boards in Gothenburg and at *Statistiska centralbyrån*. Participants agreed upon the conditions and were directed to the instructions of the experiment. The data handling procedures ensured full pseudonymity of all participants. Further details and additional information on the recruitment, data collection, and experimental implementation are provided in Appendix A.

In total, 408 finance professionals and 550 people from the general population completed the experiment. The experiment was conducted in Swedish and took on average 45 minutes to complete. The average payment to participants was 238.9 Swedish Krona (SEK; $SD = 122.3$), which is approximately \$30 given the exchange rate at the beginning of 2019. The experimental data was collected between January 4 and February 10, 2019.

Figure 1: Flow chart of the experiment. This figure illustrates the sequence of tasks for subjects in our experiment. First, participants were randomly assigned to one treatment and completed 25 investment decisions. Then, subjects from the general population could delegate their investment decision to an agent in a delegation decision stage. Finally, all subjects completed several side tasks, including self-reported items on economic preferences and supplementary survey questions, a financial literacy test, and a numeracy inventory.

The sequence of tasks within the experiment is graphically summarized in Figure 1 and described below. For detailed information on the main task, please refer to Appendix B. Details on the side tasks and questionnaires are provided in Appendix D. Analyses on subjects’ decision times across subject pools, treatments, and tasks are summarized in Appendix F, outlining high data quality due to moderate variance across all sub-samples.

Register data. In addition to the data collected in the online experiment, we obtained register data from SCB for each participant who completed the experiment. In particular, we received data on demographics (e.g., age, gender, income), occupational history (e.g., workplace, firm size), subjects’ education, their wealth history, and military records (e.g., scores of the military suitability tests). See Appendix A for further details on these variables. In the analysis of experimental results, we only use part of the registry data as control variables, in particular, participants’ gender (binary indicator for female), age (in years), net income from major employment in 2017 (in thousand SEK’s), and maximum education level (dichotomous indicators for high school education or less, university education smaller or equal to three years, and university education larger than three years).

Experimental treatments. Depending on the subject pool, participants were randomly assigned to one of the treatments listed in Table 1. Common to all treatments, both for finance professionals and for the general population sample, is the 25-item allocation decision task, which is described in detail below.

After having completed all items of the allocation decision task, participants from the general population (principals) had the opportunity to delegate their decisions to an agent. If principals opted for delegating their decisions, the experimental payoff depended on the agent’s rather than their own decisions.⁴

Depending on the treatment, the principals’ delegation was either to an investment algorithm programmed by the experimenters (*GP-ALGO*), a finance professional with aligned, i.e., linear, incentives (*GP-ALIGNED*), or a finance professional receiving a flat payment of 200 SEK for deciding on behalf of one or more clients (*GP-FIXED*). Note that, compared to the baseline condition *GP-FIXED*, treatment *GP-ALIGNED* modifies the incentive structure of the agent, while holding the type of agent constant. Treatment *GP-ALGO* modifies the type of agent from a human to an investment algorithm.

⁴ Note that we designed the experiment in a way that each participant made the investment decision first, but was informed about the opportunity to delegate the investment decisions only afterwards. Thus, principals do not actually delegate their decisions, but rather decide whether their own or the agent’s decisions are relevant for their payment. While in real-world applications, people usually do not make investment decisions prior to choosing whether or not to delegate, there are practical reasons for this design choice: First, the opportunity to delegate without prior decisions potentially leads to a high number of delegation in order to receive an experimental payment without spending any effort. Such considerations, however, are not in the focus of this project. Second, our design allows examining the allocation decisions of participants who chose to delegate. This way we can study whether or not delegation pays off for those who delegate as well as those who stick to their own decisions. Moreover, we can study risk communication between principals and agents by comparing clients’ and professionals’ investment decisions conditional on risk levels. However, we cannot account for other potential motives of delegation decisions such as principals’ unwillingness to get informed in financial matters. Thus, potential “clients” that do not want to engage in financial matters at all might have dropped out initially. For a comprehensive response rate analysis and a discussion of potential self-selection effects, please refer to Appendix E.

Table 1: Treatment overview. This table illustrates the randomly assigned between-subjects treatments for both samples, finance professionals and participants from the general population. The sample sizes per condition are indicated in Figure 1.

Finance professionals <i>... make decisions ...</i>		General population <i>... can delegate decisions to ...</i>	
<i>FP-OWN</i>	on one's own account	<i>GP-ALGO</i>	investment algorithm
<i>FP-ALIGNED</i>	for third party (linear incentives)	<i>GP-ALIGNED</i>	finance professional (linear incentives)
<i>FP-FIXED</i>	for third party (flat payment)	<i>GP-FIXED</i>	finance professional (flat payment)

If principals chose to delegate, they were asked to specify the risk (on a scale from 1 [no risk] to 4 [maximum return]) they wanted to be taken on their behalf by the agent,⁵ as well as their (maximum) willingness to pay for delegating the investment decisions (between 0 and 50 SEK, in steps of 5 SEK). At the end of the experiment, a “price” for delegating the decision to the agent (between 0 and 50 SEK) was randomly determined: If a participant’s willingness to pay was higher than this random number, his/her decisions were delegated to the agent at the randomly determined price (i.e., the agent’s decisions were payoff relevant for the principal); if not, no delegation took place and the principal’s decisions were relevant for the payment in the experiment.

Finance professionals were randomly assigned to one of three treatments in which they either made decisions on their own account (*FP-OWN*), or on behalf of subjects from the general population sample. When deciding on principals’ account, finance professionals either faced aligned incentives (i.e., they received exactly the same monetary payoff as their client; *FP-ALIGNED*), or were paid a flat fee of 200 SEK (*FP-FIXED*). Moreover, when deciding on behalf of others, finance professionals were asked to comply with a randomly assigned risk level (between 1 [no risk] and 4 [maximum return]). In case a participant from the general population delegated his/her decisions, he/she was matched with a participant from the finance professional sample (whose decisions has then been payoff relevant for the principal), based on the particular treatment *and* the stated risk level. All details about the delegation decision itself, the risk levels as a means to communicate the desired riskiness of the allocation decisions, the matching modalities, as well as the payment procedures were common knowledge.

⁵ The investment algorithm was programmed to construct investment portfolios, given the particular risk level, as follows: In each investment decision, the minimum variance portfolio and the maximum return portfolio were mapped to the endpoint options of the risk level scale, i.e., 1 and 4, respectively. Thus, risk level 1 was always associated with a sure payoff, whereas risk level 4 always involved a 100% investment in the asset with the highest expected return. For risk levels 2 and 3, portfolio weights were determined in equally sized steps between these fixed endpoints. For instance, if payoffs were 2.40 SEK / 0.00 SEK for asset A and 0.00 SEK / 0.80 SEK for asset B, then the risk-free portfolio was characterized by an investment of 25% in A and 75% in B, whereas the maximum return portfolio corresponded to an investment of 100% in A. Risk levels 2 and 3 were associated with portfolios investing 50% and 75% in A, respectively. To the participants the algorithm was described to be “programmed in such a way that it maximizes your expected profit conditional on the risk level you indicate below”.

Allocation decision task. The workhorse of our experiment is the allocation decision task as used by Banks et al. (2018). The task consists of 10 decisions with two binary assets in a first block, and 15 decisions with five binary assets in a second block. Participants were first presented with the task instructions for the first block. After reading the instructions, participants could only continue once they had correctly answered three comprehension questions. After the first ten decisions, participants were informed that five rather than two assets would be available for the remaining 15 decisions. Again, after the instructions, participants had to correctly answer three comprehension questions before proceeding with the task. The order of the two blocks was fixed for all subjects, but the order of decisions was randomized in each of the two blocks.

For each of the 25 items, participants were informed about the assets' return per 1 SEK invested depending on the outcome of a coin toss. The returns for each asset in the 25 investment decisions are depicted in Table B1, and the corresponding opportunity sets are illustrated in Figure B1 in Appendix B. Participants were endowed with 100 SEK per item and had to allocate the entire endowment on the available assets. At the end of the experiment, one of their own or—in case a client opts for delegating the decisions—one of the agent's decisions was randomly chosen, and a simulated coin toss determined the participant's payoff. Returns were paid on top of the endowment, i.e., payments could not fall below 100 SEK.

Decision-making quality measures. The allocation decision task used in this experimental set-up allows quantifying decision-making quality based on four different measures. In particular, closely following Banks et al. (2018), in addition to the expected return (ER) and the standard deviation (SD) of the chosen portfolios we determine a decision-making quality index that comprises the following measures:

- For each decision, we determine violations of the first order stochastic dominance principle ($FOSD$; Hadar and Russell, 1969). In particular, we calculate the difference between the expected return of the chosen portfolio and the highest possible expected return of a portfolio that guarantees the same minimum payoff as the chosen one. On the participant level, the measures of $FOSD$ are averaged across all decisions, except for two opportunity sets for which expected returns of all assets were identical.
- To quantify violations of the General Axiom of Revealed Preferences ($GARP$), we utilize the Money Pump Index (MPI ; Echenique et al., 2011), i.e., the monetary amount a potential arbitrageur could make by exploiting a subject's violations in revealed preferences. On the participant level, we calculate the mean money pump cost over all pairwise combinations of opportunity sets.
- Participants' failure to minimize risk (FMR ; Banks et al., 2018) is calculated based on the decisions in the two opportunity sets for which returns of all allocations were identical, such that the risk-free portfolio (second-order) dominates all other feasible portfolios. A subject's measure of FMR is calculated as the mean standard deviation over the two opportunity sets.
- Participants' financial competence (FC ; Banks et al., 2018) is measured based on the portfolio choices in each of four opportunity sets that were identical in the two-assets- and the five-assets-frame and/or that were mirrored versions of another opportunity set. A participant's measure of FC is defined as the mean absolute difference in expected returns across all identical opportunity sets.

For each participant, the predicted values of a principal component analysis of the four measures, *FOSD*, *GARP*, *FMR*, and *FC*, constitute our decision-making quality index (*DMQI*). Detailed descriptions on how each of the decision-making measures is defined are provided in Appendix C.

Questionnaires. After the allocation decisions (but prior to the choice whether or not to delegate), all participants were asked to self-assess the overall level of risk taken across the 25 items of the allocation decision task on a scale from 1 to 4, i.e., on the same scale as when choosing the risk level in delegating the risky decisions. In addition, we included the following set of non-incentivized survey items at the end of the experiment: All participants were asked about (i) their self-assessed risk attitude in general and in financial decisions (Dohmen et al., 2011; Falk et al., 2016), (ii) their willingness to abstain from something today for a future benefit (Falk et al., 2016), (iii) their trust in mankind in general, in persons from the finance industry, and in financial algorithms, (iv) their proneness to shift blame on others (Wilson et al., 1990), and (v) their level of prosociality in a hypothetical charitable giving setting (Falk et al., 2018). Furthermore, we included a 5-item questionnaire on delegation and advice-seeking in financial decisions, which was only posed to participants that indicate that they have been active in the financial market. Afterwards, all participants had four minutes to answer an 8-item Rasch-validated numeracy inventory (Weller et al., 2013), including two questions on cognitive reflection. In addition, participants had to provide their self-assessment of the number of correct answers in the numeracy questionnaire as well as of their ranking compared to a random sample of the Swedish population. These assessments allow us constructing two measures of overconfidence (overestimation and overplacement). Finally, participants had three minutes to answer a 6-item financial literacy questionnaire based on van Rooij et al., 2011. For further details regarding the survey items, please refer to Appendix D.

3. Results

In the following, we first answer research question 1 by examining principals' decisions to delegate across treatments and identifying potential drivers of the choice whether or not to delegate. In a second step, in order to address research question 2, we examine differences in decision-making quality between subject pools and treatments, serving as a basis for analyzing the effectiveness of delegation. Finally, we examine the communication of risk between principals and agents, providing us with answers to the third research question.

Descriptive results regarding the samples of finance professionals and the general population, the responses to the questionnaires, and the decision-making quality measures are presented in Tables E1 to Table E4 in Appendix E. Note that detailed descriptions on each of the decision-making measures used in this section are provided in Appendix C.

Throughout the presentation of the results, we indicate standardized effect sizes in terms of marginal effects at the means (*MEM*) for non-linear models, and in terms of (absolute) values of Cohen's d for linear models. d is approximated by $d \approx \beta / (SE \cdot \sqrt{n})$, with β denoting the respective regression coefficient and SE referring to the corresponding standard error.

Result 1 – Delegation Decisions. *Delegation rates among principals are highest to the investment algorithm, followed by professionals facing aligned incentives. Principals’ propensity to delegate their decisions increases with trust in the respective agent and their willingness to shift blame on others, but decreases with their own decision-making quality (DMQI). When delegating their decisions, principals, on average, request the agent to take more risk than they perceive to be taken in their own decision.*

Support: Panel (A) in Figure 2 reports estimates of logit regressions of the binary delegation choice on treatment indicators. We find that delegation rates increase from a base level of 16.9% in treatment *GP-FIXED* to 25.9% in treatment *GP-ALIGNED* ($MEM = 0.090$, $p = 0.034$, $n = 550$) and to 37.9% in treatment *GP-ALGO* ($MEM = 0.210$, $p < 0.001$, $n = 550$), respectively. This result suggests that principals’ delegation decisions depend on both whom they are delegating to (i.e., an algorithm or a finance professional) and what incentives the agent faces (fixed or aligned compensation). However, the principals’ willingness to pay for delegating their decisions to the agent does not significantly differ between the treatments (see Table G1 in Appendix G).

Panel (B) in Figure 2 illustrates the principals’ desired levels of risk conditional on the risk perception of their own investment decision when delegating their decisions to the agent. On average, principals tend to ask the agent to take higher levels of risk ($m = 2.84$, $sd = 0.69$) than they perceive they implemented themselves when deciding on their own behalf ($m = 2.58$, $sd = 0.76$; paired-sample t -test: $t(147) = 4.081$, $p < 0.001$, $n = 148$).

Figure 2: Delegation frequency and desired vs. perceived risk level of principals. Panel (A) depicts the share of principals opting for delegating their investment decisions to the agent conditional on the treatment. Error bars indicate standard errors of the mean (SEM); p -values are based on a logit regression of delegation on treatment indicators (see model (1) in Table 2). Panel (B) shows principals’ desired risk levels when delegating their investment decisions to the agent conditional on the perceived riskiness of their own decisions ($n = 148$).

In a second step, we investigate whether behavioral and cognitive measures systematically impact principals' decision whether to delegate their decisions to the agent. As indicated by the estimates reported in model (3) of Table 2, the odds of delegating one's decision (relative to the delegation rate of 16.9% in the *FP-FIXED* condition) are expected to decrease by 22.8% ($MEM = 0.045$, $p = 0.006$, $n = 550$) for a one standard deviation increase in principals' decision-making quality index (*DMQI*). Principals' trust in the agent turns out to have the largest effect on the delegation decision: a one standard deviation increase in (self-reported) trust, on average, implies an increase in the odds of delegating one's decision to the agent by 98.8% ($MEM = 0.115$, $p < 0.001$, $n = 550$). Similarly, the likelihood for delegating one's investment decisions tends to increase with a higher propensity for shifting blame to others ($MEM = 0.043$, $p = 0.016$, $n = 550$). Notably, neither numeracy skills and financial literacy scores, nor our measures of overconfidence, nor participants' (self-reported) risk tolerance show any explanatory power with respect to principals' delegation decisions. All results are robust to the inclusion of control variables (see models (2) and (4) in Table 2).⁶ In addition, we report that none of the drivers of delegation decisions show any significant impact on participants' willingness to pay for delegating their choices to the agent (see Table G1 in Appendix G).

Result 2 – Decision-Making Quality. *Finance professionals deciding on their own account show higher decision-making quality compared to subjects from the general population only for moderate levels of risk tolerance and above. Moreover, professionals' decision-making quality does not significantly differ when deciding on behalf of clients, neither when being paid a flat fee, nor when facing aligned incentives. On average, delegating the decisions does not pay off for principals. While finance professionals do indeed yield slightly higher returns than the general population (conditional on the risk level), their portfolios also imply higher portfolio risk.*

Support: Given our experimental setup, clients can—assuming that individual-level risk preferences are perfectly mapped by the agents' decision—only benefit from delegating their investment decisions, if the agents show superior decision-making quality. Thus, in a first step, we compare the decision-making quality index (*DMQI*) between finance professionals deciding on their own behalf and the general population. A two-sample *t*-test suggests that finance professionals are indeed less prone to poor decisions ($d = 0.281$, $t(685) = 2.942$, $p = 0.003$, $n = 687$; see model (1) in Table G2 in Appendix G). However, by design, errors in decision-making are less likely, if the decision-maker is risk tolerant.⁷ Indeed, finance professionals deciding on their own account are significantly less risk averse, in terms of both the mean portfolio risk (*SD*) taken in the 25 investment decisions (two-sample *t*-test; $d = 0.506$, $t(685) = 5.302$, $p < 0.001$, $n = 687$) as well as self-reported risk attitudes in financial matters (two-sample *t*-test; $d = 0.818$, $t(685) = 8.570$, $p < 0.001$, $n = 687$). Figure 3 shows the linear prediction of decision-making quality based on ordinary least squares regressions of *DMQI* on a subject pool indicator, the portfolio risk (*SD*), and the interaction of

⁶ For descriptive results on self-rated trust levels, self-reported risk tolerance, numeracy skills, financial literacy, overestimation, and overplacement, please refer to Figures E2–E4 in Appendix G.

⁷ For instance, consider a risk neutral decision-maker: Choosing an allocation in the task is straightforward as he/she will simply invest the entire endowment in the asset yielding the highest expected return. On the other hand, consider a highly risk averse decision-maker: In order to hedge risks, the decision-maker has to choose well-balanced portfolios. Apparently, the likelihood of violating the principle of first order stochastic dominance (*FOSD*) and/or the generalized axiom of revealed preferences (*GARP*) is considerably larger for allocations in the interior of the opportunity sets, compared to boundary allocations.

Table 2: Determinants of delegation decisions. This table reports marginal effects estimates from logit regressions of the binary choice whether to delegate the investment decisions to the agent on treatment indicators, a set of experimental measures, and self-reported measures. Robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.005$.

	(1)	(2)	(3)	(4)
<i>Treatment Indicators:</i>				
<i>GP-ALIGNED</i>	0.090*	0.084*	0.062	0.053
	(0.043)	(0.043)	(0.042)	(0.042)
<i>GP-ALGO</i>	0.210**	0.209**	0.196**	0.190**
	(0.045)	(0.045)	(0.043)	(0.043)
<i>Experimental Measures:</i>				
Decision Making Quality Index			−0.044*	−0.041*
			(0.016)	(0.015)
Financial Literacy Score (Std.)			−0.006	0.001
			(0.033)	(0.033)
Numeracy Score (Std.)			−0.040	−0.047
			(0.043)	(0.044)
Overestimation (Std.)			−0.032	−0.033
			(0.021)	(0.021)
Overplacement (Std.)			−0.010	−0.017
			(0.025)	(0.025)
<i>Self-Reported Measures:</i>				
Risk Tolerance (Std.)			0.008	0.007
			(0.019)	(0.019)
Blame Shifting (Std.)			0.041*	0.044*
			(0.018)	(0.019)
Trust in Agent (Std.)			0.115**	0.122**
			(0.018)	(0.018)
<i>Controls</i>				
	<i>no</i>	<i>yes</i>	<i>no</i>	<i>yes</i>
Wald χ^2	19.788	25.863	59.687	65.216
$p > \chi^2$	0.000	0.001	0.000	0.000
Pseudo R^2	0.032	0.041	0.127	0.143
Observations	550	550	550	550

Notes: Treatment *GP-FIXED* serves as reference condition. All self-reported measures are standardized scores. “Trust in Agent” refers to a combined variable of trust in finance professionals and financial algorithms, conditional of the treatment. “Blame Shifting” refers to the mean of two standardized survey items on shifting blame on others and resisting the temptation to shift blame on others. “Controls” include gender (binary indicator for female), age (in years), net income from major employment in 2017 (in thousand SEK’s), and maximum education level (dichotomous indicators for high school education or less, university education smaller or equal to three years, and university education larger than three years).

the dummy variable and *SD*, separated for treatments (see model (3) in Tables G2 and G3 in Appendix G, respectively). As illustrated in panel (A) of Figure 3, controlling for risk attitudes when comparing the *DMQI* between the pools reveals that finance professionals significantly outperform the general population sample in terms of decision-making quality only for moderate to high levels of portfolio risk.

In a second step, we examine if finance professionals' decision-making quality is systematically impacted by whether they decide on behalf of clients or on their own account, and whether the incentive scheme affects the proneness to errors in decision-making. A plain comparison of *DMQI* shows that—compared to deciding for clients and receiving a flat payment (*FP-FIXED*)—finance professionals tend to perform better when facing aligned incentives (*FP-ALIGNED*; $d = 0.100$, $t(405) = 2.019$, $p = 0.044$, $n = 408$), and when deciding on their own account (*FP-OWN*; $d = 0.132$, $t(405) = 2.656$, $p = 0.008$, $n = 408$); see model (1) in Table G3 in Appendix G). The difference in *DMQI* between the treatments *FP-ALIGNED* and *FP-OWN* is insignificant (Wald test; $F(1, 405) = 0.426$, $p = 0.514$, $n = 408$). However, as illustrated in panels (B) and (C) in Figure 3, the rather small differences in decision-making quality between finance professionals deciding on clients' behalf and the general population sample turn out being insignificant once we again control for varying levels of portfolio risk (see model (2) in Table G3 in Appendix G for details). Finally, panel (D) in Figure 3 depicts the linear predictions of *DMQI* conditional on mean portfolio risk for the three treatments *FP-OWN*, *FP-ALIGNED*, and *FP-FIXED*, emphasizing that finance professionals' decision-making quality is neither systematically affected by whether decisions are made on one's own behalf or on clients' accounts, nor by the compensation scheme the decision-maker faces (see model (3) in Table G3 in Appendix G for details).

As indicated by Result 1, clients take into consideration the agents' incentive structure when deciding whether or not to delegate their investment decisions. The latter finding, however, suggests that clients' expectations regarding the agents' performance—as reflected in the difference between delegation rates in treatment *GP-FIXED* and *GP-ALIGNED*—are not justified by the performance data, as professionals do not perform systematically better when facing aligned incentives.⁸

Table 3 summarizes the average decision-making quality (*DMQI*), expected portfolio returns, and portfolio risk, separated for finance professionals deciding on behalf of clients (*FP-FIXED* and *FP-ALIGNED*) and those principals who choose to delegate their decisions, conditional on the risk level principals and agents are matched on, as well as two-sample *t*-tests for each risk level. As already indicated by Result 2, finance professionals do not significantly outperform laypeople in terms of decision-making quality for any of the four risk levels. Comparing the mean expected returns and portfolio risk associated with the allocation decisions suggests that, conditional on the risk level, finance professionals tend to generate weakly (and mainly insignificantly) higher returns, but at the cost of higher portfolio risk. Thus, overall, principals' delegation decisions to professionals (agents) do not result in more efficient portfolio allocations in terms of risk-adjusted returns, not even before potential costs of delegation.

Even though delegation has a very modest effect on decision quality, it could still be effective in terms of changing the risk profile of the investor. Result 1 reveals that, on average, principals ask the agents to

⁸ In Appendix F, we report detailed analyses on decision times across subject pools, treatments, and tasks. We observe, for instance, that professionals take more time when deciding for clients compared to when deciding on their own behalf, and compared to clients' own decisions in the two-asset opportunity sets (see Table F2 for details). However, in the five-asset cases, differences in time spent between clients and professionals vanish, suggesting that clients took the more complex tasks seriously. Together with Result 2, these findings might indicate that professionals really strive for meeting clients' expectations (i.e., desired risk levels), even though it does not translate into better performance in our sample. Furthermore, we examine potential learning effects in the investment task and the impact of the time spent per investment decision on decision-making quality. While we identify a significant decrease in the average time spent per investment decision for subsequent decisions in both subject pools, we report that the time spent per decision does not significantly impact decision-making quality, neither among the general population, nor the finance professionals sample; please refer to Appendix F for details.

Figure 3: Decision-making quality (DMQI) conditional on portfolio risk. This table shows linear predictions of the decision-making quality index (DMQI) conditional on the mean standard deviation (SD) across the 25 items in the allocation decision task (normalized to 1) after ordinary least squares regressions (see Tables G2 and G3 in Appendix G). **Panels (A)–(C).** Predictions of DMQI separated for the general population sample (pooled across treatments) and finance professionals deciding on their own behalf (FP-OWN), finance professionals deciding on behalf of clients receiving a flat payment (FP-FIXED), and finance professionals deciding on behalf of clients facing aligned incentives (FP-ALIGNED), respectively. **Panel (D)** Predictions of finance professionals’ DMQI separated for the three treatments FP-FIXED, FP-ALIGNED, and FP-OWN.

take more risk than they perceive they took on their own. For those subjects requesting a higher level of risk, delegation, on average, significantly increases portfolio risk—conditional on risk levels (GP-FIXED: $d = 0.277$, $t(141) = 3.299$, $p = 0.001$, $n = 3,550$; GP-ALIGNED: $d = 0.268$, $t(152) = 3.299$, $p = 0.001$, $n = 3,825$; GP-ALGO: $d = 0.539$, $t(29) = 2.953$, $p = 0.006$, $n = 750$; for comparisons on the subject level, please refer to Figure G1 in Appendix G).⁹

⁹ For the analysis, we start with only considering those clients who delegate their decision *and* request the agent to take more risk than they perceive they took by themselves ($n_{GP-FIXED} = 10$, $n_{GP-ALIGNED} = 14$, $n_{GP-ALGO} = 26$). In a second step, we identify all agents who serve as potential “matching partner” for the clients, i.e., those finance professionals in the corresponding treatment

Table 3: Mean decision-making quality, mean expected return, and mean portfolio risk by risk level.

This table summarizes the means of the decision-making quality index (*DMQI*), expected returns, and portfolio risk of principals who choose to delegate their decisions (excluding the treatment *GP-ALGO*) and finance professionals deciding on behalf of clients, conditional on the risk levels indicated by clients and risk levels agents' are asked to comply with. n_1 and n_2 refer to the row-wise sample size of the general population and of finance professionals, respectively. Standard deviations are reported in parentheses. t -statistics and standard errors (se) are based on two-sample t -tests with $n_1 + n_2 - 2$ degrees of freedom. * $p < 0.05$, ** $p < 0.005$.

		<i>Gen. Pop.</i>		<i>Fin. Prof.</i>		<i>t-Test</i>	
		<i>m</i>	<i>sd</i>	<i>m</i>	<i>sd</i>	<i>t</i>	<i>se</i>
<i>Decision Making Quality (DMQI):</i>							
<i>RL-1</i>	($n_1 = 1, n_2 = 77$)	-0.13	.	0.08	(1.06)	.	.
<i>RL-2</i>	($n_1 = 27, n_2 = 60$)	-0.13	(0.89)	-0.21	(1.28)	0.285	[0.273]
<i>RL-3</i>	($n_1 = 43, n_2 = 68$)	-0.43	(2.45)	0.15	(0.67)	-1.861	[0.314]
<i>RL-4</i>	($n_1 = 8, n_2 = 66$)	-0.36	(1.15)	0.05	(1.76)	-0.643	[0.639]
<i>Mean Expected Return (ER):</i>							
<i>RL-1</i>	($n_1 = 1, n_2 = 77$)	141.97	.	140.61	(14.50)	.	.
<i>RL-2</i>	($n_1 = 27, n_2 = 60$)	145.25	(10.38)	147.93	(15.50)	-0.816	[3.275]
<i>RL-3</i>	($n_1 = 43, n_2 = 68$)	151.95	(10.77)	155.56	(11.77)	-1.625	[2.220]
<i>RL-4</i>	($n_1 = 8, n_2 = 66$)	155.42	(8.42)	174.37	(13.31)	-3.919**	[4.834]
<i>Mean Portfolio Risk (SD):</i>							
<i>RL-1</i>	($n_1 = 1, n_2 = 77$)	46.01	.	42.10	(36.85)	.	.
<i>RL-2</i>	($n_1 = 27, n_2 = 60$)	53.73	(21.06)	61.16	(36.92)	-0.975	[7.623]
<i>RL-3</i>	($n_1 = 43, n_2 = 68$)	72.65	(23.84)	80.67	(30.07)	-1.477	[5.423]
<i>RL-4</i>	($n_1 = 8, n_2 = 66$)	79.76	(16.46)	136.91	(41.23)	-3.864**	[14.792]

Our finding that delegating investment decisions to finance professionals does not pay off in terms of decision quality, has to be seen in light of a potential sample selection bias in our study. Some of the differences in socio-demographic characteristics between the subjects in the experiment and those that did not take part are significant (see Table E1 in Appendix E). For both subject pools, the number of participants with a University degree is clearly higher in our sample than among those that did not participate. The high level of education in our sample might explain the high levels of decision-making quality among the general population sample (see Appendix C for details). However, this biased sample selection could also be relevant in real-world markets where private investors are usually a biased sample from the general population too, e.g., with males, older and wealthier people with higher financial sophistication being over-represented (Collins, 2012; Hackethal et al., 2012; Calcagno and Monticone, 2015). Given both effects being active simultaneously, we are cautious in generalizing our findings of Result 2.

(fixed vs. aligned incentives) or the investment algorithm deciding for clients with the risk level eventually indicated by clients when delegating their decisions. For each treatment pair (e.g., *GP-FIXED*/*FP-FIXED*), the test statistics reported refer to coefficient tests of the indicator variable "agent" in ordinary least squares regressions of the portfolio risk in the 25 investment decisions on the indicator variable, controlling for delegation risk (dummy variables) with standard errors being clustered on the individual level.

Result 3 – Communication of Risk. *Professionals show difficulties in constructing portfolios conditional on the risk levels clients indicate. In particular, portfolio risk is widely dispersed for all risk levels, suggesting that principals indicating a low or high risk level when delegating their decisions might end up with very similar portfolio risks. This challenge in communicating risk between parties may be exaggerated by significant differences in the perception of risk between principals and agents.*

Support: As illustrated in panel (A) of Figure 4, the distributions of mean portfolio risk (in terms of average standard deviation) vastly overlaps for the different risk levels (see also Figure G2 in Appendix G). While the mean portfolio risk increases significantly with the risk level ($p < 0.001$ for all pairwise comparisons; see Table G4 in Appendix G for details), eyeballing the distributions reveals that almost the full range of risk might be associated with each level. This result implies that clients indicating different levels of risk when delegating their investment decisions can eventually end up with similar levels of portfolio risk. For example, 25% of the portfolios constructed by finance professionals for principals indicating risk level 2 / 4 exhibit more risk than 50% of the allocations designed for risk level 3 / 4; even more problematic, about 25% of the portfolios designated for risk level 1 / 4 implies higher risk than 25% of the allocations constructed for risk level 3 / 4.

Comparing finance professionals' risk perception of the portfolios constructed on behalf of clients ($m = 2.54$, $sd = 0.97$) with the risk levels they are asked to comply with ($m = 2.45$, $sd = 1.14$) in panel (B) of Figure 4 shows that, on average, finance professionals strive to follow the intentions of potential clients.

Figure 4: Portfolio risk conditional on risk level and clients' desired risk vs. principals' perceived risk. **Panel (A)** shows finance professionals' portfolio risk when deciding on behalf of clients conditional on the risk levels they are asked to comply with. In particular, the figure illustrates the distribution (Gaussian kernel) of the average standard deviation of the 25 allocations in the investment task (normalized to 1), conditional on the risk level (RL-1 ... RL-4) indicated by principals, on whose account finance professionals are asked to decide in treatments *FP-FIXED* and *FP-ALIGNED*. **Panel (B)** shows the risk level agents are asked to comply with when deciding on behalf of clients vs. agents' perception of the riskiness of their actual decisions ($n = 271$).

Figure 5: Portfolio risk conditional on risk perception. This figure shows cumulative distributions and boxplots of the mean standard deviation of the 25 allocations in the investment task (normalized to 1), conditional on the perceived riskiness of their choices (*RL-1* ... *RL-4*), separated for the general population sample (pooled across all treatments) and the sample of finance professionals deciding on behalf of clients (i.e., *FP-FIXED* and *FP-ALIGNED*). *p*-values reported for sample comparisons are based on two-sample Kolmogorov-Smirnov tests.

The challenge agents face when deciding on behalf of principals is further intensified by significant differences in risk perception. As illustrated in Figure 5, the distributions of portfolio risk associated with risk levels differ significantly between the general population and the finance professional sample.¹⁰ For risk levels 1 and 2, portfolio risk of subjects from the general population, on average, exceeds the risk of portfolios implemented by finance professionals deciding on behalf of clients. However, this effect reverses for risk levels 3 and 4. This result indicates that clients compose more similar portfolios across the risk spectrum than professionals do.

4. Conclusion

In this paper, we report a controlled lab-in-the field (online) experiment with finance professionals (serving as money managers) and subjects from the general population in Sweden (serving as clients). We examined (i) the motivations and characteristics of clients to delegate investment decisions, (ii) differences in decision-making quality between professionals and clients, and (iii) professionals' ability to construct portfolios that suit the risk attitudes of clients well.

¹⁰ Figure 5 depicts comparisons of the general population sample, pooled across all treatments, and the sample of finance professionals deciding on behalf of clients (*FP-FIXED* and *FP-ALIGNED*). Comparing the portfolio risk (*SD*) of the general population to the portfolio risk of finance professionals deciding on their own account (*FP-OWN*) for each of the four perceived risk levels reveals that the association of actual portfolio risk with the risk spectrum only significantly differs for risk level 4 (Kolmogorov-Smirnov tests; *RL-1*: $D = 0.354$, $p = 0.180$, $n = 60$; *RL-2*: $D = 0.195$, $p = 0.108$, $n = 311$; *RL-3*: $D = 0.198$, $p = 0.100$, $n = 232$; *RL-4*: $D = 0.379$, $p = 0.007$, $n = 84$). However, comparing the portfolio risk associated with the four levels between finance professionals deciding on their own (*FP-OWN*) and those investing on behalf of clients (*FP-FIXED* and *FP-ALIGNED*) does not reveal any significant differences (Kolmogorov-Smirnov tests; *RL-1*: $D = 0.167$, $p = 0.958$, $n = 54$; *RL-2*: $D = 0.157$, $p = 0.450$, $n = 136$; *RL-3*: $D = 0.131$, $p = 0.663$, $n = 134$; *RL-4*: $D = 0.185$, $p = 0.506$, $n = 84$).

First, we found that investors delegated to the investment algorithm significantly more often than to professionals with aligned incentives and flat incentives. Furthermore, we reported that those investors with the highest levels of trust in professionals (investment algorithms) and those who are prone to shift blame on others delegated the most, whereas we found that principals' own decision-making quality was negatively related to the delegation frequency. Second, we found that overall decision-making quality of professionals was not significantly better than that of subjects from the general population. Finally, we observed that professionals had difficulties in constructing portfolios conditional on the risk-levels clients indicated. In particular, we found strong overlaps in portfolio risks especially among three of the four risk classes.

Our study has implications for real-world delegation decisions: first, clients with low decision-making quality and/or high level of trust in professionals/algorithms are indeed the ones that delegate more frequently. This result highlights the importance of establishing trust in the finance industry in general and in money managers in particular, as it appears to be one of the major motives for delegation decisions. However, professionals' decision-making quality is only marginally and not significantly better than that of clients. We still conclude that for the clients that trust professionals or financial algorithms, delegation is probably a good choice, as they "purchase" slightly better decisions and are confident that professionals do it well. In other words, "money doctors" (Gennaioli et al., 2015) are trusted when investing money of their clients, even when the outcome is not significantly better.

Second, our results indicate that some clients use delegation as way of increasing the risk of their portfolio, but the feasibility of this objective is hampered by our finding that professionals face troubles in correctly implementing clients' expected portfolio risk-level. The issue of risk communication is particularly relevant for real-world delegation of financial decisions and related to the empirical studies of Foerster et al. (2017) and Linnainmaa et al. (2019). Both studies show that financial advisers typically invest personally just as they advise their clients. Thus, we conclude that a better match of advisers and clients in terms of risk preferences and potentially also with respect to risk perception (Holzmeister et al., 2019) might be beneficial both for clients and financial institutions.

References

- Abdellaoui, M., Baillon, A., Placido, L., & Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review*, 101(2), 695–723.
- Abdellaoui, M., Bleichrodt, H., & Kammoun, H. (2013). Do financial professionals behave according to prospect theory? An experimental study. *Theory and Decision*, 74(3), 411–429.
- Alevy, J. E., Haigh, M. S., & List, J. A. (2007). Information cascades: Evidence from a field experiment with financial market professionals. *Journal of Finance*, 62(1), 151–180.
- Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2016). Deciding for others reduces loss aversion. *Management Science*, 62(1), 29–36.
- Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2019). Risking other people's money: Experimental evidence on bonus schemes, competition, and altruism. *Scandinavian Journal of Economics*, online first, 1–27.
- Banks, J., Carvalho, L., & Perez-Arce, F. (2018). Education, decision-making, and economic rationality. *Review of Economics and Statistics*, 101(3), 428–441.
- Bartling, B., & Fischbacher, U. (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies*, 79(1), 67–87.
- Bebchuk, L., & Spamann, H. (2010). Regulating bankers' pay. *Georgetown Law Journal*, 98(2), 247–287.
- Böhm, M., Metzger, D., & Strömberg, P. (2018). Since you are so rich, you must be really smart: Talent and the finance wage premium. *Riksbank Research Paper Series No. 137*.
- Bolton, G., & Ockenfels, A. (2010). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment. *American Economic Review*, 100(1), 628–633.
- Calcagno, R., & Monticone, C. (2015). Financial literacy and the demand for financial advice. *Journal of Banking and Finance*, 50, 363–380.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, 52(1), 57–82.
- Chakravarty, S., Harrison, G., Haruvy, E., & Rutström, E. (2011). Are you risk averse over other peoples' money? *Southern Economic Journal*, 77(4), 901–913.
- Chang, T., Solomon, D., & Westerfield, M. (2016). Looking for someone to blame: Delegation, cognitive dissonance, and the disposition effect. *The Journal of Finance*, 71(1), 267–302.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- Cipriani, M., & Guarino, A. (2009). Herd behavior in financial markets: An experiment with financial market professionals. *Journal of the European Economic Association*, 7(1), 206–233.
- Collins, J. M. (2012). Financial advice: A substitute for financial literacy? *Financial Services Review*, 21(4), 307–322.
- D'Acunto, F., Prabhala, N., & Rossi, A. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5), 1983–2020.
- Dewatripont, M., & Freixas, X. (2012). Bank resolution: Lessons from the crisis. In M. Dewatripont & X. Freixas (Eds.), *The crisis aftermath: New regulatory paradigms*. London: Centre for Economic Policy Research.

- Diamond, D. W., & Rajan, R. G. (2009). The credit crisis: Conjectures about causes and remedies. *American Economic Review*, 99(2), 606–610.
- Dietvorst, B., Simmons, J., & Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, 9(3), 522–550.
- Dulleck, U., & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, 44(1), 5–42.
- Echenique, F., Lee, S., & Shum, M. (2011). The Money Pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6), 1201–1223.
- Edin, P., & Fredriksson, P. (2000). LINDA: Longitudinell individual data for Sweden. *Working Paper*.
- Eriksen, K., & Kvaløy, O. (2010). Myopic investment management. *Review of Finance*, 14(3), 521–542.
- Eriksen, K., Kvaløy, O., & Luzuriaga, M. (2017). Risk-taking on behalf of others. *CESifo Working Paper Series No. 6378*.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, 133(4), 1645–1692.
- Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2016). The Preference Survey Module: A validated instrument for measuring risk, time, and social preferences. *IZA Discussion Paper Series No. 9674*.
- Financial Crisis Inquiry Commission. (2011). *The Financial Crisis Inquiry Report: Final report of the National Commission on the causes of the financial and economic crisis in the United States*. Washington, DC: U.S. Government Printing Office.
- Foerster, S., Linnaimaa, J. T., Melzer, B. T., & Previtero, A. (2017). Retail financial advice: Does one size fit all? *Journal of Finance*, 72(4), 1441–1482.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- French, K. R. (2008). Presidential Address: The cost of active investing. *The Journal of Finance*, 63(4), 1537–1573.
- Füllbrunn, S., & Luhan, W. (2015). Am I my peer's keeper? Social responsibility in financial decision making. *Ruhr Economic Papers #551*.
- Gennaioli, N., Shleifer, A., & Vishny, R. (2015). Money doctors. *Journal of Finance*, 70(1), 91–114.
- Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance*, 51(3), 783–810.
- Guiso, L., Sapienza, P., & Zingales, L. (2004). The role of social capital in financial development. *American Economic Review*, 94(3), 526–556.
- Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance*, 63(6), 2557–2600.

- Hackethal, A., Haliassos, M., & Jappelli, T. (2012). Financial advisors: A case of babysitters? *Journal of Banking and Finance*, 36(2), 509–524.
- Hadar, J., & Russell, W. (1969). Rules for ordering uncertain prospects. *American Economic Review*, 59(1), 25–34.
- Haigh, M. S., & List, J. A. (2005). Do professional traders exhibit myopic loss aversion? An experimental analysis. *Journal of Finance*, 60(1), 523–534.
- Harvey, C., Rattray, S., Sinclair, A., & Van Hemert, O. (2017). Man vs. machine: Comparing discretionary and systematic hedge fund performance. *Journal of Portfolio Management*, 43(4), 55–69.
- Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644–1655.
- Holzmeister, F., Huber, J., Kirchler, M., Lindner, F., Weitzel, U., & Zeisberger, S. (2019). What drives risk perception? A global survey with financial professionals and lay people. *Management Science*, forthcoming.
- Ifcher, J., & Zarghamee, H. (2019). Behavioral economic phenomena in decision-making for others. *Journal of Economic Psychology*, forthcoming.
- Inderst, R., & Ottaviani, M. (2012a). Competition through commissions and kickbacks. *American Economic Review*, 102(2), 780–809.
- Inderst, R., & Ottaviani, M. (2012b). Financial advice. *Journal of Economic Literature*, 50(2), 494–512.
- Inderst, R., & Ottaviani, M. (2012c). How (not) to pay for advice: A framework for consumer financial protection. *Journal of Financial Economics*, 105(2), 393–411.
- Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, 23(2), 389–416.
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.
- Kaustia, M., Alho, E., & Puttonen, V. (2008). How much does expertise reduce behavioral biases? The case of anchoring effects in stock return estimates. *Financial Management*, 37(3), 391–412.
- Kirchler, M., Lindner, F., & Weitzel, U. (2018a). Delegated decision making and social competition in the finance industry. *Working Papers in Economics and Statistics, University of Innsbruck*.
- Kirchler, M., Lindner, F., & Weitzel, U. (2018b). Rankings and risk-taking in the finance industry. *Journal of Finance*, 73(5), 2271–2302.
- Kling, L., König-Kersting, C., & Trautmann, S. T. (2019). Investment preferences and risk perception: Financial agents versus clients. *AWI Discussion Paper Series No. 674*.
- Lachance, M.-E., & Tang, N. (2012). Financial advice and trust. *Financial Services Review*, 21(3), 209–226.
- Linnainmaa, J. T., Melzer, B., & Previtero, A. (2019). The misguided beliefs of financial advisors. *The Journal of Finance*, forthcoming.
- Logg, J., Minson, J., & Moore, D. (2017). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103.
- Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629–650.

- Mullainathan, S., Noeth, M., & Schoar, A. (2012). The market for financial advice: An audit study. *NBER Working Paper No. 17929*.
- Rajan, R. G. (2006). Has finance made the world riskier? *European Financial Management*, 12(4), 499–533.
- Rud, O. A., Rabanal, J. P., & Horowitz, J. (2018). Does competition aggravate moral hazard? A multi-principal-agent experiment. *Journal of Financial Intermediation*, 33, 115–121.
- Schwaiger, R., Kirchler, M., Lindner, F., & Weitzel, U. (2019). Determinants of investor expectations and satisfaction: A study with financial professionals. *Journal of Economic Dynamics and Control*, *forthcoming*.
- Shefrin, H. (2007). *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Oxford: Oxford University Press.
- Sutter, M. (2009). Individual behavior and group membership: Comment. *American Economic Review*, 99(5), 2247–2257.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking and Reasoning*, 20(2), 147–168.
- van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101(2), 449–472.
- Vieider, F. M., Villegas-Palacio, C., Martinsson, P., & Mejía, M. (2016). Risk taking for oneself and others: A structural model approach. *Economic Inquiry*, 54(2), 879–894.
- Weitzel, U., Huber, C., Huber, J., Kirchler, M., Lindner, F., & Rose, J. (2019). Bubbles and financial professionals. *Review of Financial Studies*, *forthcoming*.
- Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, 26(2), 198–212.
- Wilson, G., Gray, J., & Barrett, P. (1990). A factor analysis of the Gray-Wilson personality questionnaire. *Personality and Individual Differences*, 11(10), 1037–1044.

Appendices

Delegated Decision-Making in Finance

Felix Holzmeister[†] Martin Holmén^{‡,*} Michael Kirchler^{†,‡}
Matthias Stefan[†] Erik Wengström^{§,¶}

[†] University of Innsbruck, Department of Banking and Finance

[‡] University of Gothenburg, Department of Economics, Centre for Finance

[§] Lund University, Department of Economics

[¶] Hanken School of Economics, Department of Finance and Economics

* Corresponding author: martin.holmen@gu.se

Contents

A	Data collection and Recruitment	1
B	Allocation Decision Task	3
C	Decision-Making Quality Measures	5
D	Questionnaires and Side Tasks	10
E	Descriptive Results.	15
F	Descriptives and Analyses of Time Spent	25
G	Supplementary Results	27

List of Tables

B1.	Parameters of the 25 opportunity sets.	3
C1.	Principal component analysis of the four decision-making quality measures.	9
D1.	Survey questions.	12
D2.	Numeracy inventory based on Weller et al. (2013).	13
D3.	Financial literacy inventory based on van Rooij et al. (2011).	14
E1.	Sample characteristics by subject pools.	17
E2.	Descriptive statistics of the decision-making quality measures.	18
E3.	Differences in decision-making quality measures.	18
E4.	Correlations between the decision-making quality measures.	19
E5.	Descriptive statistics and comparisons between pools for the survey items.	20
E6.	Descriptive statistics for numeracy, financial literacy, and overconfidence.	21
F1.	Descriptive statistics of time spent per task.	26
F2.	Differences in time spent.	26
G1.	Willingness to pay for delegation.	27
G2.	Decision-making quality by subject pool.	28
G3.	Finance professionals' decision-making quality by treatments.	29
G4.	Finance professionals' risk taking when deciding on behalf of clients.	30

List of Figures

B1.	Opportunity sets in the allocation decision task.	4
C1.	Violation of the principle of first order stochastic dominance (<i>FOSD</i>).	6
C2.	Violation of the generalized axiom of revealed preferences.	7
E1.	Self-reported risk tolerance.	22
E2.	Self-reported trust.	23
E3.	Numeracy and financial literacy scores.	24
E4.	Overestimation and overplacement.	24
G1.	Portfolio risk of clients' asking the agent to take more risk when delegating.	31
G2.	Number of portfolios with similar portfolio risk across risk levels.	32

A. Data collection and Recruitment

Experimental software. The experimental software—computerized in *oTree* (Chen et al., 2016)—which includes all instructions, treatment variations, as well as the Swedish/English translations has been pre-registered at <https://osf.io/ubpr3/>. Demo versions of the experiment and all treatments (in English) are available via <http://hea-2019-01-en.herokuapp.com>.

Recruitment. *Statistiska centralbyrån* (Statistics Sweden; *SCB*) sent out hard copy invitations to participate in the anonymous online experiment. The receivers of the invitations logged in to our experiment using a personalized participant code, which was linked to a key only known to *SCB*. The participant code indicated whether a particular subject was recruited from the finance professional pool or the general population pool. After the data collection has been completed, we sent the identifiers of those participants who completed the experiment to *SCB* and *SCB* used their keys to match the experimental data with the requested register data (which is described in detail below). Participants were informed that the data gathered in the experiment is matched with their register data in the invitation letters and on the first screen of the experiment.

Payments. To ensure full privacy of the data collected during the experiment, payouts were handled by the third party survey firm *Enkätfabriken*. Once participants completed the online experiment, they were redirected to a dedicated form on the website of *Enkätfabriken*. Participants used the same participant code as in the experiment. For payment purposes, *Enkätfabriken* collected participants’ names, email addresses, “personnummer” (personal identity number), and bank account details. The information collected was handled only by *Enkätfabriken* and has been used exclusively for sake of ordering the bank remittances.

Registry data. In addition to the data collected in the online experiment, we obtained the following register data from *Statistiska centralbyrån* (Statistics Sweden; *SCB*) for each participant who completed all tasks in the experiment:

- *Demographics*: year born, age, gender, county, municipality, and assembly of residence, marital status, year in marital status, family status, birth country, children living at home age 0–3, 4–6, 7–10, 11–15, 16–17, ≥ 18 , highest finished education level, education orientation, education group, education county, graduation year, primary source of income, work place municipality and county, work place industry 1990–1992, 1993–2001, 2002–2010, and 2007–2014, occupation 2002–2013 and 2014, net income of own business 1991–2003, 2003–2014, and 2004–2014, capital income, disposable income 1990–2004 and 2004–2014, disposable income of family 1990–2004 and 2004–2014, country of birth, date of immigration.
- *Firm / workplace*: number of employees at firm / workplace, number of men / women at firm / workplace, number of men / women with short / long education at firm / workplace, total salaries paid by firm / workplace.
- *Education*: high school, high school program, high school grades point average, high school graduation year, university, university program, university major, university graduation year.
- *Assets*: net wealth, total debt, bank account, listed equity, fixed income funds, other funds, bonds and other securities, taxable insurances, houses, apartments, holiday homes.

- *Military records*: command suitability, non cognitive abilities score, muscle strength, physical capacity for work, length, weight, cognitive scores 1 and 2 in language and logic, one in spatial understanding, and one in technical understanding.
- *Parents*: adoptive / biological mother / father, occupation mother / father, primary income source mother / father, net income from own business mother / father, net wealth mother / father.

In the present manuscript, we only use a small part of the available registry data as control variables in our analyses of observed behavior, in particular, participants' gender (binary indicator for female), age (in years), net income from major employment in 2017 (in thousand SEK's), and maximum education level (dichotomous indicators for high school education or less, university education smaller or equal to three years, and university education larger than three years).

After the experiment reported in this paper, participants were invited to a second, independent experiment for which the obtained registry data plays a more pivotal role. This is why we have obtained so many variables which are not used in this paper. Yet, given our research questions, the restricted use of the register data has been pre-registered at the outset (see <https://osf.io/ubpr3/> for the pre-registration). For details about the second experiment, please refer to the respective pre-registration at <https://osf.io/6rdp8/>.

B. Allocation Decision Task

For each of the 25 items, participants were asked to allocate an endowment of 100 SEK on either two or five assets. Participants were informed about the assets' returns per 1 SEK invested, depending on whether a coin toss shows up heads or tails. The returns used in the experiment were adopted from Banks et al. (2018), multiplied by a factor of 1.5, and rounded to one decimal place. Table B1 summarizes the parameters of the 25 sets; Figure B1 graphically depicts the 25 opportunity sets.

Table B1: Parameters of the 25 opportunity sets. This table indicates returns (in SEK) per 1 SEK invested for the different assets in the 25 opportunity sets, depending on whether the coin toss shows up heads or tails. Within the blocks of two and five assets, the decision problems were randomized in order.

<i>Set</i>	Asset A		Asset B		Asset C		Asset D		Asset E	
	<i>Heads</i>	<i>Tails</i>	<i>Heads</i>	<i>Tails</i>	<i>Heads</i>	<i>Tails</i>	<i>Heads</i>	<i>Tails</i>	<i>Heads</i>	<i>Tails</i>
#1	0.00	1.20	3.60	0.00						
#2	3.60	0.00	0.00	1.80						
#3	4.80	0.00	0.00	1.20						
#4	2.30	0.00	0.00	4.50						
#5	0.00	2.40	2.40	0.00						
#6	1.20	0.00	0.00	4.80						
#7	0.00	2.30	4.50	0.00						
#8	0.00	3.60	1.80	0.00						
#9	0.00	2.70	3.00	0.00						
#10	1.20	0.00	0.00	3.60						
#11	0.30	2.70	0.90	0.90	1.20	0.00	0.60	1.80	0.00	3.60
#12	0.80	1.50	2.40	0.00	0.40	2.10	1.80	0.80	0.00	3.00
#13	2.30	0.60	0.40	1.50	0.00	2.40	1.50	0.90	3.00	0.00
#14	0.50	4.10	1.80	0.00	0.00	5.40	0.90	2.70	0.50	0.50
#15	2.70	0.30	3.60	0.00	0.00	1.20	0.90	0.90	1.80	0.60
#16	2.00	1.20	3.50	0.40	4.50	0.00	0.00	3.00	1.10	2.30
#17	1.40	0.20	0.00	1.80	0.50	1.40	0.80	0.80	1.80	0.00
#18	2.70	0.50	3.60	0.00	0.90	1.40	0.00	1.80	1.80	0.90
#19	0.00	2.40	2.40	0.00	1.80	0.60	0.60	1.80	1.20	1.20
#20	0.00	4.50	3.00	0.00	2.00	0.80	0.40	3.50	1.50	2.30
#21	0.00	3.60	2.70	0.90	3.60	0.00	1.50	1.50	0.60	2.70
#22	2.40	0.40	1.80	0.80	0.00	2.40	3.60	0.00	0.90	1.80
#23	0.30	2.70	1.50	0.60	1.20	1.80	2.40	0.00	0.00	3.60
#24	5.40	0.00	2.70	0.90	0.50	0.50	0.00	1.80	4.10	0.50
#25	0.50	2.70	1.80	0.00	1.40	0.90	0.90	1.80	0.00	3.60

Figure B1: Opportunity sets in the allocation decision task. In each panel of this figure, the vertical (horizontal) axis indicates the return per 1 SEK invested if the coin shows up heads (tails). Each dot indicates a single asset. The labels *FC-1*, *FC-2*, *FMR-1*, and *FMR-2* denote particular opportunity sets used for constructing the decision-making quality measures “financial competence” (*FC*) and “failure to minimize risk” (*FMR*).

C. Decision-Making Quality Measures

In each opportunity set $j \in \{1, 2, \dots, 25\}$, each participant i is endowed with 100 SEK to be allocated on assets $k \in \{1, 2, \dots, 5\}$. Let $a_{i,j,k}$ denote the fraction of the endowment allocated on asset k such that $\sum_k a_{i,j,k} = 1$.

The return per SEK invested in asset k if the coin comes up heads is denoted as $h_{j,k}$; the return per SEK invested if it comes up tails is denoted as $t_{j,k}$. Thus, the return of participant i 's allocation in opportunity set j will either be

$$\begin{aligned} H_{i,j} &= \sum_k a_{i,j,k} \cdot h_{j,k} && \text{if the coin comes up heads, or} \\ T_{i,j} &= \sum_k a_{i,j,k} \cdot t_{j,k} && \text{if the coin comes up tails.} \end{aligned}$$

Let the tuple $\mathbf{x}_{i,j} = (H_{i,j}, T_{i,j})$ denote the portfolio of participant i in opportunity set j . Following Banks et al. (2018), we define four measures of decision-making quality: (i) violations of first order stochastic dominance (*FOSD*), (ii) violations of the generalized axiom of revealed preferences (*GARP*), (iii) financial competence (*FC*), and (iv) failure to minimize risk (*FMR*). Each of these measures is defined in detail below.

Expected Return. The expected portfolio return of participant i 's investment in opportunity set j , i.e., the expected return from allocating the endowment on the available assets, is given by

$$ER_{i,j} = \frac{H_{i,j} + T_{i,j}}{2}.$$

Participant i 's mean expected return, ER_i , is calculated as the average of $ER_{i,j}$ across 23 of the 25 opportunity sets, as the expected returns are identical for all portfolios in the two remaining opportunity sets (set #5 and #19; see Table B1 and Figure B1), i.e., $ER_i = 1/23 \cdot \sum_{j=1}^{23} ER_{i,j}$.

Standard Deviation. As a measure of portfolio risk, we calculate the standard deviation of participant i 's portfolio in opportunity set j , i.e., the standard deviation of $H_{i,j}$ and $T_{i,j}$ occurring with a probability of 50% each:

$$SD_{i,j} = \sqrt{\frac{H_{i,j}^2 + T_{i,j}^2}{2} - \left(\frac{H_{i,j} + T_{i,j}}{2}\right)^2}.$$

The average portfolio risk for individual i , SD_i , is defined as the mean standard deviation across all 25 opportunity sets, i.e., $SD_i = 1/25 \cdot \sum_{j=1}^{25} SD_{i,j}$.

Violations of First Order Stochastic Dominance (*FOSD*). Suppose participant i chooses portfolio $\mathbf{x}_{i,j} = (H_{i,j}, T_{i,j})$ in opportunity set j , i.e., an allocation returning $H_{i,j}$ or $T_{i,j}$ with 50% probability each, as illustrated in Figure C1. The figure also depicts an alternative portfolio $\mathbf{x}'_{i,j} = (H'_{i,j}, T'_{i,j})$. Apparently, portfolio $\mathbf{x}'_{i,j}$ first order stochastically dominates $\mathbf{x}_{i,j}$ because $T'_{i,j} = H_{i,j}$ and $H'_{i,j} > T_{i,j}$.

Following Banks et al. (2018), we use the difference between the maximum expected return of a portfolio that provides the same minimum payoff as the chosen portfolio and the expected return of the chosen

Figure C1: Violation of the principle of first order stochastic dominance (FOSD). The chosen portfolio $x_{i,j}$ is first order stochastically dominated by the alternative portfolio $x'_{i,j}$ because $T'_{i,j} = H_{i,j}$ and $H'_{i,j} > T_{i,j}$. That is, portfolio $x'_{i,j}$ yields a higher expected return than $x_{i,j}$ while guaranteeing the same minimum payoff.

portfolio as a measure of how closely participant i 's choice in opportunity set j complies with the principle of FOSD (Hadar and Russell, 1969), i.e., $FOSD_{i,j} = (H'_{i,j} + T'_{i,j}) \cdot 1/2 - (H_{i,j} + T_{i,j}) \cdot 1/2$ in the example depicted in Figure C1.

In general, given a chosen portfolio $x_{i,j} = (H_{i,j}, T_{i,j})$, let $h_j^* = \max_k h_{j,k}$ be the maximum return across all assets k if the coin comes up heads and $t_j^* = \max_k t_{j,k}$ if the coin comes up tails. By investing the fraction

$$w = \frac{\min(H_{i,j}, T_{i,j})}{\min(h_j^*, t_j^*)}$$

on the asset paying $\min(h_j^*, t_j^*)$ and 0 SEK otherwise, and investing the fraction $(1 - w)$ on the asset paying $\max(h_j^*, t_j^*)$ and 0 SEK otherwise, participant i maximizes the expected return but still guarantees a minimum return of $\min(H_{i,j}, T_{i,j})$. Thus, our measure of FOSD is:

$$FOSD_{i,j} = \left(w \cdot \frac{\min(h_j^*, t_j^*)}{2} + (1 - w) \cdot \frac{\max(h_j^*, t_j^*)}{2} \right) - \frac{(H_{i,j} + T_{i,j})}{2}.$$

To assess participant i 's average violations of FOSD, we average the measure over all choices, except for the two opportunity sets for which any portfolio will yield the same expected returns (set #5 and #19; see Table B1 and Figure B1), i.e., $FOSD_i = 1/23 \cdot \sum_{j=1}^{23} FOSD_{i,j}$.

Violations of the General Axiom of Revealed Preferences (GARP). According to the Generalized Axiom of Revealed Preferences, for any two opportunity sets m and n ($m \neq n$), if participant i reveals to prefer $\mathbf{x}_{i,m}$ over $\mathbf{x}_{i,n}$, then $\mathbf{x}_{i,n}$ is not strictly preferred to $\mathbf{x}_{i,m}$. Figure C2 illustrates an example of a GARP violation where participant i chooses $\mathbf{x}_{i,m}$ in opportunity set m when $\mathbf{x}_{i,n}$ is affordable, and chooses $\mathbf{x}_{i,n}$ in opportunity set n when $\mathbf{x}_{i,m}$ is affordable.

Let p_j denote the ratio of maximum returns for heads and tails in opportunity set j , respectively, i.e., $p_j = h_j^*/t_j^*$. In the example shown in Figure C2, $\mathbf{x}_{i,m}$ is revealed preferred to $\mathbf{x}_{i,n}$ at price p_m and $\mathbf{x}_{i,n}$ is revealed preferred to $\mathbf{x}_{i,m}$ at price p_n .

Figure C2: Violation of the generalized axiom of revealed preferences. Portfolio $\mathbf{x}_{i,m}$ is revealed preferred to portfolio $\mathbf{x}_{i,n}$ at price p_m and $\mathbf{x}_{i,n}$ is revealed preferred to $\mathbf{x}_{i,m}$ at price p_n . $MPI_{i,m,n} = \alpha_{i,m,n} + \beta_{i,m,n}$ with $\alpha_{i,m,n} = p_m \cdot (\mathbf{x}_{i,m} - \mathbf{x}_{i,n})$ and $\beta_{i,m,n} = p_n \cdot (\mathbf{x}_{i,n} - \mathbf{x}_{i,m})$.

The extent of violations of GARP is measured with the Money Pump Index (MPI), which is based on the idea that an arbitrageur can exploit violations in revealed preferences (Echenique et al., 2011): The arbitrageur could make profit by buying portfolio $\mathbf{x}_{i,m}$ at price p_n and then selling it at price p_m ; likewise, the arbitrageur could buy portfolio $\mathbf{x}_{i,n}$ at price p_m and sell it at price p_n . The Money Pump Index is the total profit the arbitrageur could make, i.e.,

$$\begin{aligned} MPI_{i,m,n} &= \alpha_{i,m,n} + \beta_{i,m,n} \\ MPI_{i,m,n} &= p_m \cdot (\mathbf{x}_{i,m} - \mathbf{x}_{i,n}) + p_n \cdot (\mathbf{x}_{i,n} - \mathbf{x}_{i,m}). \end{aligned}$$

We calculate the money pump for each violation of GARP, i.e., for $25 \cdot (25 - 1) \cdot 1/2 = 300$ pairwise combinations of opportunity sets. For each participant i , we determine the average money pump index, over all pairwise combinations, i.e., $MPI_i = 1/300 \cdot \sum_{m=1}^{25} \sum_{n=1}^{25} MPI_{i,m,n} \forall m > n$.

Financial competence (FC). Four opportunity sets were identical in the two-asset- and the five-asset-frame, i.e., four sets were presented in both the two-asset- (sets #1, #2, #8, and #10) and the five-asset-frame (sets #11, #15, #18, and #25). Moreover, two of the four opportunity sets presented in the two-asset- and five-asset-frame, respectively, were constructed as mirror images of one another, i.e., only the payoffs for heads and tails were interchanged. Thus, two opportunity sets (denoted as FC_1 and FC_2 in Figure B1) were effectively presented four times each ($\#1 = \#10 = \#11 = \#15$ and $\#2 = \#8 = \#18 = \#25$).

Let $J_1 = \{\#1, \#10, \#11, \#15\}$ and $J_2 = \{\#2, \#8, \#18, \#25\}$. Thus, J_1 and J_2 are sets containing identical opportunity sets presented as mirror images and/or in the two-asset- or the five-asset-frame, respectively. The financial competence of individual i is defined as the average absolute differences between the expected returns across the identical opportunity sets in J_1 and J_2 , i.e.,

$$FC_i = \frac{1}{12} \cdot \left(\sum_{k,l \in J_1} |ER_{i,k} - ER_{i,l}| + \sum_{m,n \in J_2} |ER_{i,m} - ER_{i,n}| \right) \quad \forall k > l \ \& \ m > n.$$

Note that our definition of FC_i differs from the measure used by Banks et al. (2018), who average the absolute differences in expected returns across the two frames, but not across the mirrored versions of the sets. However, the notion of “financial competence”, arguably, is not only reflected by choosing the same portfolio in an opportunity set presented in frames with varying complexity, but also by choosing the same portfolio if heads and tails are interchanged.

Failure to minimize risk (FMR). In two opportunity sets (#5 and #19; see Figure B1), the expected return per 1 SEK invested was the same for all assets k , such that all feasible portfolios will share the same expected return. Choosing a fully-hedged portfolio (i.e., a zero-risk portfolio), thus, (second-order) dominates all other feasible portfolios in these two opportunity sets. The failure to minimize risk for subject i in opportunity set j , $FMR_{i,j}$, is measured as the standard deviation $SD_{i,j}$ of the particular portfolio allocation, which is then averaged over the two opportunity sets, i.e.,

$$FMR_i = \frac{1}{2} \cdot \sum_{j=1}^2 SD_{i,j}.$$

Decision-making quality index (DMQI). As a measure of decision-making quality, Banks et al. (2018) calculate the arithmetic mean of the four measures $FOSD$, MPI , FC , and FMR . We suggest an alternative approach for two reasons. First, the measures on the subject level are not based on comparable aggregation levels (for instance, the FMR index is constructed as the mean over two opportunity sets, whereas MPI is averaged across 300 pairwise combinations of opportunity sets). This implies that shortcomings in decision-making are weighted differently for each of the four indices. Second, there seems to be no obvious reason why the four measures—which can differ considerably in magnitude on an individual level—should be equally weighted.

Thus, rather than determining the mean of the four measure on the subject level, we utilize the predicted values of a principal component analysis of the four measures to constitute the $DMQI$ index for each participant. The predictions of the principal component analysis summarized in Table C1 serve as

a unified decision-making quality index, denoted as $DMQI$, on the participant level. Note that, in theory, the predicted values have a mean of zero and a standard deviation of unity. Thus, positive values can be interpreted as above average while negative values indicate that a participants' decision-making quality is below average.

Table C1: Principal component analysis of the four decision-making quality measures. This table outlines the four principal components of the $DMQI$, i.e., (i) violations of first order stochastic dominance ($FOSD$), (ii) money pump index (MPI), (iii) financial competence, and (iv) failure to minimize risk (FMR); $n = 958$. **Panel (A)** reports the eigenvalue and the proportion of explained variance for each of the four components. **Panel (B)** reports unrotated components.

Panel (A)				
<i>Component</i>	<i>Eigenvalue</i>	<i>Difference</i>	<i>Proportion</i>	<i>Cumulative</i>
<i>Comp. #1</i>	2.034	1.046	0.509	0.509
<i>Comp. #2</i>	0.989	0.279	0.247	0.756
<i>Comp. #3</i>	0.710	0.444	0.178	0.933
<i>Comp. #4</i>	0.267	.	0.067	1.000

Panel (B)				
<i>Component</i>	<i>Comp. #1</i>	<i>Comp. #2</i>	<i>Comp. #3</i>	<i>Comp. #4</i>
<i>FOSD</i>	0.621	−0.099	−0.318	0.710
<i>MPI</i>	0.618	0.026	−0.360	−0.698
<i>FC</i>	0.457	−0.228	0.859	−0.048
<i>FMR</i>	0.154	0.968	0.179	0.081

D. Questionnaires and Side Tasks

After the main experiment, participants were asked to answer a set of Likert items—all scaled from 0 (minimum) to 10 (maximum)—which are summarized in Table D1 below. The questions on risk tolerance and patience are based on Dohmen et al. (2011) and Falk et al. (2016, 2018); and the two statements addressing the proneness to shift blame are based on the inventory introduced by Wilson et al. (1990). In addition to the survey items reported in Table D1, participants were exposed to a hypothetical charitable giving setting based on Falk et al. (2018), asking how much they would donate to a good cause if they had unexpectedly received 10,000 SEK.¹¹ The 14 items were displayed on five separate screens: the first screen contained the questions regarding risk preferences in general and their willingness to give up something today in order to benefit more in the future; the second screen included the item of risk tolerance in financial matters as well as the three statements on trust; the third screen showed the hypothetical charitable giving task; the fourth screen comprised the two questions on blame shifting; and the fifth screen involved the five questions related to financial investments and the use of expertise.

Once the above questionnaires had been completed, participants answered eight questions allowing to determine their numeracy skills. As a measure of subjects' numeracy, we use the number of correct answers. The numeracy task is based on the Rasch-validated inventory proposed by Weller et al. (2013). Two of the eight questions in the original set are well-known items from the Cognitive Reflection Test (CRT) introduced by Frederick (2005). Since this three-item test has been widely spread on the Internet, many people will know the questions and the corresponding answers. Therefore, the two items on cognitive reflection skills have been replaced by items from the CRT proposed by Toplak et al. (2014). For answering the eight questions, participants faced a time constraint of four minutes. Since the items, by construction of the test, differ considerably in difficulty, the order of the questions has been randomized to avoid systematic effects arising from the time constraint. The questions used in the numeracy task are listed in Table D2.

After submitting their answers to the numeracy questions, participants were asked to self-assess their performance in the task in two different ways: (i) "How many of the eight questions you answered on the previous screen did you answer correctly?" (0 to 8), and (ii) "Compared to a random sample of the Swedish population, how did you score in terms of correct answers? Please estimate your position in the ranking." (Top 10%, Top 20%, ..., Bottom 20%, Bottom 10%). While the first question allows for determining participants' overestimation of their own skills (as the difference between their estimates and actual performance), the second question allows for quantifying subjects' tendency to "overplace" their performance relative to others. Question (ii) asks participants to evaluate their performance relative to a *random* sample of the Swedish population. However, our sample is not representative with respect to the level of education due to self-selection effects. For this reason we take a detour to derive a sensible measure of overplacement: The validated inventory proposed by Weller et al. (2013) is constructed in such a way that scores are approximately normally distributed among a general population sample. The fact that the numeracy scores in our general population sample are significantly different from a normal distribution (Shapiro-Wilk-Test; $W = 0.987, p < 0.001, n = 550$) somewhat confirms our conjecture of a self-selection effect in our sample. Thus, in a first step, we draw random integers from a normal distribution with a mean of 4.07 and a standard deviation of 1.83, the first and second moment reported for Study 2 in Weller et al. (2013), validating their Rasch-based measure. In a second step, we determine the percentiles associated with each possible score between 0 and 8. Finally, we deduct the percentile (from the sampled normal

¹¹ The question was presented to participants as follows: "Imagine the following situation: Today you unexpectedly received 10,000 SEK. How much of this amount would you donate to a good cause?"

distribution) corresponding to participants' numeracy score from their estimated decile, i.e., their answer to question (ii), to assess the degree of participants' overplacement.¹² As a final task of the experiment, participants were asked to answer six single-choice questions based on van Rooij et al. (2011), allowing to determine their financial literacy. In particular, three of the questions stem from their basic literacy inventory (*Q1–Q3*, corresponding to (2), (3), and (5) in van Rooij et al. (2011)), and three questions are based on the advanced literacy inventory (*Q4–Q6*, corresponding to (12), (16), and (7) in van Rooij et al. (2011)). As an index of financial literacy, we use the sum of participants' correct answers. The questions used in the financial literacy task are depicted in Table D3.

Descriptive results relating to the questionnaires are provided in Table E5 and Figures E1 and E2; summary results of the side experiments on numeracy skills, financial literacy, and the two measures of overconfidence are provided in Table E6 and Figures E3 and E4 in Appendix E.

¹² As we ask participants to estimate their performance relative to the general population in deciles rather than percentiles, we use the *minimum* difference to either of the bounds of the interval they implicitly provide as our measure of overestimation. That is, if the percentile (from the sampled normal distribution) lies within the interval subjects estimate, the measure takes value 0; if the percentile is smaller than the lower bound (upper bound) of the estimated interval, we evaluate the percentile to the lower bound (upper bound) of the interval.

Table D1: Survey questions. This table summarizes the Likert items, all participants answered after the main experimental task. In particular, the table depicts the variable description as referred to in the main text, the wording of the question/statement, and the corresponding labelling of the minimum and maximum values for each item. The three items indicated with an asterisk were only displayed if the question “Frequent Investments” was not answered with 0 (“does not describe me at all”).

<i>Variable</i>	<i>Question / Statement</i>	<i>Likert Scale</i>	
		<i>min (0)</i>	<i>max (10)</i>
<i>Risk Tolerance (in General)</i>	Are you generally a person who is willing to take risks or do you try to avoid taking risks?	not at all willing to take risks	very willing to take risks
<i>Patience (in General)</i>	How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?	not at all willing to give up something today	very willing to give up something today
<i>Risk Tolerance in Financial Matters</i>	I am generally willing to take risks in financial matters.	does not describe me at all	describes me perfectly
<i>Trust (in General)</i>	I generally trust other people.	does not describe me at all	describes me perfectly
<i>Trust in Finance Professionals</i>	I generally trust employees from the finance industry.	does not describe me at all	describes me perfectly
<i>Trust in Investment Algorithms</i>	I generally trust robo-advisors (i.e. computer programs) in financial matters.	does not describe me at all	describes me perfectly
<i>Blame Shifting (Others)</i>	If you hurt yourself accidentally, do you sometimes blame somebody who happens to be nearby even though you realize, on reflection, that they were not responsible?	I never blame others	I often blame others
<i>Blame Shifting (Temptation)</i>	Can you easily resist the temptation to blame others for the accidents that happen to you?	I can resist easily	I cannot resist at all
<i>Frequent Investments</i>	I frequently invest in stocks and mutual funds myself (not through the national pension system).	does not describe me at all	describes me perfectly
<i>Delegate to Fin. Profs.*</i>	I delegate my investment decisions (e.g., purchase of stocks, bonds, investment funds, real estate) to financial advisors at banks or other institutions and refrain from taking decisions myself.	does not describe me at all	describes me perfectly
<i>Delegate to Inv. Algos.*</i>	I delegate my investment decisions (e.g., purchase of stocks, bonds, investment funds, real estate) to robo-advisors at banks or other institutions and refrain from taking decisions myself.	does not describe me at all	describes me perfectly
<i>Use Expertise of Fin. Profs.*</i>	I use the expertise of financial advisers for my investments/pension savings.	does not describe me at all	describes me perfectly
<i>Responsibility in Financial Matters</i>	I am solely responsible for financial decisions in my household.	does not describe me at all	describes me perfectly

Table D2: Numeracy inventory based on Weller et al. (2013). This table summarizes the questions used to assess participants' numeracy and the correct answers to each of the questions. For answering all items, participants were given a maximum of four minutes. The inventory proposed by Weller et al. (2013) includes two questions from Frederick (2005). As these are likely to be known by many people, items *Q2* and *Q3* have been replaced by questions from Toplak et al. (2014).

<i>ID</i>	<i>Question</i>	<i>Correct Answer</i>
<i>Q1</i>	Suppose you have a close friend who has a lump in her breast and must have a mammogram. Of 100 women like her, 10 of them actually have a malignant tumor and 90 of them do not. Of the 10 women who actually have a tumor, the mammogram indicates correctly that 9 of them have a tumor and indicates incorrectly that 1 of them does not. Of the 90 women who do not have a tumor, the mammogram indicates correctly that 81 of them do not have a tumor and indicates incorrectly that 9 of them do have a tumor. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?	50 percent
<i>Q2</i>	If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?	4 days
<i>Q3</i>	A man buys a pig for 600 SEK, sells it for 700 SEK, buys it back for 800 SEK, and sells it finally for 900 SEK. How much has he made?	200 SEK
<i>Q4</i>	In a lottery, the chance of winning a car is 1 in 1000. What percent of lottery tickets win a car?	0.1 percent
<i>Q5</i>	In a lottery, the chances of winning a 10.000 SEK prize are 1%. What is your best guess about how many people would win a 10.000 SEK prize if 1000 people each buy a single lottery ticket?	10 people
<i>Q6</i>	Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number?	500 times
<i>Q7</i>	If the chance of getting a disease is 20 out of 100, this would be the same as having a ... chance of getting the disease.	20 percent
<i>Q8</i>	If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?	100 people

Table D3: Financial literacy inventory based on van Rooij et al. (2011). This table summarizes the questions used to assess participants' literacy in financial matters and the corresponding choice options to each of the questions. Correct answers are highlighted in *italics*. For answering all items, participants were given a maximum of three minutes.

<i>ID</i>	<i>Question</i>	<i>Choices</i>
<i>Q1</i>	Suppose you had 1,000 SEK in a savings account and the interest rate is 20% per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?	<ul style="list-style-type: none"> ○ <i>more than 2,000 SEK</i> ○ exactly 2,000 SEK ○ less than 2,000 SEK ○ do not know
<i>Q2</i>	Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?	<ul style="list-style-type: none"> ○ more than today ○ exactly the same ○ <i>less than today</i> ○ do not know
<i>Q3</i>	Suppose that in the year 2025, your income after tax has doubled and prices of all goods have doubled too. In 2025, how much will you be able to buy with your income?	<ul style="list-style-type: none"> ○ more than today ○ <i>exactly the same</i> ○ less than today ○ do not know
<i>Q4</i>	When an investor spreads his money among different assets, does the risk of losing money in general:	<ul style="list-style-type: none"> ○ increase ○ <i>decrease</i> ○ stay the same ○ do not know
<i>Q5</i>	If the interest rate falls, what should happen to bond prices?	<ul style="list-style-type: none"> ○ <i>rise</i> ○ fall ○ stay the same ○ none of the above ○ do not know
<i>Q6</i>	Which of the following statements is correct? If somebody buys the stock of firm B in the stock market:	<ul style="list-style-type: none"> ○ <i>he owns a part of firm B</i> ○ he has lent money to firm B ○ he is liable for the firm B's debt ○ none of the above ○ do not know

E. Descriptive Results

In the following, we present a set of descriptive results for all measures elicited in the experiment. Many of these variables only enter our analyses as controls. Yet, while several results presented below back up our main findings, we also deem it interesting to compare our two subject pools—participants from the general population and professionals from the finance industry—along these measures.

Response rate analysis and self-selection. For the finance professionals group, the response rate analysis shows that men responded to a greater extent than women, and that finance individuals in the age group 45–59 years responded to a slightly lesser extent than other ages. Furthermore, the non-response analysis shows that those with the lowest income responded to a somewhat higher degree compared to the others, and that those with a post-secondary education level of three years or more responded to greater extent than others. In the case of country of birth, the response rate was slightly higher for those born in Sweden compared to other countries. In the finance group there was a certain difference between the different job codes where the response frequency was slightly lower (5%) in the group of traders and portfolio managers (job code “2414”) compared with analysts and advisors (code “2413”) and brokers (code “3311”) (6.4%).

For the general population group, the response rate analysis shows similar patterns regarding gender, i.e., men responded to a greater extent than women. The response rate was lowest among the elderly. Furthermore, the response rate analysis indicates that those with the lowest and highest income responded to a somewhat higher extent compared to other income groups. When it comes to the level of education, those with a post-secondary education of three years or more tend to be over-represented in our sample. In the case of country of birth, the response rate was slightly higher for the ones born in Sweden compared to other countries.

A detailed summary of participants demographics compared to the characteristics of the sample invited is presented in Table E1. In particular, Table E1 reports the number of respondents and non-respondents per category of several socio-demographic characteristics, separated for both samples, as reported by SCB. Moreover, we report χ^2 -tests comparing whether participants in our samples differ significantly from those who have been invited by SCB but did not participate in the experiment. As discussed in Appendix A, we report self-selection effects in terms of gender, age, country of birth, income, and education for the general population sample, and self-selection effects with respect to gender, age, and education for the finance professionals sample. For a discussion of how this translates into the decision-making quality measures, see the discussion below.

Decision-making quality measures. Table E2 reports summary statistics of the decision-making quality measures derived from the main task (see Appendix B and C for details), separated for the general population sample and the finance professionals. Since we conjecture that the varying incentive structure between treatments in the finance professionals’ sample gives rise to differences in decision-making quality, the summary statistics are separated for the three treatments. Notably, all measures—except for the index capturing financial competence (*FC*)—are considerably left-skewed for both samples and across all treatments.

On average, decision-making quality is remarkably high, both for finance professionals *and* the general population sample. Comparing the means of the decision-making measures in our samples to the means reported by Banks et al. (2018)¹³ shows that our general population sample performs significantly better than the participants in Banks et al. (2018) in all four measures (two-sample *t*-tests; $p < 0.001$ for all comparisons). This, too, indicates that our sample is highly sophisticated, which might be one of the reasons why we do not find strong differences in decision-making quality between the finance professionals and the general population in our results.

¹³ Banks et al. (2018) report the following means (*m*) and standard deviations (*sd*) for their (“pre-reform”) general population sample ($n = 1,416$) from the US: $m = -\$3.68$, $sd = \$1.93$ (*FOSD*); $m = -\$2.77$, $sd = \$3.87$ (*MPI*); $m = -\$6.98$, $sd = \$3.97$ (*FC*); and $m = -\$3.32$, $sd = \$4.91$ (*FMR*). Two pivotal differences need to be taken into account, to come up with sensible comparisons: (i) in Banks et al. (2018), participants receive an endowment of of \$25 to be allocated on the assets; in contrast, in our setting, participants were endowed with 100 SEK; (ii) all assets’ payoffs from Banks et al. (2018) have been multiplied by a factor of 1.5 in our experiment. Thus, multiplying the means and standard deviations reported by Banks et al. (2018) by a factor of $100/25 \cdot 1.5 = 6$ provides sensible benchmarks for the decision-making quality measures in our setting. In particular, the summary statistics provided in Table E2 can be compared to the following adjusted values from Banks et al. (2018): $m = -\$22.08$, $sd = \$11.58$ (*FOSD*); $m = -\$16.62$, $sd = \$23.22$ (*MPI*); $m = -\$41.88$, $sd = \$23.82$ (*FC*); and $m = -\$19.92$, $sd = \$29.46$ (*FMR*).

Table E1: Sample characteristics by subject pools. This table depicts the number (in %) of respondents (“Resp.”), i.e., those who participated in our experiment, and non-respondents (“No Resp.”), i.e., those who were invited but did not participate, for a number of different characteristics, separated for the general population and the finance profession sample. χ^2 -tests (with $k-1$ degrees of freedom) and the corresponding p -values are reported.

	General Population			Finance Professionals		
	Resp.	No Resp.	χ^2 / p	Resp.	No Resp.	χ^2 / p
<i>Gender:</i>						
Male	55.35	49.36	9.322	75.30	68.47	10.169
Female	44.65	50.64	(0.002)	24.70	31.53	(0.001)
<i>Age:</i>						
20 – 29 years	11.55	10.28	37.789	11.85	8.73	14.062
30 – 39 years	31.69	23.18	(< 0.001)	31.12	28.79	(0.015)
40 – 49 years	26.62	26.39		28.51	30.04	
50 – 59 years	20.99	26.74		17.27	22.83	
60 – 69 years	9.15	13.41		10.04	8.60	
70 – 79 years	0.00	0.00		1.20	1.00	
<i>Country of Birth:</i>						
Sweden	88.17	82.84	13.248	89.76	88.95	0.311
Abroad	11.83	17.16	(< 0.001)	10.24	11.05	(0.577)
<i>Citizenship:</i>						
Swedish	97.04	95.64	3.132	97.59	96.53	1.604
Foreign	2.96	4.36	(0.077)	2.41	3.47	(0.205)
<i>Marital Status:</i>						
Married	46.90	46.26	2.247	52.21	56.31	4.910
Unmarried	41.41	40.49	(0.523)	40.36	35.46	(0.179)
Divorced	11.27	12.42		7.03	7.79	
Widowed	0.42	0.83		0.40	0.45	
<i>Income:</i>						
< 124,999 SEK	3.24	2.70	25.646	2.01	1.53	2.985
125,000 – 199,999 SEK	5.63	5.79	(< 0.001)	2.41	2.16	(0.560)
200,000 – 279,999 SEK	12.82	15.25		3.01	3.41	
280,000 – 369,999 SEK	24.08	31.16		5.22	6.85	
> 370,000 SEK	54.23	45.11		87.35	86.06	
<i>Education:</i>						
No High School	1.83	8.89	198.587	0.80	1.08	32.058
High School	28.45	46.89	(< 0.001)	7.83	17.06	(< 0.001)
University (< 3 years)	19.86	14.95		11.45	11.32	
University (> 3 years)	49.86	28.61		79.72	69.95	
Unknown, n/a	0.00	0.66		0.20	0.59	

Table E2: Descriptive statistics of the decision-making quality measures. This table reports the means and standard deviations (in parentheses) as well as the median and interquartile ranges (*IQR*; in brackets) for to four decision-making quality measures (i) violations of first order stochastic dominance (*FOSD*), (ii) money pump index (*MPI*), (iii) financial competence (*FC*), (iv) failure to minimize risk (*FMR*), as well as the unified measure of decision-making quality (*DMQI*), separated for the general population sample (pooled across all treatments) as well as the three treatments conducted among finance professionals.

	<i>GP-*</i>		<i>FP-FIXED</i>		<i>FP-ALIGNED</i>		<i>FP-OWN</i>	
	<i>m / sd</i>	<i>q50 / iqr</i>	<i>m / sd</i>	<i>q50 / iqr</i>	<i>m / sd</i>	<i>q50 / iqr</i>	<i>m / sd</i>	<i>q50 / iqr</i>
<i>FOSD</i>	-4.78 (4.55)	-4.04 [3.96]	-4.35 (5.63)	-3.01 [5.72]	-3.26 (3.63)	-2.59 [3.63]	-2.70 (3.12)	-1.71 [3.84]
<i>MPI</i>	-9.71 (24.40)	-3.63 [7.83]	-13.27 (35.04)	-2.05 [11.38]	-7.38 (17.61)	-2.33 [7.44]	-5.30 (10.88)	-0.89 [6.64]
<i>FC</i>	-14.55 (10.23)	-12.96 [11.38]	-12.46 (10.20)	-10.05 [14.83]	-11.37 (8.58)	-11.28 [12.38]	-11.70 (10.36)	-9.46 [15.40]
<i>FMR</i>	-7.54 (20.54)	0.00 [0.00]	-15.05 (31.20)	0.00 [7.80]	-14.36 (30.39)	0.00 [12.00]	-11.60 (28.70)	0.00 [1.50]
<i>DMQI</i>	-0.08 (1.22)	0.16 [0.91]	-0.11 (1.51)	0.18 [1.07]	0.16 (0.90)	0.22 [0.84]	0.24 (0.74)	0.35 [0.91]
Obs.	550		132		139		137	

Table E3: Differences in decision-making quality measures. This table reports the *t*-statistics from two-sample *t*-tests between the general population sample (pooled across all treatments) and the finance professionals sample separated for the treatment conditions for the four decision-making quality measures (i) violations of first order stochastic dominance (*FOSD*), (ii) money pump index (*MPI*), (iii) financial competence (*FC*), (iv) failure to minimize risk (*FMR*), as well as the unified measure of decision-making quality (*DMQI*). Standard errors (*se*) are reported in parentheses. Means, standard deviations, medians, and interquartile ranges for each all measures in all treatments are reported in Table E2. * $p < 0.05$, ** $p < 0.005$.

	<i>GP-*</i> vs. <i>FP-FIXED</i>	<i>GP-*</i> vs. <i>FP-ALIGNED</i>	<i>GP-*</i> vs. <i>FP-OWN</i>	<i>FP-FIXED</i> vs. <i>FP-ALIGNED</i>	<i>FP-FIXED</i> vs. <i>FP-OWN</i>	<i>FP-ALIGNED</i> vs. <i>FP-OWN</i>
	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>
<i>FOSD</i>	-0.647 (0.592)	-3.014** (0.490)	-4.298** (0.473)	-1.912 (0.572)	-2.990** (0.552)	-1.366 (0.408)
<i>MPI</i>	0.815 (3.752)	-0.962 (2.949)	-1.766 (2.784)	-1.763 (3.344)	-2.540* (3.140)	-1.178 (1.765)
<i>FC</i>	-1.305 (1.155)	-2.446* (1.063)	-1.967 (1.150)	-0.955 (1.143)	-0.603 (1.254)	0.294 (1.144)
<i>FMR</i>	2.236* (3.030)	2.063* (2.952)	1.160 (2.868)	-0.184 (3.742)	-0.944 (3.654)	-0.776 (3.559)
<i>DMQI</i>	0.292 (0.166)	-1.613 (0.137)	-2.321* (0.132)	-1.788 (0.151)	-2.459* (0.145)	-0.867 (0.100)
Obs.	315	322	320	271	269	276

Table E4: Correlations between the decision-making quality measures. This table reports Pearson correlations between the decision-making quality measures (i) violations of first order stochastic dominance (*FOSD*), (ii) money pump index (*MPI*), (iii) financial competence (*FC*), (iv) failure to minimize risk (*FMR*), as well as the unified measure of decision-making quality (*DMQI*), separated for the general population sample (lower triangular matrix) and the finance professionals subject pool (upper triangular matrix). *p*-values are reported in parentheses.

	<i>FOSD</i>	<i>MPI</i>	<i>FC</i>	<i>FMR</i>	<i>DMQI</i>
<i>FOSD</i>		0.776 (< 0.001)	0.334 (< 0.001)	−0.084 (0.090)	0.740 (< 0.001)
<i>MPI</i>	0.703 (< 0.001)		0.327 (< 0.001)	0.075 (0.131)	0.871 (< 0.001)
<i>FC</i>	0.418 (< 0.001)	0.379 (< 0.001)		−0.131 (0.008)	0.685 (< 0.001)
<i>FMR</i>	0.281 (< 0.001)	0.261 (< 0.001)	0.240 (< 0.001)		0.222 (< 0.001)
<i>DMQI</i>	0.766 (< 0.001)	0.861 (< 0.001)	0.762 (< 0.001)	0.451 (< 0.001)	

Table E5: Descriptive statistics and comparisons between pools for the survey items. This table reports the means and standard deviations (in parentheses) for all survey items included in the experiment, separated for the general population and the finance professionals subject pool. The column “*t*-test” reports the differences in means and the *t*-values (in brackets) from two-sample *t*-tests based on $n = 958$. * $p < 0.05$, ** $p < 0.005$.

	<i>Gen. Pop.</i>	<i>Fin. Prof.</i>	<i>t-Test</i>
Altruism/Hypothetical Charitable Giving	0.79 (1.37)	0.85 (1.69)	−0.061 [0.099]
Blame Shifting - Others	1.12 (1.56)	1.19 (1.59)	−0.065 [0.103]
Blame Shifting - Temptation	1.55 (2.11)	1.78 (2.20)	−0.239 [0.140]
Risk Tolerance in General	4.79 (2.14)	5.80 (1.94)	−1.017** [0.135]
Risk Tolerance	4.28 (2.34)	6.09 (2.13)	−1.806** [0.147]
Patience in General	6.03 (2.00)	7.21 (1.81)	−1.179** [0.125]
Trust in General	5.71 (2.36)	5.79 (2.21)	−0.080 [0.150]
Trust in Finance Professionals	4.16 (2.33)	4.63 (2.37)	−0.478** [0.154]
Trust in Investment Algorithms	4.02 (2.25)	4.04 (2.45)	−0.024 [0.153]
Frequent Investments	3.54 (3.31)	6.69 (3.25)	−3.149** [0.215]
Responsibility in Financial Matters	5.60 (3.67)	6.85 (3.33)	−1.249** [0.231]
Use Expertise of Finance Professionals	3.58 (3.19)	2.21 (2.81)	1.376** [0.214]
Delegate to Finance Professionals	3.36 (3.11)	1.32 (2.16)	2.039** [0.192]
Delegate to Investment Algorithms	1.71 (2.37)	0.85 (1.67)	0.865** [0.147]
Observations	550	408	958

Notes: All items, except for “Altruism,” were answered on Likert scales ranging from 0 (minimum) to 10 (maximum). The variable “Altruism” refers to the amount transferred (up to 10,000 SEK) in a hypothetical charitable giving setting. For reasons of comparison, the variable is re-scaled to thousands SEK.

Table E6: Descriptive statistics and comparisons between pools for numeracy, financial literacy, and overconfidence. This table reports the means and standard deviations (in parentheses) for participants' numeracy and financial literacy scores, their self-estimates regarding their numeracy scores (in terms of estimates of the score and their relative performance compared to the Swedish general population), and the two measures of overconfidence (overestimation and overplacement), separated for the general population and the finance professional subject pool. The column "*t*-test" reports the differences in means and the *t*-values (in brackets) from two-sample *t*-tests based on $n = 958$. * $p < 0.05$, ** $p < 0.005$.

	<i>Gen. Pop.</i>	<i>Fin. Prof.</i>	<i>t-Test</i>
<i>Skills:</i>			
<i>Numeracy Score</i>	4.44 (1.63)	5.31 (1.59)	−0.865** [0.106]
<i>Financial Literacy Score</i>	4.29 (1.20)	5.39 (0.94)	−1.099** [0.071]

Notes: Overestimation refers to the difference between participants' estimate of their numeracy and their actual numeracy score. Overplacement refers to the (minimum) difference between participants' estimate of the decile, their performance in the numeracy task belongs to, and the percentiles of the numeracy scores evaluated based on a normal distribution (see Appendix D for further details).

Figure E1: Self-reported risk tolerance. This figure depicts the mean levels of risk tolerance (self-reported on scales from 1 to 10) in general and in financial matters, respectively, separated for the general population sample and the finance professionals sample. p -values reported above the bars indicate comparisons between the subject pools and are based on two-sample t -tests ($n = 958$). p -values reported in the boxes are based on paired-sample t -tests with sample sizes of 550 (general population) and 408 (finance professionals), respectively.

Figure E2: Self-reported trust. This figure depicts the mean levels of trust (self-reported on scales from 1 to 10) in the general population (T_{GP}), in finance professionals (T_{FP}), and in investment algorithm (T_{RA}), respectively, separated for the general population sample and the finance professionals sample. p -values reported above the bars indicate comparisons between the subject pools and are based on two-sample t -tests ($n = 958$). p -values reported in the boxes are based on paired-sample t -tests with sample sizes of 550 (general population) and 408 (finance professionals), respectively.

Figure E3: Numeracy and financial literacy scores. This figure shows empirical cumulative density distributions (based on Gaussian kernels with a bandwidth of 1) for participants' numeracy and financial literacy scores, respectively. Kolmogorov-Smirnov-tests are reported in the light gray boxes ($n = 958$).

Figure E4: Overestimation and overplacement. This figure shows empirical cumulative density distributions (based on Gaussian kernels with a bandwidth of 1 in the left panel and 0.2 in the right panel) for the estimates of participants' overestimation and overplacement measures, respectively. Kolmogorov-Smirnov-tests are reported in the light gray boxes ($n = 958$).

F. Descriptives and Analyses of Time Spent

In the following, we examine the time spent per experimental task in the online experiment. Throughout the analysis, we truncate the time spent per task at the 99% percentile to avoid that outliers distort the results. In particular, for each task, durations exceeding this threshold are replaced by the value of the 99% percentile. Descriptive statistics of the time spent per task, separated for the general population and the finance professionals subject pools, are presented in Table F1. On average, the times spent in the experimental tasks appear to be sufficiently long to be confident that participants in both samples took the experiment very seriously, which is also confirmed by the high levels of decision-making quality (see Appendix C for details). Differences in the time spent between the two pools are reported in Table F2.

With respect to the main task, we examine learning effects by means of ordinary least squares regressions of the time spent on the 25 decisions on a linear time trend (with standard errors clustered at the subject level). The regressions reveal that the time spent per decision decreases with the progressing round numbers, in the decisions with both two and five assets, respectively. For the first two-asset item, participants from the general population take, on average, 57.1 seconds; for the subsequent decisions, the time spent, on average, decreases by 5.1 seconds per item ($t(548) = 13.916$, $p < 0.001$, $n = 5,500$). Finance professionals take, on average, 72.7 seconds for the first two-asset decision; for the following nine decisions with two assets, the time spent, on average, decreases by 6.5 seconds per item ($t(406) = 8.776$, $p < 0.001$, $n = 6,120$). Likewise, learning is observed for consecutive investment decisions with five assets. For the first five-asset item, participants from the general population take, on average, 3.5 minutes; for the subsequent decisions, the time spent, on average, decreases by 13.1 seconds per item ($t(548) = 2.065$, $p = 0.039$, $n = 5,500$). Finance professionals take, on average, 2.6 minutes for the first five-asset decision; for the following fourteen decisions with five assets, the time spent, on average, decreases by 7.7 seconds per item ($t(406) = 2.844$, $p = 0.005$, $n = 6,120$).

In addition, we investigate whether decision-making quality is systematically affected by time participants take to decide on the 25 investment decisions. Notably, ordinary least squares regression of the decision-making quality index ($DMQI$) on the time spent on the investment task (i.e., the sum of the time spent in the investment task with two and five assets) reveal that subjects' proneness to poor investment decisions is not significantly driven by the time they spend on each decision, neither in the general population sample ($\beta = 0.005$, $t(548) = 1.314$, $p = 0.189$, $n = 550$), nor in the finance professionals sample ($\beta = 0.003$, $t(406) = 1.261$, $p = 0.208$, $n = 408$).

Table F1: Descriptive statistics of time spent per task. This table reports the means and standard deviations (in parentheses) as well as the median and interquartile ranges (*IQR*; in brackets) for the time spent per experimental task (measured in minutes), separated for the general population sample (all treatments) as well as the three treatments conducted among finance professionals.

	<i>GP-*</i>		<i>FP-FIXED</i>		<i>FP-ALIGNED</i>		<i>FP-OWN</i>	
	<i>m / sd</i>	<i>q50 / iqr</i>	<i>m / sd</i>	<i>q50 / iqr</i>	<i>m / sd</i>	<i>q50 / iqr</i>	<i>m / sd</i>	<i>q50 / iqr</i>
Investment Task w/ Two Assets	5.53 (4.17)	4.30 [3.10]	7.00 (5.37)	5.13 [4.56]	7.26 (5.23)	5.07 [5.98]	5.16 (3.94)	4.25 [3.08]
Investment Task w/ Five Assets	15.24 (10.73)	11.99 [10.52]	18.40 (14.85)	13.57 [14.41]	19.11 (15.44)	13.73 [14.02]	15.65 (12.07)	11.77 [10.30]
Questionnaires (Self-Reported)	2.67 (1.54)	2.30 [1.20]	2.49 (1.29)	2.18 [1.01]	2.55 (1.53)	2.20 [1.08]	2.52 (1.16)	2.15 [1.27]
Numeracy Inventory (8 Items)	3.65 (0.55)	4.00 [0.62]	3.68 (0.57)	4.00 [0.58]	3.58 (0.64)	4.00 [0.92]	3.60 (0.62)	4.00 [0.77]
Financial Literacy Test (6 Items)	2.05 (0.56)	1.99 [0.85]	1.74 (0.59)	1.63 [0.77]	1.76 (0.52)	1.63 [0.72]	1.77 (0.60)	1.65 [0.90]
Observations	550		132		139		137	

Table F2: Differences in time spent. This table reports the *t*-statistics from two-sample *t*-tests between the general population sample (pooled across all treatments) and the finance professionals sample separated for the treatment conditions for the time spent per experimental task (measured in minutes). Standard errors (*se*) are reported in parentheses. Means, standard deviations, medians, and interquartile ranges for the time spent per experimental task in all treatments are reported in Table F1. * $p < 0.05$, ** $p < 0.005$.

	<i>GP-*</i> vs. <i>FP-FIXED</i>	<i>GP-*</i> vs. <i>FP-ALIGNED</i>	<i>GP-*</i> vs. <i>FP-OWN</i>	<i>FP-FIXED</i> vs. <i>FP-ALIGNED</i>	<i>FP-FIXED</i> vs. <i>FP-OWN</i>	<i>FP-ALIGNED</i> vs. <i>FP-OWN</i>
	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>	<i>t / se</i>
Investment Task w/ Two Assets	-2.447* (0.558)	-2.975** (0.544)	0.988 (0.483)	-0.390 (0.644)	3.218** (0.573)	3.754** (0.558)
Investment Task w/ Five Assets	-1.218 (1.509)	-1.676 (1.523)	0.676 (1.353)	-0.388 (1.842)	1.670 (1.648)	2.076* (1.670)
Obs.	315	322	320	271	269	276

G. Supplementary Results

Table G1: Willingness to pay for delegation. This table reports estimates from ordinary least squares regressions of participants' willingness to pay for delegating their decisions to the agent on treatment indicators, a set of experimental measures, and self-reported measures. Robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.005$.

	(1)	(2)	(3)	(4)
<i>Treatment Indicators:</i>				
<i>GP-ALIGNED</i>	2.043 (3.252)	2.052 (3.475)	2.091 (3.205)	1.920 (3.334)
<i>GP-ALGO</i>	3.492 (3.076)	3.091 (3.242)	3.284 (3.024)	2.713 (3.148)
<i>Experimental Measures:</i>				
Decision Making Quality Index			−0.377 (0.774)	−0.422 (0.790)
Financial Literacy Score (Std.)			3.217 (2.294)	3.192 (2.295)
Numeracy Score (Std.)			−3.598 (3.102)	−3.618 (3.150)
Overestimation (Std.)			−0.814 (1.634)	−0.769 (1.641)
Overplacement (Std.)			−0.134 (1.991)	−0.223 (1.992)
<i>Self-Reported Measures:</i>				
Risk Tolerance (Std.)			−0.262 (1.454)	−0.344 (1.497)
Blame Shifting (Std.)			0.340 (1.294)	0.471 (1.340)
Trust in Agent (Std.)			1.907 (1.544)	2.191 (1.565)
<i>Constant:</i>				
<i>GP-FIXED</i>	31.290** (2.482)	29.429** (7.994)	26.793** (4.233)	24.086* (8.752)
<i>Controls</i>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>yes</i>
<i>F</i>	0.648	0.313	0.706	0.565
<i>p > F</i>	0.524	0.947	0.718	0.897
Adj. <i>R</i> ²	−0.005	−0.033	−0.023	−0.050
Observations	148	148	148	148

Notes: Treatment *GP-FIXED* serves as reference condition. All self-reported measures are standardized scores. “Trust in Agent” refers to a combined variable of trust in finance professionals and financial algorithms, conditional of the treatment. “Blame Shifting” refers to the mean of two standardized survey items on shifting blame on others and resisting the temptation to shift blame on others. “Controls” include gender (binary indicator for female), age (in years), net income from major employment in 2017 (in thousand SEK’s), and maximum education level (dichotomous indicators for high school education or less, university education smaller or equal to three years, and university education larger than three years).

Table G2: Decision-making quality by subject pool. This table reports estimates from ordinary least squares regressions of the decision-making quality index (*DMQI*) on an indicator variable for finance professionals (*FP-OWN*), the mean standard deviation across the 25 decisions in the allocation decision task (*SD*), and the interaction of *FP-OWN* and *SD*. Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.005$.

	(1)	(2)	(3)
<i>Pool Indicator:</i>			
<i>FP-OWN</i>	0.320** (0.109)	0.297* (0.111)	-0.051 (0.227)
<i>Moderator Variable:</i>			
<i>SD</i>		0.001 (0.001)	0.000 (0.001)
<i>Interaction Effect:</i>			
<i>FP-OWN</i> $\times$ <i>SD</i>			0.004 (0.002)
<i>Constant:</i>			
<i>GP-*</i>	-0.077 (0.049)	-0.161 (0.094)	-0.058 (0.110)
<i>F</i>	8.655	4.885	4.300
$p > F$	0.003	0.008	0.005
Adj. R^2	0.011	0.011	0.014
Observations	687	687	687

Notes: The sample for the regressions includes all participants deciding on their own behalf, i.e., all participants from all general population treatments (*GP-**) and finance professionals in treatment *FP-OWN*.

Table G3: Finance professionals' decision-making quality by treatments.

This table reports estimates from ordinary least squares regressions of the decision-making quality index (*DMQI*) on indicator variable for the treatments *FP-ALIGNED* and *FP-OWN*, the mean standard deviation across the 25 decisions in the allocation decision task (*SD*), and the interaction of *FP-OWN* and *SD*. Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.005$.

	(1)	(2)	(3)
<i>Treatment Indicators:</i>			
<i>FP-ALIGNED</i>	0.269* (0.133)	0.263* (0.132)	0.358 (0.245)
<i>FP-OWN</i>	0.356* (0.134)	0.315* (0.134)	0.230 (0.257)
<i>Moderator Variable:</i>			
<i>SD</i>		0.003* (0.001)	0.003 (0.002)
<i>Interaction Effects:</i>			
<i>FP-ALIGNED</i> $\times$ <i>SD</i>			-0.001 (0.003)
<i>FP-OWN</i> $\times$ <i>SD</i>			0.001 (0.003)
<i>Constant:</i>			
<i>FP-FIXED</i>	-0.112 (0.096)	-0.333* (0.125)	-0.340 (0.174)
<i>F</i>	3.824	5.009	3.130
$p > F$	0.023	0.002	0.009
Adj. R^2	0.014	0.029	0.026
Observations	408	408	408

Table G4: Finance professionals' risk taking when deciding on behalf of clients. This table reports estimates from ordinary least squares regressions of the mean standard deviation across the 25 items of the investment task (*SD*) on an indicator variable for the treatment *FP-ALIGNED*, indicators for the risk level professionals are asked to take into consideration when deciding on clients' behalf, a set of experimental measures, and self-reported measures. Robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.005$.

	(1)	(2)	(3)	(4)
<i>Treatment Indicator:</i>				
<i>FP-ALIGNED</i>	2.387 (4.403)	2.284 (4.351)	2.360 (4.447)	2.671 (4.412)
<i>Given Risk Level:</i>				
Risk Level=2	18.951** (6.323)	18.572** (6.316)	19.722** (6.236)	19.241** (6.241)
Risk Level=3	38.613** (5.582)	37.865** (5.827)	39.499** (5.598)	39.502** (5.818)
Risk Level=4	94.786** (6.609)	93.585** (6.544)	95.742** (6.632)	95.082** (6.585)
<i>Experimental Measures:</i>				
Decision Making Quality Index			2.465 (2.134)	2.500 (1.930)
Financial Literacy Score (Std.)			9.341 (4.824)	6.044 (5.146)
Numeracy Score (Std.)			-2.299 (5.311)	-5.122 (5.518)
Overestimation (Std.)			-5.235* (2.605)	-5.824* (2.571)
Overplacement (Std.)			2.239 (3.243)	1.385 (3.353)
<i>Self-Reported Measures:</i>				
Risk Tolerance (Std.)			1.945 (2.630)	2.875 (2.761)
Blame Shifting (Std.)			-1.814 (2.580)	-2.401 (2.565)
<i>Constant:</i>				
<i>FP-FIXED</i>	40.894** (4.779)	33.745* (15.283)	22.090* (10.186)	15.431 (19.686)
<i>Controls</i>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>yes</i>
<i>F</i>	56.676	33.283	23.477	19.787
<i>p > F</i>	0.000	0.000	0.000	0.000
Adj. <i>R</i> ²	0.486	0.498	0.497	0.507
Observations	271	271	271	271

Notes: Treatment *GP-FIXED* serves as reference condition. All self-reported measures are standardized scores. "Blame Shifting" refers to the mean of two standardized survey items on shifting blame on others and resisting the temptation to shift blame on others. "Controls" include gender (binary indicator for female), age (in years), net income from major employment in 2017 (in thousand SEK's), and maximum education level (dichotomous indicators for high school education or less, university education smaller or equal to three years, and university education larger than three years).

Figure G1: Portfolio risk of clients' asking the agent to take more risk when delegating. This figure shows the average portfolio risk across the 25 investment decisions (SD; normalized to 1) of those clients that choose to delegate *and* ask the agent to take more risk than they think they took in their own decision, separated for the treatments *GP-FIXED*, *GP-ALIGNED*, and *GP-ALGO* (red dots). The blue dots indicate the mean portfolio risk across the 25 investment decisions (SD) of agents that serve as potential matching partners, i.e., those in the corresponding treatment deciding for clients with the risk level that matches their desired risk level when delegating. Error bars indicate standard errors of the mean (SEM) and are clustered on the individual level for agents. Asterisks indicate significant differences per principal-agent level and are based on two-sample *t*-tests (with clustered standard errors); * $p < 0.05$, ** $p < 0.005$. Aggregate comparisons between clients' and agents' portfolio risk per treatment are reported in the gray boxes. *t*-statistics are based on ordinary least squares regressions of portfolio risk on an indicator variable for "agent", controlling for risk level indicators, with standard errors being clustered on the individual level.

	<i>RL-1</i>	<i>RL-2</i>	<i>RL-3</i>	<i>RL-4</i>	<i>HHI</i>
0.00 - 0.10	0.786	0.179	0.036	0.000	0.651
0.10 - 0.20	0.529	0.324	0.118	0.029	0.400
0.20 - 0.30	0.382	0.382	0.206	0.029	0.336
0.30 - 0.40	0.250	0.271	0.396	0.083	0.299
0.40 - 0.50	0.162	0.270	0.405	0.162	0.290
0.50 - 0.60	0.107	0.071	0.500	0.321	0.370
0.60 - 0.70	0.000	0.118	0.294	0.588	0.446
0.70 - 0.80	0.100	0.200	0.200	0.500	0.340
0.80 - 0.90	0.000	0.111	0.111	0.778	0.630
0.90 - 1.00	0.095	0.048	0.000	0.857	0.746

Figure G2: Number of portfolios with similar portfolio risk across risk levels.

This figure shows the fraction of finance professionals' portfolios (when deciding on behalf of principals, i.e., in treatments *FP-FIXED* and *FP-ALIGNED*) across equally-sized classes of portfolio risk (normalized to 1) over the four risk levels. The color coding increases with the cell's magnitude. The column *HHI* refers to the Herfindahl-Hirschman-Index, a diversity index defined as $HHI = \sum_k s_k^2$ with s_k denoting the share in risk level $k = \{1, 2, 3, 4\}$. *HHI* takes a minimum value of 0.25 (if $s_k = 0.25 \forall k$) and a maximum value of 1 (if $s_k = 1 \forall k$).