Can Simulated Experience Be Harnessed to Help People Make Investment Decisions?

Tomás Lejarraga[1,2], Kavitha Ranganathan[3], and Dirk U. Wulff[2,4]

[1]Universitat de les Illes Balears

[2]Max Planck Institute for Human Development

[3]T. A. Pai Management Institute

[4]University of Basel

Author Note

Tomás Lejarraga https://orcid.org/0000-0001-6135-2243 Kavitha Ranganathan https://orcid.org/0000-0002-6877-9079 Dirk U. Wulff https://orcid.org/0000-0002-4008-8022

We thank Christine Laudenbach for her helpful comments and for endorsing the implementation of the risk communication tools in our studies. We also thank Ralph Hertwig for his comments, and Laura Wiles for editing the manuscript. This work was supported by a grant from the Biäsch Foundation for the Advancement of Applied Psychology to Dirk U. Wulff (2019–23).

Correspondence concerning this article should be addressed to Tomás Lejarraga, Departament d'Economia de l'Empresa, Universitat de les Illes Balears, Palma, Spain, E-mail: tomas.lejarraga@uib.eu

Abstract

To make profitable investment decisions, investors must know and understand their risks. They can learn about these risks in different ways. Evidence suggests that investors who learn from a "risk tool" simulator perceive financial risk more accurately, feel more informed and confident, and thus take on more financial risk. We attempt a conceptual replication of these findings, exploring whether they extend from kind to wicked environments and to investors with some investment experience. We conducted three studies online, amounting to 3,455 participants, and observe that neither the risk tool nor any of the other risk communication interventions lead to different risk taking or to different subjective representations of risk relative to a control condition in which participants received no intervention.

Keywords: risk tool, risk simulator, decisions from experience

Can Simulated Experience Be Harnessed to Help People Make Investment Decisions?

Introduction

To make profitable investment decisions, investors must know and understand the risks they face. Investors' comprehension of risk can differ considerably depending on how that risk is learned or communicated to them. Finance professionals typically communicate risks using descriptions—for instance, in the form of financial reports, mutual fund brochures, insurance brochures, investor education programs, and market research reports. These documents describe risks using a summary of historical returns and their respective probabilities. Investors could also acquire knowledge about financial risks in other ways; namely, by observing the development of stock prices or through their own investment experience. For instance, day traders decide to purchase a stock by simply observing prior stock movements, or individual investors increase their subscription to initial public offering (IPO) auctions subsequent to previous successful experience (e.g., Kaustia & Knüpfer, 2008). Indeed, the way in which knowledge about risks is acquired has a dramatic influence on investors' understanding of risk and their willingness to accept it. For example, our previous work shows experimentally that people who learn about a stock market crash from experience are more likely to stay out of the market than people who learn about the same crash from descriptions (i.e., the "depression babies" effect), even when wealth effects are kept constant (Lejarraga et al., 2016).

In general, people seem not only to be persuaded more by their experience than by described sources of information, but to have more accurate subjective assessments of risks, even if experience is simulated (Hertwig & Wulff, 2022). Kaufmann et al. (2013) and Bradbury et al. (2015) show that simulations of the stock market, which are closer to the notion of witnessing rather than experiencing through action, help investors understand risk accurately, and lead them to invest more in the stock market than investors who learn from described sources. On the other hand, Lejarraga and Gonzalez (2011) show that

exhaustive descriptive information is often neglected when the participants have the chance to also experience the information by sampling; that is, people who have experience and descriptions make choices as though descriptions had been omitted. Consistent with this observation, Barron et al. (2008) show that people are more likely to ignore a described warning if they have already experienced a series of safe outcomes, but they are more likely to be persuaded by the warning if they have no previous experience. These studies converge in that people tend to overweight the information they gain from experience over that which is provided in a descriptive form; this can lead to more accurate risk assessments, as long as the experiences are representative of the environment (for reviews, see Hertwig & Wulff, 2022; Wulff et al., 2018).

We build on previous research in two ways. First, we conduct a conceptual replication of the study undertaken by Kaufmann et al. (2013). This involves exposing experimental investors to four different risk communication interventions and observing how these interventions affect participants' risk taking, factual knowledge about the encountered decision environment, and subjective assessments of confidence and satisfaction. Second, we extend the study conducted by Kaufmann et al. (2013) by manipulating whether or not investors had prior experience of the decision environment before they were exposed to the risk communication intervention. Specifically, and akin to Barron et al. (2008), one group of people build up experience with the decision environment by making a series of decisions with feedback prior to being exposed to the risk communication intervention whereas the other group, akin to Kaufmann et al. (2013), are given the risk communication intervention straight away.

Using this setup, we will be able to provide a better estimate of how effective the four risk interventions are in informing people about financial risks in situations involving no prior experience and to explore how and whether prior experience impacts the effectiveness of the four interventions.

Overview of Studies

To study the relative effectiveness of the risk communication interventions and how prior experience interacts with them, we conducted two preregistered experimental studies[1]. Both studies had the same between-subjects design with nine conditions comprising one control condition and eight treatment conditions that result from crossing two factors: (a) prior experience with two levels (with and without) and (b) the intervention used to learn about the options, with four levels (description, experience, distribution, and the "risk tool") (Figure 1). Participants played an investment game in which their task was to allocate funds between a risky and a safe investment option for a number of periods. The risky option was a stock offering a variable rate of return and the safe option was a deposit offering a fixed rate of return.

In Study 1, the stock had a higher expectation value than the deposit. In Study 2, the stock involved the possibility of a large but rare loss. Thus, in terms of expectation, the stock was less attractive than the deposit. Because rare events are unlikely to be encountered in a small sample of experiences, the setting in Study 2 can be conceptualized as a "wicked" investment environment, and Study 1 as a "kind" environment (Hogarth, 2001; Hogarth et al., 2015).

Prior experience

Participants who were assigned to a *prior-experience* condition were endowed with a portfolio of £100 and were asked to make investment decisions in 20 periods. Initially, participants knew that one option was a stock and the other a deposit, but were not informed about the return distributions of these options. Immediately after each investment, the obtained return was automatically added to their running portfolio, providing some feedback to participants about the options. After the initial 20 investment periods, participants in the eight treatment conditions (but not participants in the control

[1] Study 1 was preregistered in https://https://osf.io/pv2hu and Study 2 in https://osf.io/ekzma

condition) were presented with one of the four communication tools and were allowed to use the tool to explore the investment options for as long as they wanted. After a required minimum exploration, participants returned to the investment game to continue investing for another 20 periods.

Participants who were assigned to a *no-prior-experience* condition entered the experiment without any prior experience; namely, they started the investment game by using one of the communication tools directly. After exploring the options with one tool, participants began the investment game for 20 investment periods. To keep wealth constant between the prior-experience and the no-prior-experience conditions, participants in the latter were yoked to participants in the former: We recorded all the portfolio amounts of participants in the prior-experience conditions after period 20 and used them as starting portfolios for participants in the no-prior-experience condition. Therefore, portfolios in the prior and no-prior-experience conditions were constant at the start of period 21.

After participants finished playing the investment game, they completed a task survey including the following questions (Kaufmann et al., 2013):

- How risky do you perceive the stock (the risky asset) to be? (1 = *not risky at all*, 7 = *very risky*)
- How confident do you feel about investing in the risky asset? (1 = *completely unconfident*, 7 = *completely confident*)
- If we put £100 in the risky asset, what is the expected return of the £100 after five years? (Give your best estimate.)
- If we put £100 in the risky asset, in how many out of 100 cases will the return fall below £100 after five years?
- If we put £100 in the risky asset, in how many out of 100 cases will the return fall be above £150 after five years?

- How informed do you feel about the two assets (the deposit and the stock)? (1 = *completely uninformed*, 7 = *completely informed*)

After the survey, participants were shown their final account balance and were asked: How satisfied are you with your return? (1 = *completely unsatisfied*, 7 = *completely satisfied*).

Participants then completed the Berlin Numeracy Test (Cokely et al., 2012), a survey of financial behavior (reference, year), an investment quiz (Cohn et al., 2015), and a question about general propensity to take risks (Goebel et al., 2019, SOEP). Finally, participants completed a demographic survey, including questions about their income and wealth.

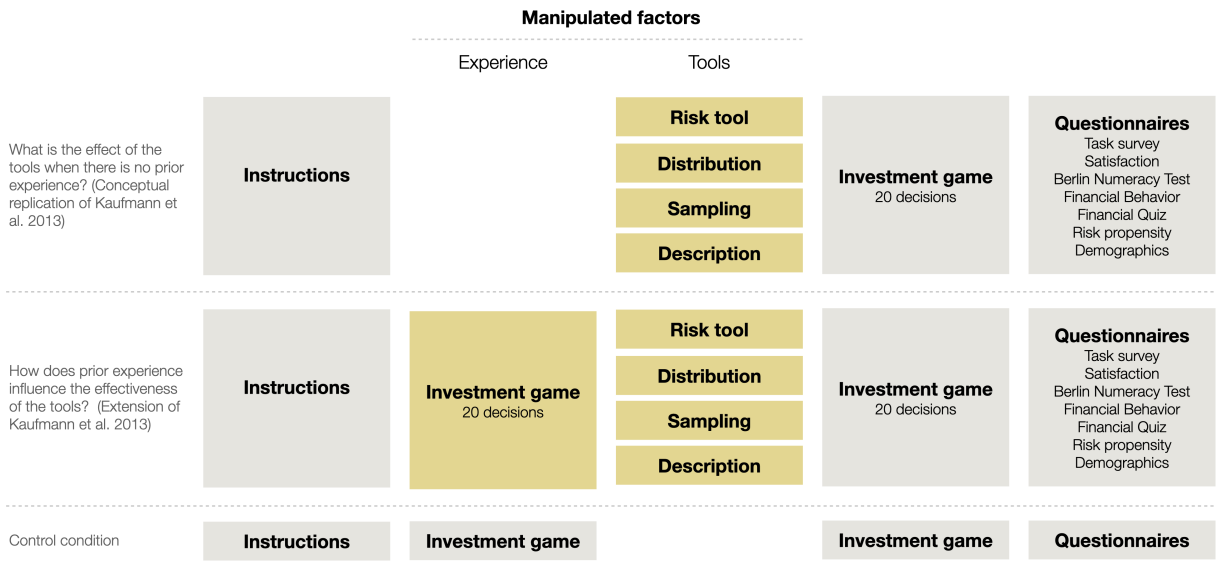

Figure 1

Experimental design and procedure. Yellow boxes indicate manipulated factors.

Risk Communication Tools

The *description* tool describes the options in full, consistent with Kaufmann et al. (2013). For example, in Study 1, participants read "The deposit is a safe asset. It has a

guaranteed return of 0.83% for sure. If for 20 periods you invest the full £100 in the deposit, you will have a return of £117.91. The stock is a risky asset. It has an expected return of 2.16% with a standard deviation of 7.42%. If you invest the full £100 in the stock, you will have an expected final outcome of £153.30". In Study 2, the description of the deposit did not change, but the description of the stock was "... It has an expected return of 0.72% with a standard deviation of 12.53%. If for 20 periods you invest the full £100 in the stock, you will have an expected final outcome of £115.43." The description tool also allows participants to see the expected return of a specific investment allocation and its corresponding 70% and 95% confidence intervals in relative frequencies. For example, in Study 2, if participants distributed £100 equally between options, they received the following message: "In 70 out of 100 cases your return will be between £97.35 and £105.29 and in 95 out of 100 cases between £91.96 and £108.74." Participants could sample different allocations using a slider that determined the proportion of funds allocated to each option. Participants were forced to sample initially an investment where all funds were allocated to the safe deposit, then an allocation where all funds were allocated to the risky stock, and finally an allocation involving any mix of stock and deposit. After these three forced samples, participants were allowed to use the tool for as long as they wished, sampling as many different allocations as they wanted.

The *risk tool* was programmed following Kaufmann et al. (2013). Participants could choose an investment allocation and see the outcome of their decision plotted on a histogram. They could simulate as many outcomes as they wanted using different simulation modes. They could simulate one outcome at a time, or they could simulate outcomes automatically, in either slow or fast motion mode. As outcomes accumulate in the histogram, the graph becomes increasingly representative of the underlying distribution that generates the outcomes. As in the description tool, participants were initially forced to sample—using the three modes—a fully safe investment, a fully risky investment, and finally any mix of safe and risky investments of their choice. Only after these three forced

samples did the risk tool allow them to sample as many different investment allocations as they wanted before they could return to the investment game.

The *distribution* tool showed the distribution of potential returns of participants' investments by plotting the density function in a graph. Participants could change their investment allocation and observe how the distribution of potential returns changed. As in the other tools, participants were initially forced to sample a fully safe, a fully risky, and a mixed allocation. Only then did the distribution tool allow them to sample as many different investment allocations as they wanted before they were allowed to return to the investment game.

The *experience* tool allowed participants to sample outcomes from their investment allocations. In contrast to the risk tool, in which outcomes were plotted in a graph, the experience tool showed a single numerical outcome for each allocation. Participants could sample as many outcomes as they wished for a particular allocation, and they could also explore outcomes for a different allocation. As in the other tools, participants were initially forced to sample a fully safe, a fully risky, and a mixed allocation before they were allowed to return to the investment game.

The four tools were designed to be consistent with the tools used in Kaufmann et al. (2013), and they were reviewed and endorsed by the first author of that article. Our implementations are available in (TO BE ADDED).

Investment options

In both studies, the risky option was a stock offering a variable rate of return, and the safe option was a deposit offering a fixed rate of return. The deposit was constant in Study 1 and 2. To keep our study consistent with Kaufmann et al. (2013), we transformed their annual rate of 3.35% into quarterly rates of 0.83%.

The stock differed in Study 1 and 2. Specifically, in Study 1, the stock offered 2.16% (SD of 7.42%) and was computed as the quarterly rate of return that was equivalent to the

annual rate used by Kaufmann et al. (2013) (8.95%)—which, in turn, was the average annual return from the Morgan Stanley Capital International (MSCI) USA Index between 1973 and 2008. Thus, in Study 1, the stock was more attractive in expectation than the deposit.

In Study 2, we introduced the possibility of a large but rare loss. Thus, the stock offered a draw from the same distribution of Study 1 (2.16%, $SD = 7.42\%$) with a probability of 0.98, and with 0.02 probability, a draw from a rare negative event distribution resulting in a loss of 70% ($SD = 7.42\%$). Thus, the overall expected return of the stock was 0.72% ($SD = 12.53\%$), making the deposit more attractive than the stock.

Participants

We followed the "2.5 rule" to plan our study samples (Simonsohn, 2015). Specifically, our planned samples were 2.5 times the average sample size per condition in Study 2 and 3 of Kaufmann et al. (2013), which was 69. The resulting sample for each of our nine conditions is 172.5, which we rounded to 173.

Data were collected online using Prolific (https://www.prolific.co/). Eligible participants were UK residents, had learned English as their first language, and were aged 18 or older at the time of data collection. Participants were compensated with a participation fee (£1.25 in Study 1 and £1.40 in Study 2) and a bonus that depended on their investment decisions in the experiment and was on average £0.42 and £0.27 in Study 1 and 2, respectively. On average, participants received £1.62 in Study 1 and £1.7 in Study 2 for approximately 15 minutes of participation.

Overall, there were 1,745 participants in Study 1 and 1,773 in Study 2. We used six attention checks to filter out inattentive participants. Those who failed two checks were excluded from the sample, but results are robust to more stringent exclusion criteria (see Additional Analyses). The final samples analyzed involved 1,560 and 1,546 participants in Study 1 and 2, respectively.

Results

Figure 2 shows the proportion of funds invested in the stock across conditions and studies. To compare investments across conditions, we compute the mean proportion of funds allocated to the stock across periods for each individual. We then report means and 95% confidence intervals across individuals. Figure 4 shows the distribution of responses to all knowledge questions.

Impact of Risk Communication Tools on Investments

Investments in a Kind Environment

Investments in Study 1 did not differ according to the tool used. The mean proportion of funds invested in the stock was 0.53 (95% CI: 0.50, 0.57) for those who used the risk tool, 0.55 (0.52, 0.58) for those who used the distribution tool, 0.5 (0.47, 0.54) for those who used the sampling tool, and 0.51 (0.47, 0.54) for those who used the description tool. None of these conditions differed from the control condition, in which participants used no tool and invested on average 0.52 (0.49, 0.55) in the stock. Compared with previous results (Kaufmann et al., 2013), participants who used the risk tool (0.53, 0.50, 0.57) did not take more risks than participants who used other tools (0.52, 0.5, 0.54)

Investments were also unchanged by previous experience. Participants who had no previous experience invested a mean of 0.52 (0.51, 0.54) in the stock, the same as the 0.52 (0.5, 0.53) by participants who had already invested for 20 periods beforehand. Specifically, after prior experience, participants who used the risk tool invested 0.52 (0.49, 0.55) in the stock. Similar levels of risk were taken by participants in the other conditions. Those who used the distribution tool invested 0.51 (0.48, 0.55), those who used the sampling tool invested 0.53 (0.5, 0.56), and those who used the description tool invested 0.5 (0.47, 0.54) in the stock. Similarly, having experienced 20 investment periods did not make one tool better or worse than another.

Investments in a Wicked Environment

A similar pattern of results emerged in the wicked investment environment of Study 2. Again, investments did not differ across different tools. The mean proportion of funds invested in the stock was 0.56 (0.52, 0.6) for those who used the risk tool, 0.53 (0.49, 0.56) for those who used the distribution tool, 0.52 (0.48, 0.55) for those who used the sampling tool, and 0.47 (0.43, 0.51) for those who used the description tool. Again, none of these conditions differed markedly from the control condition, in which participants invested on average 0.53 (0.5, 0.57) in the stock. Furthermore, participants who used the risk tool (0.56, 0.52, 0.6) did not take more risks than participants who used other tools (0.5, 0.48, 0.53). Here again, we do not replicate results reported by Kaufmann et al. (2013)

After prior experience, participants who used the risk tool invested 0.48 (0.45, 0.52) in the stock. Those who used the distribution tool invested 0.49 (0.46, 0.53), those who used the sampling tool invested 0.54 (0.5, 0.58), and those who used the description tool invested 0.4 (0.36, 0.45) in the stock.

In contrast to the kind environment of Study 1, participants who had prior experience took less risk (0.48, 0.46, 0.5) than those who had no previous experience (0.52, 0.51, 0.54). In this wicked environment, more experience increases the chances that participants learn about the possibility of a large loss. Results suggest that participants with prior experience became more cautious. However, having had prior experience did not systematically impact the effect of tools. One possible exception is the description tool. Participants who had prior experience and used the description tool took the least financial risk (0.4, 0.36, 0.45) compared to the rest of the tools (0.5, 0.48, 0.53). Overall, the description tool led to the most cautious investments by participants who both did, and did not, have prior experience.

In sum, compared to previous findings, we observe no systematic effect of risk communication tools across Study 1 and 2, except for the description tool being more effective at revealing the risk of a large but rare loss. Similarly, experience did not influence

risk taking except in revealing the possibility of a rare event.

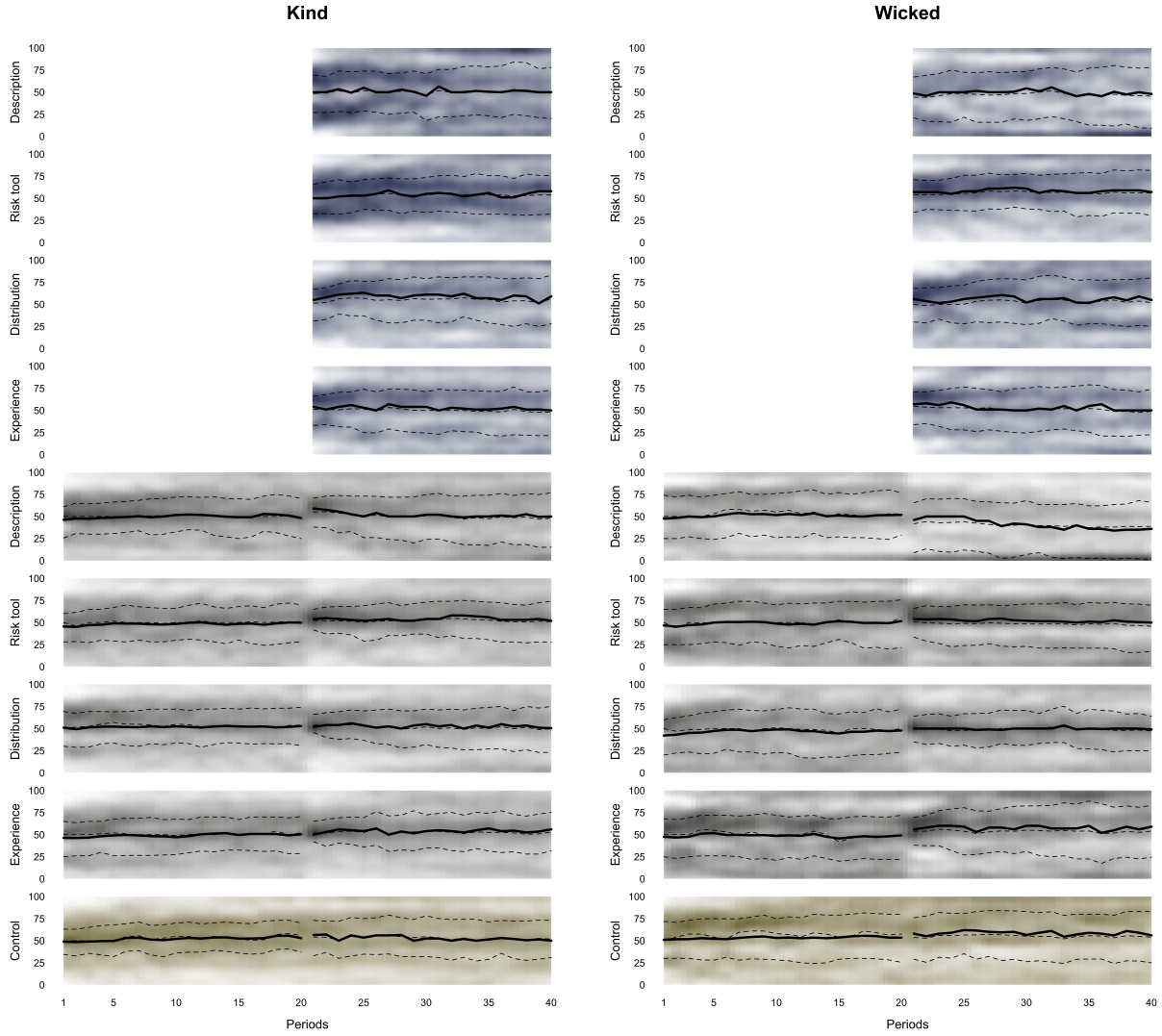

Figure 2

Proportion of funds invested in the risky option across periods, by condition and study. Gray plots indicate the conditions with prior experience. Blue plots indicate the conditions without prior experience. Gold plots show the control conditions. The thick line in each plot denotes the mean proportion of funds invested in the risky stock. The dashed lines indicate the 95% confidence intervals of the mean. Darker regions in the plot indicate higher density than lighter regions.

Impact of Risk Communication Tools on Subjective Assessments of Risk

Assessments in a Kind Environment

In previous work, participants who used the risk tool perceived less risk than those who used other tools. In our study, we observed no effect of tools on risk perception. Participants who used the risk tool scored 4.5 (4.29, 4.71) on risk perception, no different to participants who used the description tool (4.55, 4.33, 4.78), the distribution tool (4.65, 4.44, 4.86), and the sampling tool (4.51, 4.30, 4.73).

Similarly, previous findings indicated that the risk tool led to the highest feeling of being informed; however, this observation was not replicated in our studies. Participants who used the risk tool (3.33, 3.12, 3.55) did not feel more or less informed than participants who used other tools. Participants who used the description tool scored 3.60 (3.37, 3.82), those who used the distribution tool scored 3.21 (2.98, 3.44), and those who used the sampling tool scored 3.40 (3.17, 3.62).

In terms of confidence, the tools had no differential impact. The risk tool led to a mean confidence score of 3.98 (3.75, 4.20), the description tool to 4.01 (3.76, 4.26), the distribution tool to 3.92 (3.70, 4.15), and the sampling tool to 4.10 (3.87, 4.32). Tools also had no differential impact on satisfaction. The risk tool led to a mean satisfaction score of 5.24 (5.03, 5.46), the description tool to 5.19 (4.99, 5.39), the distribution tool to 4.80 (4.56, 5.04), and the sampling tool to 5.18 (4.97, 5.39).

With respect to the perceived chances of ending with a positive return, tools again had no different impact. The risk tool led to a mean chance score of 39.3 (35.2, 43.3), the description tool to 40.5 (36.1, 44.8), the distribution tool to 37.8 (34.0, 41.7), and the sampling tool to 35.1 (30.9, 39.3). The same pattern emerges for the perceived chances of ending with a loss. The risk tool led to a mean chance score of 38.9 (35.7, 42.1), the description tool to 37.9 (33.9, 41.9), the distribution tool to 40.1 (36.7, 43.4), and the sampling tool to 37.8 (34.1, 41.5). Again, the tools had no effect on perceived return. Excluding one outlier, the risk tool led to a mean perceived return of 272 (198, 346), the

description tool to 228 (187, 270), the distribution tool to 322 (225, 418), and the sampling tool to 226 (181, 272).

Overall, the tools had no impact on subjective assessments of the investment problem faced by participants in the kind environment.

Assessments in a Wicked Environment

Participants across tools judged the wicked environment to be more risky than participants in the kind environment; however, no clear pattern emerged across tools and subjective assessments.

In the wicked environment, participants who used the risk tool scored 4.66 (4.43, 4.89) on risk perception, no different to participants who used the description tool (4.85, 4.62, 5.08) or the distribution tool (4.81, 4.59, 5.03), but lower than those using the sampling tool (4.99, 4.78, 5.20).

Compared to previous findings, participants who used the risk tool felt less informed (3.49, 3.27, 3.72) than participants who used the description tool (3.94, 3.70, 4.17), but no differently informed than participants who used the distribution (3.36, 3.14, 3.58) or sampling (3.36, 3.12, 3.61) tools.

The risk tool led to higher confidence (3.92, 3.68, 4.16) than the distribution tool (3.53, 3.30, 3.76), but no more confidence than the description (3.74, 3.50, 3.98) or sampling (3.70, 3.48, 3.92) tools.

In terms of satisfaction, the description tool led to the highest scores. Participants who used the description tool not only felt most informed but were also more satisfied (4.72, 4.43, 5.02) than participants who used the risk tool (4.41, 4.13, 4.69), the distribution (4.34, 4.08, 4.61) or sampling (4.40, 4.12, 4.69) tools.

With respect to the perceived chances of ending with a positive return, the description tool led to the highest chances (40.8, 35.8, 45.9) compared to the distribution tool (34.5, 30.5, 38.6), the sampling tool (35.2, 30.9, 39.5), and the risk tool (33.9, 29.6,

38.3). The perception of the chance of ending with a loss did not differ across tools. The risk tool led to a mean chance score of 37.1 (33.4, 40.7), the description tool to 41.0 (36.8, 45.2), the distribution tool to 37.7 (34.3, 41.2), and the sampling tool to 39.1 (35.4, 42.9). Again, there was no effect on perceived return. Excluding one outlier, the risk tool led to a mean perceived return of 240 (158, 323), the description tool to 197 (138, 257), the distribution tool to 252 (186, 317), and the sampling tool to 160 (132, 188).

Interim Discussion

In previous work, the risk tool led—relative to other tools—to lower risk perception and to higher levels of feeling informed and confidence, which translated into more financial risk taking. Our results do not replicate these observations, either in a kind or in a wicked environment. Both investments and subjective assessments of the investment decision did not differ markedly or systematically across tools and environments. Although the risk tool led to the perception of lowest risk in the wicked environment, this observation was not accompanied by a higher feeling of being informed or by being more confident. Across results, the risk tool did not emerge as the tool that promotes investments in either kind or wicked environments. In contrast to previous work, there was an indication that the description tool led participants to feel informed, both in the kind and wicked environments.

Validity Analyses

Our results indicate that there is no systematic effect of the different risk communication tools on the amount of risk that people take or on their subjective assessments of financial risk. Here we examine the possibility that these null results might be caused by the possibility of participants' giving inconsistent responses and that they may have been unresponsive to the structure of the investment game.

Consistency

We examined whether participants who see themselves as greater risk takers do indeed take more financial risks, as well as those who judge losses as being less likely, those who judge the stock as being less risky, and those who are more confident in the stock. Results indicate high consistency in responses. Participants who see themselves as greater risk takers took more risks in the investment task, as measured by the positive correlation between the score in the SOEP's general risk item and the proportion of funds invested in stocks by each participant ($r = .27$). And indeed, higher financial risk was also taken by those who judged the stock as being less risky ($r = -.183$), those who perceived losses as being less likely ($r = -.18$), and those who were more confident in the stock ($r = .38$).

Responsiveness

Participants were also responsive to the outcomes of their investments. After a gain, participants were 1.7 times more likely to increase their investment in the stock as compared to decreasing it whereas, after a loss, participants were 1.2 times more likely to decrease their investment in the stock as compared to increasing it. Furthermore, in Study 2, participants who invested at least 50% of their portfolio in the stock were 2.4 times more likely to decrease their investment in the stock relative to increasing it after experiencing the rare extreme loss.

Finally, participants were sensitive to the structure of the investment decision. Overall, risk taking in Study 1, where the stock was a relatively attractive prospect, was significantly higher than in Study 2, where the deposit was better ($\delta = -1.38$, $95\%CI = [-1.52, -1.25]$). This pattern is also present in participants' risk perception—with higher risk perceived in Study 2 ($\delta = .433$, $[.425, .441]$), confidence ($\delta = -.417$, $[-.425, -.408]$), and feelings of being informed ($\delta = .046$, $[.038, .055]$).

Robustness of Null Effects

Our results show that the four risk communication tools used in our experiments have no differential impact in the way people take financial risk and form subjective representations of that risk. This pattern of null results was robust to whether or not participants had prior experience (as in Study 1 and 2) and to whether or not the risky stock entailed the possibility of a rare and large loss (Study 2). Here, we examine the robustness of these null results across a variety of potential moderators; namely, the amount of attention that participants paid to the task (based on the whether participants passed five or six attention checks), the degree of engagement with the tools (according to a composite measure of the number of allocations sampled and time spent with the tool), participants' financial expertise (based on a composite measure of the Berlin Numeracy Test, the survey of investment frequency, and the investment quiz), and different instructions for the risk tool [2].

To examine the moderating effects of these factors, we estimated linear models for each response variable (risk taking, risk perception, confidence, and feeling informed) and each of several different moderators. Figure 4 shows, on the y-axis, the additional explained variance (change in adjusted R^2) of a model that includes the interaction term relative to a model with only the main effects of tool and moderator. The x-axis shows the moderators that were examined and response variables that are each represented by a

[2] To rule out the possibility that the null effects observed in Study 1 and 2 were caused by participants failing to understand the instructions of the task, we conducted a third experiment in which we provided more detailed instructions about the risk tool, and compared responses with those of participants in the description tool with unchanged instructions. If instructions were ambiguous in Study 1 and 2, and are clearer in the current version of the risk tool, responses should differ across the two conditions. However, responses in this third experiment (in terms of risk taking and subjective assessments of risk) did not differ across participants who used the risk tool with improved instructions and the description tool with the older instructions. In short, the results of this third study suggest that the null results in Study 1 and 2 were not driven by ambiguous instructions.

color-coded square. Results show that the null effects of the tools are robust across many moderators and response variables. For example, tools have no different impact independently of whether participants are attentive, engaged or expert, or whether the risk tool offers more detailed instructions. Of the 24 moderating interactions, 22 are not significant, which means that null effects are robust. Only prior experience and the type of environment interact with the tools to influence risk taking; namely, participants using the sampling tool took more financial risk when they had prior experience than when they did not. For participants using other tools, no such increased risk taking was observed under the prior experience condition. Furthermore, participants in wicked environments who used the sampling and the risk tools took more risk than those who used other tools; however, neither prior experience nor the type of environment had any effects on how the tools influence risk perception, confidence, and feeling informed.

Comparison With Kaufmann et al. (2013)

The results of our studies speak unambiguously. They stand in stark contrast with those of Kaufmann et al. (2013). Whereas Kaufmann et al. (2013) showed that the risk tool promotes more informed and confident investors who are willing to take more risk, our results show no traces of such an effect. Why do the results differ so markedly? One possible explanation is that the incentives for participants were different. In Kaufmann et al. (2013)'s Experiment 1, 10 of the 133 University of Mannheim students who participated in the study were compensated with Amazon.com gift cards whose worth was proportional to their investment performance. The gift cards were worth between €10 and €18. Thus, participants were paid, on average, €1.05. In their Experiment 2, 190 participants from the pool of the Yale School of Management earned a $5 gift card and had a 5% chance to earn an unspecified additional performance-based compensation. In Experiment 3, 362 participants—also from the Yale School of Management—had a 50% chance to earn a $5 Amazon.com gift card and a 2.5% chance to earn an unspecified

additional pay. On average these participants received a minimum of $2.5. In Experiment 4, 212 participants from Amazon MTurk earned $1.30 and a "20% chance to earn additional performance-based pay". And in Experiment 5, 5 out of 39 students from the University of Mannheim, Germany, who acted as participants, earned an "Amazon.com gift card for the amount of the financial market simulation divided by 100", which translates into €1.42, on average. Therefore, with the exception of participants from the pool of the Yale School of Management, who earned between $2.5 and $5, participants in Europe earned between €1.05 and €1.42 for the experiment, and those in MTurk $1.30. These amounts are slightly below the compensation that we paid our participants in Prolific; namely, £1.62 in Study 1 and £1.7 in Study 2. Moreover, because our experiments were conducted in Prolific, our compensation schemes had to meet Prolific's principle of "ethical rewards", which required fair minimum pay for participation. In short, the compensation of participants across the original experiments and our studies does not seem to differ sufficiently to be the cause of the distinct results.

Another possible explanation might lie in the composition of the sample of participants. In their Experiment 1, Kaufmann et al. (2013) used a sample of participants from the University of Mannheim (61% male), with an average age of 22 (18 to 50), of whom approximately 30% reported owning stocks. Experiments 2 and 3 used a similar sample of participants (41% male) recruited from the pool of the Yale School of Management, aged on average 34 (18 to 70), with a median income of $40,000 ($0 to $199,000), of whom 50% were college educated and 45% owned stocks. Experiment 4 used an MTurk sample of US participants (49% male), aged on average 36 (20 to 68), with a median income of $39,000 ($0 to $200,000), of whom 51% were college educated and 31% owned stocks. Experiment 5 used University of Mannheim students (59% male), aged on average 24 (18 to 43), with approximately 36% who owned stocks. In other words, the samples used by Kaufmann et al. (2013) are characterized by being diverse in terms of age, gender (47% male across experiments), education, and income, and in which more than a

third of the participants owned stocks. Our experiments used two Prolific samples from the same participant pool. In both of our studies, 3,106 participants were, on average, 37.8 years old (18 to 86), 35.7% were male, and 43% reported dealing with investment instruments at least once a year. The mean income was approximately £23,000 (£0 to £120,000). Overall, these numbers reflect only moderate differences between the samples used by Kaufmann et al. (2013) and in our studies, implying that differences in sample composition do not offer a convincing explanation of the differences in results.

General Discussion

The power of simulations to communicate risk effectively has become evident in recent years (Hertwig & Wulff, 2022). Simulations help people make probability judgments more accurately (Hogarth & Soyer, 2011), judge the risk of medications such as opioids more precisely (Wegwarth et al., 2022), and engage in financial risk with more knowledge and confidence (Bradbury et al., 2015; Kaufmann et al., 2013). As much as simulations seem to promote effective risk communication, it is important to understand under which conditions they work well. Here, we conducted a conceptual replication of the work by Kaufmann et al. (2013) with minor variations that, a priori, should not have caused different results from the original study. Namely, we created our own implementations of the risk communication tools—including the main simulator, the risk tool—according to the original design and endorsed by the original author. Furthermore, instead of testing the effectiveness of the risk tool (and of other tools) in a one-shot investment decision, we examined the tools in a set of 20 investment periods (cf. Wulff et al., 2015). Even though the differences were arguably minimal, across three studies amounting to 3,455 participants, the risk tool did not lead to more risk taking or to more accurate subjective representations of risk as observed in previous studies. Also, having some experience with the investment problem did not have any influence on the effectiveness of the risk tool, nor on the influence of other tools on behavior.

If our implementation of the risk communication tools had no effect on investment behavior or on subjective representations of risk, one might suspect that our samples included a large share of inattentive participants. Indeed, online data collection has been questioned on the basis of participants' lack of attention (e.g., McCrea et al., 2015). For this reason, we employed six different attention checks and filtered the data accordingly. Several signs indicate that the data are reliable: Across both studies, participants who perceived more risk also took less risk, more confident participants took more risks, and overall participants showed systematic reactions to the outcomes of their investments. Moreover, the general pattern of null results does not change when we restrict all the analysis to the most attentive, most engaged, and most expert participants.

The consistent pattern of null results poses new questions worth exploring in the future. Are minor implementation decisions critical to how well the risk tool and other simulators communicate risk? Is the risk tool effective in a one-shot decision situation but not when people know they will engage in several investment decisions? Finally, is the risk tool effective only for a specific profile of potential investors but not for the general public, as we explored in these studies? New studies should continue to seek answers to these open questions and help us understand when, how, and for whom financial simulators work well.

Appendix

Preliminary Study

We conducted a preliminary study before Study 1 in which we observed that the tools were not adequate representations of the tools used by Kaufmann et al. (2013), for example, participants were not required to explore a minimum number of samples before they could continue investing. Also, this experiment did not include effective attention checks. We therefore improved the implementation of the tools for both of our studies in line with the original implementations and added attention checks. Results from this experiment were largely consistent with the results reported in the main text. Although we

do not report the details of this experiment here, the data are available in (TO BE ADDED).

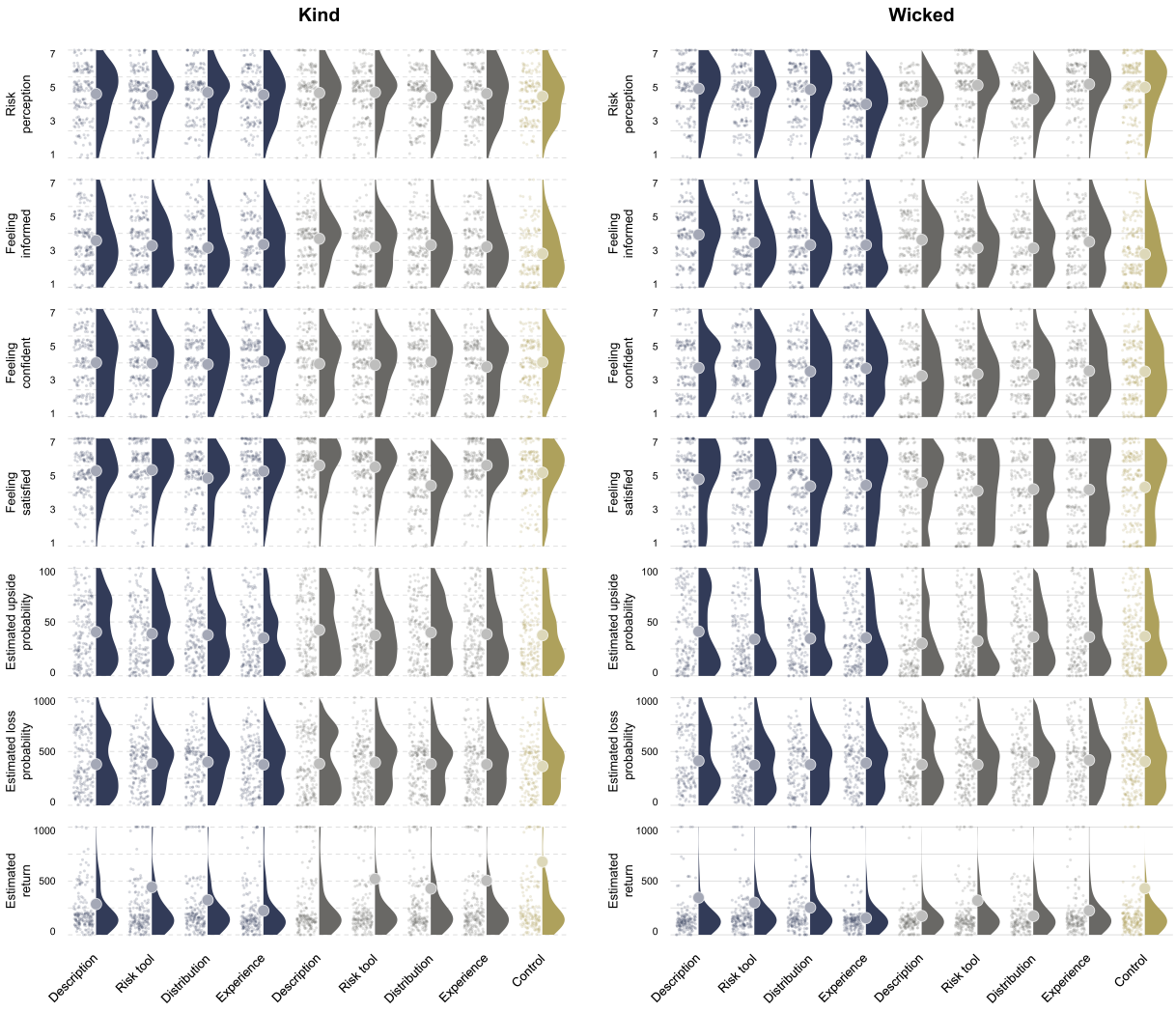

Figure 3

Subjective risk assessments by condition and study. The results are displayed as raincloud plots that show the original data (points), the density distribution (polygon), and the average value across participants (large circle).

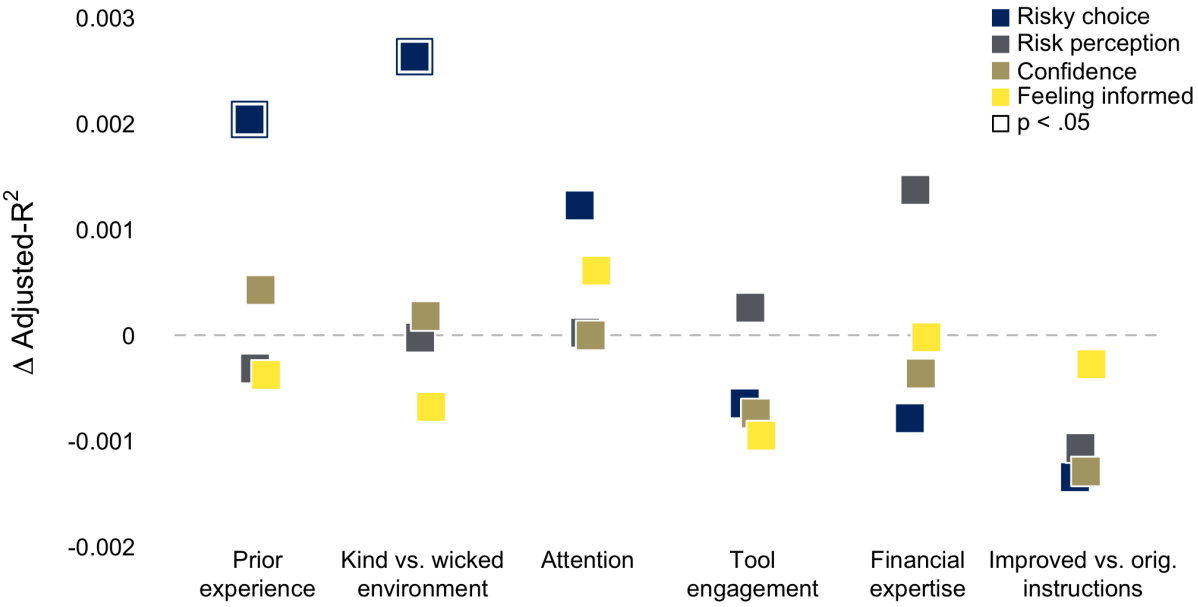

Figure 4

Moderator analysis of the effect of risk intervention tools. Each square shows for each combination of moderator (x-axis) and outcome variable (color; see legend) the change in adjusted R^2 that results from adding the interaction between the moderator and a variable coding the risk communication tool to a regression with only main effects. Prior experience contrasts participants with and without access to the investment task before entering the risk communication tool. Kind vs. wicked environment contrasts participants in Study 1 and 2. Attention contrasts participants who have failed 1 or 0 attention checks. Tool engagment is a composite of the time spent exploring the risk communication tool and the number of allocations explored. Financial expertise is a composite of the number of correct answers in the investment quiz, self-reported investment frequency, and the number of correct answers in the Berlin Numeracy Test. Improved vs. original instructions contrasts the data of Study 1 and 2 with a new study with improved instructions (see Footnote 1).

References

Barron, G., Leider, S., & Stack, J. (2008). The effect of safe experience on a warnings' impact: Sex, drugs, and rock-n-roll. *Organizational Behavior and Human Decision Processes*, *106*(2), 125–142.

Bradbury, M. A., Hens, T., & Zeisberger, S. (2015). Improving investment decisions with simulated experience. *Review of Finance*, *19*(3), 1019–1052.

Cohn, A., Engelmann, J., Fehr, E., & Maréchal, M. A. (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review*, *105*(2), 860–85.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The berlin numeracy test. *Judgment and Decision making*.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., & Schupp, J. (2019). The german socio-economic panel (soep). *Jahrbücher für Nationalökonomie und Statistik*, *239*(2), 345–360.

Hertwig, R., & Wulff, D. U. (2022). A description–experience framework of the psychology of risk. *Perspectives on psychological science*, *17*(3), 631–651.

Hogarth, R. M. (2001). *Educating intuition*. University of Chicago Press.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science*, *24*(5), 379–385.

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology: General*, *140*(3), 434.

Kaufmann, C., Weber, M., & Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management science*, *59*(2), 323–340.

Kaustia, M., & Knüpfer, S. (2008). Do investors overweight personal experience? evidence from ipo subscriptions. *The Journal of Finance*, *63*(6), 2679–2702.

Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. *Organizational Behavior and Human Decision Processes*, *116*(2), 286–295.

Lejarraga, T., Woike, J. K., & Hertwig, R. (2016). Description and experience: How experimental investors learn about booms and busts affects their financial risk taking. *Cognition*, *157*, 365–383.

McCrea, S. M., Penningroth, S. L., & Radakovich, M. P. (2015). Implementation intentions forge a strong cue–response link and boost prospective memory performance. *Journal of Cognitive Psychology*, *27*(1), 12–26.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science*, *26*(5), 559–569.

Wegwarth, O., Ludwig, W.-D., Spies, C., Schulte, E., & Hertwig, R. (2022). The role of simulated-experience and descriptive formats on perceiving risks of strong opioids: A randomized controlled trial with chronic noncancer pain patients. *Patient Education and Counseling*, *105*(6), 1571–1580.

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). How short-and long-run aspirations impact search and choice in decisions from experience. *Cognition*, *144*, 29–37.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological bulletin*, *144*(2), 140.