Downside financial risk is misunderstood

Philip W. S. Newall*

Abstract

The mathematics of downside financial risk can be difficult to understand: For example a 50% loss requires a subsequent 100% gain to break-even. A given percentage loss always requires a greater percentage gain to break-even. Instead, many non-expert investors may assume for example that a 50% gain is sufficient to offset a 50% loss. Over 3,498 participants and five experiments, the widespread illusion that a sequence of equal percentage gains and losses produces a zero overall return was demonstrated. Participants continued to err frequently, even with percentage returns of +/-100%, or when financially incentivized. Financial literacy, numeracy, and deliberation were all shown to independently contribute to accurate performance. These results have implications for promoting the understanding of downside financial risk.

Keywords: financial risk, downside risk, numeracy, percentages, financial literacy, deliberation

1 Introduction

"Stocks plunge 508 points, a drop of 22.6%", was how the New York Times reported the fall in the Dow Jones Industrial Average on Black Monday, the record one-day stock market crash on October 19th 1987. Financial returns are usually reported as percentages, in order to standardize price changes across different markets, investments, and investors. An unusual feature of percentage returns is that a subsequent *29.2%* increase is required to reverse Black Monday's *22.6%* decrease. Recent past stock market returns may be more salient for novice investors. An investor may look at the two most recent US bear markets and be surprised by the required subsequent returns to break even. The peak-to-trough fall from 2000-2003 of 43.7% required a 77.5% return to break even, while the 2007-2009 fall of 50.8% required a 103.4% return to break even (Shiller, 2016). A return sequence of +x% and –x% results in a negative overall return — it doesn't matter which return comes first — with the size of the negative overall return increasing with the size of x (Bodie, Kane & Marcus, 2001). This paper explores the causes and prevalence of the potentially widespread misunderstanding of this aspect of downside financial risk.

Personal investors will often interact with percentage information. For example, investors will often read rule-of-thumb advice that a maximum tolerable loss for diversified stock market investments is around 50%, and that this should be accounted for when setting allocation limits between stocks and safer assets such as bonds (Garrett, 2012;

Swedroe & Balaban, 2012). But only investors who understand the asymmetry between positive and negative percentage returns can rationally process this information. Many investors are instead likely to assume that returns of +50% and –50% cancel out. Investors may overestimate their risk tolerance if they mistakenly underestimate the devastating impact of large percentage losses. For example, novice participants in the recent dot com and real estate investment bubbles may have had their confidence bolstered by several early years of large percentage gains, unaware of how easily a sequence of losses can reverse these gains. Although this bias may be reduced in decisions from experience compared to decisions from description (Hertwig, Barron, Weber & Erev, 2004), this can be an expensive lesson to learn via experience.

A return sequence of +x% and –x% results in a negative overall return, with overall losses increasing as x increases (Bodie et al., 2001). While returns of +10% and –10% result in a total return of –1%, a return sequence of +50% and –50% results in a total return of –25%. It is easiest to understand the correct answer by simulating the value of an investment over the two years. For returns of +/-10%, an initial investment of $100 increases to $110 in year one, before decreasing by $11 to $99. The second percentage change multiplies both the initial value and the first percentage change. For returns of +/-50% a stock worth $100 increases to $150, before falling to $75. I hypothesize that, instead of performing this two-step multiplicative procedure, many investors will perform a one-step additive procedure and assume that returns of +x% and –x% = 0.

Research indicates that investors struggle to perform the multiplication required in compound interest calculations, involving annual percentage *increases* (Eisenstein & Hoch, 2007; McKenzie & Liersch, 2011). For example, an annual growth of 10% over three years requires multiplying 1.1*1.1*1.1 by the initial account balance. These experi-

---

*Behavioural Science Centre, Stirling Management School, University of Stirling, Stirling, FK9 4LA, U.K. Email: p.w.newall@stir.ac.uk.

ments suggest that a number of participants perform a simpler additive calculation, where the first year's return is added to the anticipated final balance for every year (e.g., 1 + 0.1 + 0.1 + 0.1). When dealing with a sequence of pure increases this leads to a systematic underestimation of the final balance (Eisenstein & Hoch, 2007; McKenzie & Liersch, 2011).

Prior consumer research suggests that percentages are challenging generally, and not just in the financial domain (Chen & Rao, 2007; Chen, Marmorstein, Tsiros & Rao, 2012; Kruger & Vargas, 2008). For example, Kruger and Vargas (2008) find systematic framing effects between saying product A is 50% more expensive than product B, compared to saying that product B is 33% cheaper than product A. (Imagine prices of $150 for product A and $100 for product B.) Chen and Rao (2007) find that successive percentage discounts or surcharges on consumer products are processed additively, rather than multiplicatively.

Downside financial risk highlights the numerous inputs to good financial decisions. Financial literacy is often-investigated as a cause of poor financial behaviors (Mitchell & Lusardi, 2011). But a recent meta-analysis of the literature suggests that financial literacy interventions have little average effect on financial behaviors (Fernandes, Lynch & Netemeyer, 2014). Downside financial risk is primarily a numerical problem, however, and numeracy has been highlighted as another predictor of financial behaviors (Cokely & Kelley, 2009; Cole, Paulson & Shastry, 2014; Estrada-Mejia, de Vries & Zeelenberg, 2016; Ghazal, Cokely & Garcia-Retamero, 2014). This paper explores associations between the understanding of downside financial risk, financial literacy, and numeracy.

Five experiments are reported below to explore the (mis-)understanding that a return of –x% more than wipes out a return of +x%. I hypothesize that participants will most commonly err by answering that this return sequence results in a zero overall return (all Experiments). One important issue is the extent to which this error is independent of return size volatility, x (Experiments 1, 2, and 4). Experiment 3 tests the robustness of this effect to financial incentives. Potentially relevant individual difference measures are investigated (Experiments 1 and 5). Finally, Experiments 4 and 5 test a debiasing intervention, and Experiment 5 tests potential causal mechanisms through which this intervention operates.

2 Experiment 1

2.1 Method

In all five experiments, participants aged 18 and over and from the US were collected via Amazon Mechanical Turk, and paid $.10. A total of 981 participants were recruited for Experiment 1, to create a highly-powered initial study. For this experiment, 53.3% of the sample had a college degree, 57.9% were female, and the average age was 36.4.

Return size volatility was manipulated within-participants:

> Suppose a stock increases 10% [50%] in year one, decreases by 10% [50%] in year two, and does not pay any dividends for the duration. Is the stock's final value more than, equal to, or less than its initial value?

In a counterbalanced design, half of participants answered the +/-10% question at the start of the experiment, and the +/-50% question at the end. The other half of participants answered these questions in the opposite order. These questions were put as far apart as possible to minimize recall effects.

Other experimental measures were collected between these two questions: demographic information of age, gender, and education; financial literacy and numeracy scales (which were potentially-relevant individual difference variables). The multiple-choice format of the Berlin numeracy test (Cokely, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012), and a 13-part financial literacy scale (Fernandes et al., 2014) were selected as scales possessing good psychometric features. The mean financial literacy score was 8.6 out of 13, while the mean numeracy score was 1.6 out of 4 (compared to mean scores of 7.3 out of 13 and 2.1 out of 4 reported in those two papers, respectively).

2.2 Results and discussion

Table 1 presents results of the downside risk questions. The modal participant answered neither question correctly: 50.8% answered both downside risk questions incorrectly. Just over a third of participants answered both downside risk questions correctly (33.9%), while the remainder answered one question correctly (15.3%).

Table 2 presents results of each individual question, broken down by percentage size and question order. The results were stable: The percentage of correct responses ranged in a narrow band, from 38.8% for +/-10% first, to 44.3% for +/-50% last. For the first question of the survey, the modal response was "*equal to its initial value*", for both +/-10% and +/-50% questions. The correct answer "*less than its initial value*", was the other most common response. Therefore, in all the regression analyses that follow, multinomial logistic regression was used to predict the likelihood of a response shifting from "equal to" to "less than" its initial value.

A multinomial logistic regression was run to model the factors predicting equal-to/less-than responses. Standard errors were clustered per-participant because in this within-participants design there were two observations of the dependent variable per-participant. Table 3 shows the results. Financial literacy and numeracy scales were standardized

Table 1: Overall percentage of correct responses to downside risk questions.

Number of questions answered correctly	Percentage of participants
0	50.8%
1	15.3%
2	33.9%

Table 2: Responses to each downside risk question.

	First		Last	
Response	+/-10%	+/-50%	+/-10%	+/-50%
More than its initial value	14.6%	12.5%	13.5%	13.4%
Equal to its initial value	43.9%	43.6%	38.9%	38.4%
Less than its initial value	38.8%	41.3%	41.9%	44.3%
Don't know	2.7%	2.6%	5.7%	3.9%

before being included in the regression. Both of these scales were highly predictive of the correct response. Question order was also significantly related to the correct response, showing that participants had a slight tendency toward learning the correct response over the course of the experiment. Return size volatility, however, was not significant, showing that performance was identical across +/10% and +/-50% questions. Gender was the only significant demographic predictor of the correct response, with males on-average scoring higher than females. Statistical significance of these relationships was unchanged in a series of single predictor regression models.

A mediational analysis was next performed to structurally model the potentially causal pathways linking numeracy, financial literacy and equal-to/less-than responses. The khb command in Stata was used for performing mediation on a multinomial logistic regression, which produces test statistics based on the Sobel test (Kohler, Karlson & Holm, 2011). Standard errors were again clustered per-participant. Financial literacy was found to partially mediate the link between numeracy and task performance. Both the direct link between numeracy and task performance ($z = 7.61$, $p < .001$), and the indirect link from numeracy via financial literacy to task performance ($z = 7.71$, $p < .001$) were statistically significant. The indirect link via financial literacy explained 31.2% of the total relationship between numeracy and task performance. The mediational analysis reveals both direct and indirect effects of numeracy on task performance. That

is, part of the reason numeracy predict overall performance is because more numerate participants also tend to be more financially literate. However, it must be noted that the version of the Berlin numeracy task here showed some skewness (mean response of 1.6 out of 4). The Berlin-Schwartz numeracy scale has been shown to have better psychometric properties in Mechanical Turk samples, and may have led to different results (Cokely et al., 2012; see also Experiment 5 below and the general discussion).

The result showing that the proportion of equal-to/less-than responses was unaffected by return size volatility motivated the design of Experiment 2. This was to explore the robustness of this result to increasing return size volatility. Recall that responding "*equal to its initial value*" overestimates the final portfolio value to an increasing degree as return size volatility increases.

3 Experiment 2

Experiment 2 compared the +/-10% question with a more extreme +/-100% question (where the -100% return makes the investment worthless going forward).

3.1 Method

Another 287 participants were recruited (no participant took part in more than one experiment), and assigned to one of two conditions. Participants answered either the +/-10% question from Experiment 1, or a manipulation featuring the most extreme negative return possible (for a non-leveraged investment):

> Suppose a stock increases 100% in year one, decreases by 100% in year two, and does not pay any dividends for the duration. Is the stock's final value more than, equal to, or less than its initial value?

3.2 Results and discussion

Results are shown in Table 4. As can be seen, more people answered correctly with the +/-100% return size. But this is because a smaller percentage of participants answered "*more than its initial value*" in the +/-100% condition (2.8%) than in the +/-10% condition (17.4%). A multinomial logistic regression showed that the shift from responding "*equal to its initial value*", to "*less than its initial value*" was almost significant ($B = -0.51$, $z = -1.95$, $p = .051$, 95% CI [-1.01, 0.01]). However, the large difference in "more than its initial value" responses across the two conditions conflicts with the independence of irrelevant alternatives assumption of the multinomial logistic regression model. Therefore, a binary logistic regression model was also run, and showed that the proportion of "*equal to its*

Table 3: Multinomial logistic regression estimates from Experiment 1, comparing equal-to/less-than responses. Positive values correspond to a higher probability of responding "*less than its initial value*".

Variables	Estimate	z	p	95% Confidence Interval
Financial literacy	0.68	7.04	<.001	[0.49, 0.87]
Numeracy	0.54	7.24	<.001	[0.39, 0.69]
Question order	0.21	3.30	0.001	[0.09, 0.34]
Return size	0.10	1.57	0.117	[–0.03, 0.23]
Age	–0.01	–0.92	0.355	[–0.02, 0.01]
Education				
High school graduate	–1.19	–1.80	0.072	[–2.48, 0.11]
Some college	–0.49	–0.77	0.440	[–1.74, 0.76]
College graduate	–0.01	–0.01	0.994	[–1.26, 1.25]
Gender	–0.45	–2.94	0.003	[–0.74, –0.15]

Note: Baseline levels were the first question for question order, +/-10% for return size, some high school for education, and male for gender.

Table 4: Results of experiment 2.

Response	+/-10%	+/-100%
More than its initial value	17.4%	2.8%
Equal to its initial value	43.1%	40.6%
Less than its initial value	34.0%	53.2%
Don't know	5.6%	3.5%

Table 5: Results of experiment 3.

Response	no incentive	incentive
More than its initial value	20.3%	11.9%
Equal to its initial value	35.7%	37.3%
Less than its initial value	40.6%	48.5%
Don't know	3.5%	2.2%

initial value" responses did not differ significantly between the two conditions ($B = 0.10$, $z = 0.43$, $p = .688$, 95% CI [–0.37, 0.57]).

Although participants were more accurate in the +/-100% condition, this increase in accuracy did not come from a decrease in "*equal to its initial value*" responses. If the magnitude of overestimation is considered as an outcome measure, then participants actually performed worse in the +/-100% condition (overestimating final value by 100%) than in the +/-10% condition (overestimating by 1%).

While these are interesting results to hypothetical questions, real world decisions will have (often large) financial incentives. Therefore, the robustness of the effect to financial incentives motivated the design of Experiment 3.

4 Experiment 3

4.1 Method

Another 277 participants were recruited and assigned to one of two conditions. In a between-participants design, participants answered the +/-10% question, either as-before with

only a $0.10 baseline fee, or with an additional $0.10 incentive (giving participants the chance to earn $0.20 in total with a correct answer). Although this is not a high absolute level of financial incentives, it is a high relative level of incentive. Participants in the incentive condition saw the message, "Answer this question correctly and earn a $0.10 bonus!" immediately above the downside risk question text. All bonuses were credited within 24 hours.

4.2 Results and discussion

Results of Experiment 3 are in Table 5. The two conditions had very similar frequencies of "*equal to its initial value*" responses: 35.7% with no incentive, and 37.3% with incentives. A multinomial logistic regression showed that the financial incentive did not induce a statistically significant shift between equal-to/less-than responses ($B = 0.13$, $z = 0.50$, $p = .619$, 95% CI [–0.39, 0.66]). The only effect of financial incentives was a decrease in the least-common response "*more than its initial value*", from 20.3% to 11.9%, with most of these participants now moving toward the correct answer. This difference may also be problematic for the

Table 6: Responses to each downside risk question.

Response	+/-10%		+/-50%	
	no debiasing	debiasing prompt	no debiasing	debiasing prompt
More than its initial value	18.9%	11.6%	15.2%	10.7%
Equal to its initial value	45.3%	41.0%	44.0%	32.5%
Less than its initial value	30.7%	46.2%	36.6%	54.8%
Don't know	5.1%	1.2%	4.3%	2.0%

independence of irrelevant alternatives assumption for the multinomial logistic regression model. Therefore, a binary logistic regression was also run, which also failed to detect any significant difference in "*equal to its initial value*" responses across the two groups ($B = 0.07$, z = 0.28, p = .776, 95% CI [–0.42, 0.56]).

Experiment 3 bears two similarities with Experiment 2. In both experiments, a manipulation increased the percentage of correct responses, but without reducing the frequency of "*equal to its initial value*" responses. Both of these experiments provide evidence for at least one extra pattern of mistakes, which leads to a minority of participants responding "*more than its initial* value", in the baseline condition, but which is then substantially reduced with the +/-100% question or financial incentives. It is beyond the scope of this paper to investigate why this might be the case or whether these results are robust, but a greater understanding of the heterogeneity of mistakes in this task may help improve the understanding of underlying psychological factors. However, "*equal to its initial value*" is in both experiments the modal incorrect response, and the frequency of this incorrect response was unaffected by the experimental manipulations. This result suggests that this is the most frequent and robust incorrect response to downside financial risk.

5   Experiment 4

Experiment 4 was designed to explore the extent to which participants can be debiased and potential mechanisms that help participants avoid the "*equal to its initial value*" response. Recall that calculating the correct response requires holding the year one value in memory before calculating the impact of the second percentage change. Therefore, a debiasing prompt was included as a new experimental condition to try and promote more accurate evaluation of downside financial risk.

5.1   Method

A total of 1,014 participants were recruited. A 2x2 between-participants design was used, manipulating return size volatility (+/-10%, +/-50%) and debiasing prompt (no debiasing, debiasing prompt). Experimental materials were the same as before, with the following prompt added in the debiasing conditions after the first sentence of description:

> When answering, try to imagine what would happen to a $100 initial investment over the two years. Think about the investment's value after year one, and then its value after year two.

This is a direct prompt for participants to perform the additional step of mental calculation required to generate the correct answer.

5.2   Results and discussion

Results of Experiment 4 are in Table 6. As can be seen, "*equal to its initial value*" is the modal response with no debiasing prompt, for both return size conditions. Inclusion of a debiasing prompt improves responses, and "*less than its initial value*" becomes the modal response.

A multinomial logistic regression, with independent variables of return size volatility and debiasing prompt was run. Regression coefficients comparing equal-to/less-than responses revealed the following. The debiasing prompt was effective ($B = 0.60$, z = 4.30, p < .001, 95% CI [0.33, 0.88]). Participants were also slightly more accurate in the +/-50% condition than the +/-10% condition ($B = –0.30$, z = –2.14, p = .032, 95% CI [–0.57, –0.02]).

Although substantial error-rates remained, the debiasing prompt improved evaluation and judgment. There was in this experiment also a small effect of return size volatility, with participants being more accurate in the +/-50% condition. However, this effect was not large enough to correct for the greater magnitude of overestimation (responding "*equal to its initial value*" overestimates the final value by 1% in the +/-10% condition but by 25% in the +/-50% condition).

6   Experiment 5

Experiment 5 was designed to test the robustness of the previous experiments. In Experiment 5 an alternative question wording was used, and participants were asked about the presence of household investments. Response time, financial literacy, and the Berlin-Schwartz numeracy scale were further collected to investigate factors underlying the misunderstanding of downside financial risk.

Table 7: Responses across debiasing prompt and presence of household investments.

	no debiasing		debiasing prompt	
Response	no investments	investments	no investments	investments
More than its initial value	17.9%	21.1%	18.0%	15.0%
Equal to its initial value	39.6%	28.2%	31.4%	26.4%
Less than its initial value	37.2%	48.2%	47.4%	56.3%
Don't know	5.3%	2.5%	3.1%%	2.4%

6.1 Method

A further 939 participants were collected from Amazon Mechanical Turk. Participants had a mean financial literacy score of 8.3 out of 13 (Fernandes et al., 2014) and a mean Berlin-Schwartz numeracy score of 3.0 out of 7 (Cokely et al., 2012). More than half of the sample, 57.3%, reported some level of household investments by responding yes to the question, "Do you, or does anyone else in your household, own any stocks, bonds, or mutual funds in an investment account, or in a self-directed IRA or 401(k) retirement account?".

One goal of Experiment 5 was to see if the results of the previous experiments were robust to an alternative question wording. The following question wording was used to clarify that the second percentage change refers to the year 1 price by using the phrase "decreases by 10% of its new price":

> Suppose a stock increases 10% in year one, decreases by 10% of its new price in year two, and does not pay any dividends for the duration. Is the stock's final value more than, equal to, or less than its initial value?

Participants either answered this question, or a version of this question including the debiasing prompt from Experiment 4. Therefore, this was a two-condition between-participants experiment. Total response time was recorded for this question. Participants took on average 35.6 seconds on this question before continuing to the rest of the survey, which involved demographic, financial literacy, and numeracy blocks presented in randomized order.

6.2 Results and discussion

Table 7 shows that, "*equal to its initial value*" was the modal response for non-investors who did not receive the debiasing prompt. The correct "*less than its initial value*" was the

modal response for the three other groups. However, in all cases over a quarter of responses were for "*equal to its initial value*". A multinomial logistic regression showed that both debiasing prompt ($B = 0.34$, $z = 2.23$, $p = .026$, 95% CI [0.04, 0.64]) and the presence of household investments ($B = 0.48$, $z = 3.16$, $p = .002$, 95% CI [0.18, 0.78]) led to significant shifts from equal-to to less-than responses.

A multinomial logistic regression was next run to investigate the associations between equal-to/less-than responses and financial literacy, numeracy, and response time. These variables were first standardized before being added to the previous regression. There was a statistically significant positive link between all three of these variables and equal-to/less-than responses. The link with numeracy was the largest ($B = 0.92$, $z = 8.86$, $p < .001$, 95% CI [0.72, 1.12]), while links with financial literacy ($B = 0.55$, $z = 5.48$, $p < .001$, 95% CI [0.35, 0.74]), and response time were equal ($B = 0.52$, $z = 4.06$, $p < .001$, 95% CI [0.27, 0.77]). Numeracy had the strongest positive link with accurate performance, although all three effects were positive. Statistical significance of these relationships was unchanged in a series of single predictor regression models.

Potential mediational relationships were next tested using the khb command for mediation in non-linear probability models in Stata (Kohler et al., 2011). In all cases the shift from equal to- to less than- responses was assessed using a multinomial logistic regression. Consistent with results from Experiment 1, financial literacy was again found to partially mediate the link between numeracy and task performance. The model indicated that both the direct link between numeracy and task performance ($z = 8.72$, $p < .001$), and the indirect link from numeracy via financial literacy to task performance ($z = 5.40$, $p < .001$) were statistically significant. The indirect link via financial literacy explained 20.1% of the total relationship between numeracy and task performance, similar to that observed in Experiment 1 (i.e., 20.1% vs. 31.2%).

Next, potential mediational relationships between the debiasing prompt and financial literacy, numeracy, and response time on equal-to/less-than responses were tested. Neither financial literacy ($z = 0.77$, $p = .444$) nor numeracy ($z = 1.53$, $p = .126$) mediated the link between debiasing prompt and task performance. However, response time did fully mediate the link between debiasing prompt and equal-to/less-than responses: The indirect effect was significant ($z = 2.97$, $p = .003$) and the direct effect was no longer significant ($z = 1.32$, $p = .188$). Finally, I modeled the extent to which response time mediated the links between financial literacy and numeracy and equal to-/less than- responses. Across the sample as a whole, response time failed to mediate either financial literacy ($z = 0.50$, $p = .617$) or numeracy ($z = 0.04$, $p = .976$).

Experiment 5 has two main findings. First, many people including those who manage household investments ex-

Table 8: Results of two replication studies, using the original question wording (Prolific Academic) and the new question wording from Experiment 5 (Mechanical Turk).

Response	Prolific Academic		Mechanical Turk	
	+/-10%	+/-50%	+/-10%	+/-50%
More than its initial value	15.8%	17.1%	21.5%	17.1%
Equal to its initial value	30.1%	26.0%	31.3%	40.7%
Less than its initial value	47.3%	52.7%	43.1%	41.4%
Don't know	6.9%	4.1%	4.2%	0.7%

pressed the same robust pattern of errors found in Experiments 1–4. However, this pattern was substantially reduced, but not eliminated, by a variation in question wording that encouraged participants to process the numerical information more carefully and deliberately. This result helps support the external validity of these findings. Second, financial literacy, numeracy, and response time all contributed to the shift from equal-to to less-than responses. Interestingly, the effect of the debiasing prompt was fully mediated by response time, indicating that the debiasing prompt worked by prompting greater deliberation.

These results may continue to hold for investors in general. Newall and Love (2015) found a similar pattern of results with a question based on losses of +/-10%, after screening out Mechanical Turk participants without any household investments. Two further experiments were also conducted to establish the generality of these results. An experiment with 292 participants aged 18 and over from the US was run on Prolific Academic, another crowdsourcing website. Return size volatility was manipulated between-participants (+/-10%, +/-50%), using the original question wording. And another identical experiment was run on 284 participants from Mechanical Turk, using Experiment 5's question wording. Results of these two experiments are in Table 8. Multinomial logistic regression showed that return size volatility did not lead to a significant shift in equal-to/less-than responses in either the Prolific Academic experiment ($B = 0.26$, $z = 0.93$, $p = .354$, 95% CI [–0.29, 0.80]) or the Mechanical Turk experiment ($B = –0.30$, $z = –1.12$, $p = .262$, 95% CI [–0.83, 0.23]).

7 General discussion

These results show that downside financial risk is commonly misunderstood. The most frequent and robust error in these experiments is consistent with participants adding the ef-

fect of the two percentage changes, when these results of changes (expressed as proportions) should normatively be multiplied. This error is consistent with the strategies people seem to use underlying exponential growth bias, where sequences of percentage gains are underestimated (Eisenstein & Hoch, 2007; McKenzie & Liersch, 2011). This error is also consistent with the mistakes consumers make when evaluating percentage discounts or surcharges (Chen & Rao, 2007; Kruger & Vargas, 2008).

These results have implications for the communication of downside risk to diverse investors and others charged with communicating and managing financial products. Stock market losses are commonly communicated as percentages normalized over different starting values. For example, the daily news report may say that the Dow Jones Industrial Average has declined 2% over the last day and 10% over the last month – percentage movements normalized over different starting values, as in the experiments reported in this paper. These results suggest that percentage changes should if possible be communicated in a different way. Reporting a series of percentage changes as a single aggregate change – performing the necessary multiplication – should improve understanding (i.e., reporting yearly returns of +10% and – 10% as a two-yearly return of –1%).

These results show that investors' intuitions about downside financial risk are most inaccurate for large percentage decreases. The US stock market has fallen by around 50% twice since the turn of the century (Shiller, 2016). Personal investment guides often warn investors to prepare for maximum losses of around this magnitude with diversified stock market investments (Garrett, 2012; Swedroe & Balaban, 2012). Many investors may unwittingly overestimate the risk tolerance if they do not correctly understand that such losses can take a long time to recover from, given historical after-inflation stock market returns of around 6–7% a year (Dimson, Marsh & Staunton, 2009). It takes a subsequent 100% return to recover from a 50% loss.

These results also shed light on potential drivers of the misunderstanding of downside risk, and the extent to which it can be debiased. Poor financial literacy has been much-studied as a potential cause of poor financial behavior (Mitchell & Lusardi, 2011). But other authors argue that directly increasing financial literacy has had little positive effect on financial behavior, and that positive correlations between literacy and behavior reflects omitted variables (Fernandes et al., 2014). Numeracy has been highlighted as another potential driver of poor financial behaviors (Cokely & Kelley., 2009; Cole et al., 2014; Estrada-Mejia et al., 2016; Ghazal et al., 2014). In Experiment 1 financial literacy and numeracy were approximately equally associated with task performance. Experiment 5 used a more sensitive measure of numeracy for the given participant pool, and found that numeracy was then a better predictor of task performance than financial literacy. (It should be noted that

these designs were correlational in nature and thus caution is merited when evaluating causal claims.) Nevertheless, both constructs were robust predictors of task performance, and in both experiments financial literacy partially mediated the link between numeracy and task performance. It could be that some knowledge of financial concepts is needed to accurately apply numerical reasoning to financial problems, such as the understanding of downside financial risk.

Theoretically, financial literacy and numeracy are both important in the understanding of downside financial risk, and both may be relatively hard to improve. A simple debiasing prompt improved responses in Experiments 4 and 5. Response time data in Experiment 5 showed that this debiasing prompt was fully mediated by response time, showing that its effect occurred from inducing greater deliberation (Cokely & Kelley, 2009). Prompting greater deliberation should be further investigated as a potentially cost-effective method of improving financial decisions.

8   References

Bodie, Z., Kane, A., & Marcus, A. J. (2001). *Investments* (5th ed.). New York: McGraw Hill.

Chen, H., Marmorstein, H., Tsiros, M., & Rao, A. R. (2012). When more is less: The impact of base value neglect on consumer preferences for bonus packs over price discounts. *Journal of Marketing, 76*(4), 64–77.

Chen, H., & Rao, A. R. (2007). When two plus two is not equal to four: Errors in processing multiple percentage changes. *Journal of Consumer Research, 34*(3), 327–340.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*(1), 25–47.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4*(1), 20–33.

Cole, S., Paulson, A. & Shastry, G. K. (2014). High school curriculum and financial outcomes: The impact of mandated personal finance and mathematics courses (Working paper no. 13-064). Retrieved from Harvard Business School website: http://www.hbs.edu/faculty/Publication%20Files/13-064_c7b52fa0-1242-4420-b9b6-73d32c639826.pdf.

Dimson, E., Marsh, P., & Staunton, M. (2009). *Triumph of the optimists: 101 years of global investment returns.* New Jersey, Princeton: Princeton University Press.

Eisenstein, E., & Hoch, S. (2007). *Intuitive compounding: Framing, temporal perspective, and expertise (working paper).* Unpublished manuscript.

Estrada-Mejia, C., de Vries, M., & Zeelenberg, M. (2016). Numeracy and wealth. *Journal of Economic Psychology, 54*, 53–63.

Fernandes, D., Lynch, J., G., & Netemeyer, R., G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science, 60*(8), 1861–1883.

Garrett, S. (2012). *Personal finance workbook for dummies* (2nd ed.). New York: John Wiley & Sons.

Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making, 9*(1), 15–34.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534–539. http://dx.doi.org/10.1111/j.0956-7976.2004.00715.x.

Kohler, U., Karlson, K. B., & Holm, A. (2011). Comparing coefficients of nested nonlinear probability models. *Stata Journal, 11*(3), 420–438.

Kruger, J., & Vargas, P. (2008). Consumer confusion of percent differences. *Journal of Consumer Psychology, 18*(1), 46–61.

McKenzie, C. R., & Liersch, M. J. (2011). Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research, 48*(SPL), S1–S13.

Mitchell, O. S., & Lusardi, A. (Eds.). (2011). *Financial literacy: Implications for retirement security and the financial marketplace*. New York: Oxford University Press.

Newall, P. W. S., & Love, B., C. (2015). Nudging investors big and small toward better decisions. *Decision, 2*(4), 319–326.

Shiller, R. J. (2016). Irrational exhuberence stock market data. Retrieved from http://www.econ.yale.edu/~shiller/data/ie_data.xls.

Swedroe, L. E., & Balaban, R., C. (2012). *Investment mistakes even smart investors make and how to avoid them.* New York: McGraw Hill.