Thinking about Prices versus Thinking about Returns in Financial Markets*

Markus Glaser [a] Zwetelina Iliewa [b] Martin Weber [c]

This draft: November 17, 2016

Abstract

Prices and returns are alternative ways to present information and to elicit expectations in financial markets. But do investors make sense of prices and returns in the same way? This paper presents three studies with subjects of varying expertise, with various amounts of information and with different incentive schemes. The results are consistent across all studies: Asking subjects to forecast returns as opposed to prices results in higher expectations, whereas showing them return charts as opposed to price charts results in lower expectations. Experience is not a useful remedy but Cognitive Reflection mitigates the impact of format changes.

Keywords: expected returns, information presentation, question format, judgmental forecasting, professional forecasters

JEL Classification: D14, D18, G02, G23

---

* Preliminary results of Study 3 were previously circulated with the title "Expert Forecasts: Fast, Frugal, Flawed". We thank Peter Bossaerts, Edgar Erdfelder, Arvid Hoffmann, Lena Jaroszek, Christine Laudenbach, Christoph Merkle, Alexandra Niessen-Ruenzi, Terrance Odean, Jesper Riedler, Paul Smeets, Michael Ungeheuer and Qingwei Wang for helpful suggestions. We are also grateful to seminar participants at the University of Mannheim, the Centre for Financial Research (Cologne) and the National Academy of Sciences Leopoldina as well as to conference participants at the Experimental Finance Conference 2016 and the Research in Behavioral Finance Conference 2016. We gratefully acknowledge financial support by the German Research Foundation (DFG), Research Grants GL695/3-1, SCHR709/4-1 and WE993/15-1. All errors are our own.

[a] Institute for Capital Markets and Corporate Finance, Ludwig-Maximilians-University Munich, E-mail: glaser@bwl.lmu.de

[b] Corresponding author: Zwetelina Iliewa, Centre for European Economic Research (ZEW), L7.1, 68161 Mannheim, E-mail: iliewa@zew.de

[c] University of Mannheim, Department of Banking and Finance, L5.2, 68161 Mannheim, Centre for European Economic Research (ZEW), L7.1, 68161 Mannheim, and the Centre for Economic Policy Research, London EC1V 0DX, E-mail: weber@bank.bwl.uni-mannheim.de

Thinking about Prices versus Thinking about Returns in Financial Markets

Abstract

Prices and returns are alternative ways to present information and to elicit expectations in financial markets. But do investors make sense of prices and returns in the same way? This paper presents three studies with subjects of varying expertise, with various amounts of information and with different incentive schemes. The results are consistent across all studies: Asking subjects to forecast returns as opposed to prices results in higher expectations, whereas showing them return charts as opposed to price charts results in lower expectations. Experience is not a useful remedy but Cognitive Reflection mitigates the impact of format changes.

Keywords: expected returns, information presentation, question format, judgmental forecasting, professional forecasters

JEL Classification: D14, D18, G02, G23

1 Introduction

Prices and returns are typically considered to be interchangeable ways to present financial market information and to elicit financial market expectations in surveys. Normative decision theory assumes that expectations are invariant to changes in superficial features of the information and the way they are elicited. However, do investors understand a chart of a fund's past performance in the same way when represented by past prices or past returns? And is it the same to ask an analyst to forecast prices or forecast returns? We separate the impact the chart format and the question format have, and compare the format of prices and the format of returns in three experimental studies. Across the studies we vary the level of expertise of the subjects, the amount of information and the incentives schemes. We report sizeable differences in the subjective expectations depending on the chart format and the question format, which are consistent across all studies: asking subjects to forecast returns as opposed to prices results in higher expectations. The magnitude of the effect varies between 1.1 and 2.4 percentage points per month across studies. In contrast, showing subjects return bar charts as opposed to price level line charts results in lower expectations. Across studies the magnitude of the effect varies between 1.7 and 1.0 percentage points per month.

Examining the difference between these particular formats is important as they are used extensively and typically considered to be interchangeable in the real world. Price line charts are widely used as a standard on the platforms of major information providers of both institutional investors (e.g. Bloomberg) and retail investors (e.g. Yahoo! Finance). In contrast, the European Undertakings for Collective Investments in Transferable Securities Directive requires information on the past performance of funds in Europe, embedded in Key Investor Information Documents (KIID), to be provided in the form of return bar charts (see European Commission, 2010, Chapter 3, Section 4). Similarly, expected future price levels and expected future returns constitute two elicitation formats which are widely used in real-world surveys. For instance, the Duke/CFO Magazine Business Outlook Survey enquires about expected returns, whereas the Livingston Survey of the Federal Reserve Bank of Philadelphia asks about expected price levels.

Several previous studies, which we will later discuss in greater detail, have examined the difference between the formats of price levels and returns on subjective expectations. However, they have reached opposing results. The differences of the results may be triggered by divergent study designs. For instance, Diacon and Hasseldine (2007) and Stössel and Meier (2015) focus on the effect of the chart format and a mixture of chart formats by holding the question format constant. In contrast, Glaser, Langer, Reynders, and Weber (2007) examine the effect of the question format by holding the chart format constant. Our main contributions can be summarized as follows: Firstly, we resolve the apparent contradictions by separating the effect of the chart format (i.e. return bar chart versus price line chart) from the effect of the question format (i.e. asking for expected price levels versus expected returns). To this end, in two laboratory studies we vary the format of the chart and the format of the question in a 2×2 design. Secondly, we isolate the impact of the chart format from potential confounding factors, which may have influenced the results in previous studies, such as the data frequency of the chart. Previous studies have used a different data frequency for the two different chart formats - high frequency data for price line charts and low frequency data for return bar charts. Thirdly, we test the boundaries of the subjects' susceptibility to format changes by examining it under different conditions: different level of expertise (i.e. students in the lab versus finance professionals in their real-world environment); different amounts of information (i.e. only historical chart information versus real-world information); different incentive schemes (i.e. fixed versus performance-based incentives). Fourthly, we distinguish between potential explanations of how the question format and the chart format affect the process of investors' expectation formation.

Our further results can be summarized as follows. We show that the sizeable effects of the chart format and the question format are neither mitigated by performance-based incentives, nor by expertise, nor by real-world information. We find that the chart format and the question format have a larger effect if subjects score low on Cognitive Reflection, indicating that they tend to rely on their intuitive judgment instead of reflecting upon their answers. We list several characteristics of the intuitive number sense and discuss how they can explain our findings. Furthermore, changes in the format also affect which past returns matter for the subjective expectations. In line with the existing empirical evidence (see, e.g., Greenwood and Shleifer, 2014), subjects extrapolate from past returns in all treatments. However, when they are shown price line charts, they extrapolate only from the most recent past. In contrast, when they are shown return charts and asked to think about returns, they take the entire available past information into account. Hence, the format design affects what subjects make of the available data. Taken together, our results imply that the format design in financial markets can be used as a powerful tool to alter investors' expectations. Information platforms and financial advisors could thus either willingly or unwillingly manipulate investors' expectations by varying the format through which their investors perceive the past information and think about the future. Understanding how the information presentation affects investors is of utmost importance against the backdrop of the rising importance of automated robo advisors. They are set out to replace the human advisor by providing unbiased and accessible financial advice especially to less wealthy investors (see, e.g., Financial Conduct Authority, 2016). Robo advisors only "communicate" with their clients by means of quantitative and graphical information, hence understanding how format changes contribute to their clients' perception of the information is crucial for the success of their advice.

The remainder of this paper proceeds as follows: Section 2 outlines the experimental design of the three studies. Section 3 displays and discusses our results on the differences between the two chart formats and the two question formats. Section 4 examines Cognitive Reflection as a mediator of the main treatment effects and discusses potential explanations. Section 5 uncovers how the different formats affect which past returns investors focus on. Finally, Section 6 concludes.

2 Study Design

We present three studies, which are designed to (i) separate the impact of the two chart formats (hereafter referred to as *stimuli*) from the impact of the two question formats (hereafter referred to as *tasks*), (ii) isolate the impact of the format from confounding factors, (iii) test the boundaries of the two main effects and (iv) unveil potential explanations for their occurrence. Study 1 and 2 were conducted in April 2015 and November/December 2015 at the mLab laboratory of the University of Mannheim (Germany) using the experimental software z-Tree (Fischbacher, 2007). Study 3 was conducted online in the period from September 2012 until June 2015. Our study group for Study 3 consists of finance professionals from the pool of participants of a real-world financial market survey conducted by the Centre for European Economic Research (ZEW). The ZEW Financial Market Survey is conducted since 1991 among financial market professionals from leading financial institutions, such as banks and insurance companies, and large industrial companies in Germany. The survey has gained considerable market attention since its establishment. The monthly press releases of the survey results trigger a price reaction in highly liquid financial markets such as the German stock and bond market (see, e.g., Entorf, Gross, and Steiner, 2012). This indicates that investors expect that the survey responses contain valuable information and are not driven by irrelevant factors such as the question format.

In order to separate the two main effects - the effect of the two *stimuli* from the effect of the two *tasks* - in Study 1 and Study 2 we use a 2×2 experimental design, in which we vary the *stimuli* (i.e. price line charts versus return bar charts) and the *tasks* (asking for price levels versus asking for returns). Price line charts and return bar charts used in previous studies often differ in aspects that go beyond the format. Typically, return bar charts display the data at a lower frequency compared to price line charts (see, e.g., Glaser, Langer, Reynders, and Weber, 2007; Diacon and Hasseldine, 2007), which makes the frequency of the information a potential confounding factor in previous results. We therefore isolate the effect of the chart format from the effect of the data frequency by holding the data frequency constant. Furthermore, we isolate the impact of random patterns in the data. To this end, we show each group of four subjects (one subject from each treatment) a different sequence of randomly generated charts. Moreover, we isolate the well-documented impact of the scale of the y-axis (see, e.g., Lawrence and O'Connor,

1993) by holding the y-axis constant for each subject. In order to test the boundaries of the two main effects we examine them under different conditions - under varying levels of expertise, under varying amounts of information, and under different incentive schemes as illustrated in Figure 1.

[Insert Figure 1 about here]

Existing literature alleges that irrelevant factors should not affect beliefs and choices when the subjects are presented with large amounts of information or in case they are experienced in the task (Schoorman, Mayer, Douglas, and Hetrick, 1994). Several previous studies document opposing evidence. For instance, finance professionals are evidently affected by irrelevant anchors (Kaustia, Alho, and Puttonen, 2008), by their mood (Goetzmann, Kim, Kumar, and Wang, 2015) and by priming (Gilad and Kliger, 2008). Haigh and List (2005) and Sarin and Weber (1993) show that traders on the Chicago Board of Trade exhibit myopic loss aversion and ambiguity aversion comparable to students. Simmons and Nelson (2006) suggest that experts are more susceptible to irrelevant factors than novices because the latter do not have any prior knowledge and thus have no choice but to think through their beliefs and choices deliberately. In Study 3 we analyze the effect the question format has on professional forecasters from the panel of the ZEW Financial Market Survey. They are highly experienced in the task - forecasting the German stock market performance index DAX - and have a realistically large amount of information at their disposal.

In our baseline study - Study 1 - the subjects are provided with fixed remuneration. Previous literature shows that performance-based financial incentives are not an effective remedy against rationality violations in experiments (Camerer and Hogarth, 1999) and can even have a detrimental effect on the exerted effort (Ariely, Gneezy, Loewenstein, and Mazar, 2009). Nevertheless, it can be argued that the influence of irrelevant factors such as the format of the chart and the question (i.e. framing effects) might disappear in the presence of performance-based incentives. In order to test this hypothesis, in Study 2 we introduce a performance-based incentive scheme. Furthermore, in order to unveil potential explanations of the main effects, in Study 2 we extend the design of Study 1 as described in detail below.

In the following we describe the subjects and the design of the three studies in detail. A chronological overview of the setup of the laboratory studies as well as the instructions and questionnaires of the three studies are provided in the Online Appendix.

2.1 Subjects

Our dataset consists of 179 (Study 1) and 169 (Study 2) subjects from the pool of the mLab laboratory for economic experiments at the University of Mannheim.[1] Study 3 was conducted online with 212 subjects from the pool of participants of the ZEW Financial Market Survey. The subjects are a heterogeneous group of financial market practitioners: active (e.g., portfolio managers) and passive (e.g., professional forecasters); buy-side participants from Treasury departments in large German companies and sell-side participants such as asset analysts and investment advisors. Table 1 shows that the subjects in Study 3 are significantly older compared to the subjects from the laboratory Studies 1 and 2. The average subject in Study 1 and Study 2 is 23 years old. The subjects in Study 1 and Study 2 are similar based on almost all elicited demographic variables. There are fewer subjects in Study 2 who have completed a statistics course, which can be explained by the fact that the two laboratory experiments were conducted in different semesters. The average subject in Study 3 is 47 years old, having 23 years of experience in the finance industry. In contrast to Studies 1 and 2, the subjects in Study 3 are almost exclusively male, which is characteristic for the finance industry.

[Insert Table 1 about here]

[1] Overall 184 and 170 subjects participated in Study 1 and Study 2 respectively, but we excluded five subjects from Study 1 and one subject from Study 2 for failing to respond in the assigned format, as explained in Section 2.4.

2.2 Studies 1 and 2: Forecasting Based on Charts of Past Performance

Main Part

The main parts of Study 1 and Study 2 consist of forecasting the future development of a sequence of financial market instruments based on charts of their past performance. In a 2×2 design we test the difference between two *stimuli* - price line charts and return bar charts - and two *tasks* - forecasting the future price level and forecasting the future return. The data used for the charts shown to each subject is simulated individually for each group of four subjects - one subject from each one of the four treatments. This way we ensure that the aggregate data is the same in each treatment. At the same time, we ensure that our results are not driven by random patterns in the data because within a particular treatment each subject is exposed to different data sequences. In Study 1 and Study 2 we use overall 450 and 860 distinct simulated data sequences. Each data sequence comprises 12 random draws from the return distribution of an artificial asset, which corresponds to a chart of one year of monthly performance. In Study 1 (Study 2) each subject was shown a sequence of 10 (20) charts. Half of the charts for each subject were drawn from a normal distribution with a positive mean, $N(\mu, \sigma)$, and the other half were drawn from a normal distribution with a negative mean, $N(-\mu, \sigma)$, where $\mu = 0.014$ and $\sigma = 0.059$. The parameters μ and σ are taken from the sample distribution of the monthly returns of the German stock market total return index DAX 30 over a period of five years.

Study 2 was designed to closely replicate Study 1. However, in order to test alternative explanations of the main effects we introduce the following modifications: Firstly, we introduce performance-based incentives. In Study 1 participants receive a fixed remuneration of 4 euros. In Study 2, the remuneration depends on the forecast accuracy. Following completion of the main task, we randomly draw one asset which determines the subject's remuneration. The subjects were told that they will earn 21 euros if one of their randomly drawn forecasts are "reasonably close" to the true realization and 6 euros otherwise. We used a verbal description of the accuracy interval instead of a numerical description in order to avoid differences in the format of the numerical intervals contaminating the results. All subjects whose absolute error was lower than the absolute error of 80% of the subjects in the respective session earned 21 euros. The remaining subjects earned 6 euros. The average remuneration in Study 2 is higher than in Study 1 because the main part of Study 2 is twice as long (20 assets as opposed to 10 assets). The main part of Study 1 took on average 6.1 minutes to complete and the main part of Study 2 took on average 17.4 minutes to complete. Secondly, we extend the forecasting task by a directional question. Subjects are asked to assess on a three-point Likert-type scale whether they expect the return to be roughly positive/zero/negative (for *task return* treatment) or roughly higher/equal/lower than the current price level (for *task price* treatment). The question was elicited before the respective numerical forecast (see Appendix B.1). Thirdly, we vary the starting price of the charts. In Study 1 all price line charts in the *stimulus price* treatment are generated based on a starting price of 100 monetary units. In Study 2 the charts for half of the subjects start from a price level of 100 monetary units. The other half of the subjects obtain charts, starting from a price level of 1000 monetary units. This variation is necessary in order to distinguish between potential explanations as elaborated in Section 4.2. Both laboratory studies conclude with a questionnaire eliciting demographic characteristics: age, gender, interest and experience in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course (see Appendix B.2 for a full list). We also measure a Cognitive Reflection Test score (Frederick, 2005) and a Numeracy score (Cokely, Galesic, Schulz, and Ghazal, 2012) as elaborated in Section 4.1.

Between-Subject Treatments

In the two different *stimuli* the charts either display the simulated price level development, as illustrated in Figure 2, or the simulated return development, as illustrated in Figure 3. In both *stimuli* the scale of the y-axis is held constant for each subject. The selection of the highest and the lowest level displayed on the y-axis ensures that the range covers all observations among all charts displayed to the subject. Furthermore, we require the range to be wide enough so that at least the 90% confidence interval of the true return distribution is conceivable at the time of forecasting. The last price level is explicitly printed in the upper right corner of the charts in order to reduce the noise in estimating it. In Study 2 we reduce the salience of the return bar charts by widening the y-scale of the charts in the *stimulus return* treatment.

[INSERT FIGURE 2 ABOUT HERE]

[INSERT FIGURE 3 ABOUT HERE]

The two distinct *tasks* either ask subjects to report their expectation regarding the future price level, which we use to calculate the implicit return expectation, or ask subjects directly to report the future return they expect. Following Glaser, Langer, Reynders, and Weber (2007), we ask subjects in the treatment *task price* to report the median future price level and subjects in the treatment *task return* - to report the median future return. The exact wording of the two *tasks* is given as follows[2]:

Task Price: *Please provide a forecast of the future price level (in monetary units) of this financial market instrument one month from now. The actual realization of the future price level should lie with equal probability above or below your forecast.*

Task Return: *Please provide a forecast of the future return (in percent) of this financial market instrument over the next month. The actual realization of the future return should lie with equal probability above or below your forecast.*

In order to ensure that subjects understand the term "return" we include a definition which states that a return is defined as the percentage change of the price of the financial market instrument. The subjects were allowed to use pen and paper but they were not allowed to use calculators.

[2] The full versions of the instructions in Study 1 and Study 2 are provided in the Online Appendix. All studies were conducted in German language. The instructions in German are available upon request.

Control Task in Study 2

In Study 2 we include a control task subsequent to the main part. We ask subjects, who are assigned to the *stimulus price* treatment, to report their assessment of the past average monthly return in a given chart. We conduct the control task over a random selection of 10 charts out of the 20 charts which were previously shown to the subject over the course of the main part.

2.3 Study 3: Expert Forecasts with Real-World Information

Main Part

Study 3 differs from Studies 1 and 2 mainly in the high level of expertise of the subjects and the larger amount of information available to them. The professionals in Study 3 forecast in their day-to-day environment by making use of the information sources, which are typically available to them in the real world. Thus, in Study 3 we vary only the *tasks*. We ask the subjects to forecast the German stock market total return index DAX 30 (in the following "DAX") one month ahead. The Study consists of 12 waves in the period from September 2012 until June 2015. Figure 4 illustrates the timing and the duration of the individual waves. Over the course of the study the average monthly DAX return was 1.5% with a standard deviation of 0.045%. Thus, the sample distribution of the monthly DAX return was close to the distribution with a positive mean in Studies 1 and 2.

[Insert Figure 4 about here]

Prior to the first wave of Study 3 the subjects were randomly assigned to one of two between-subject treatment groups - a *task return* treatment, in which subjects are asked to forecast the future DAX return, and a *task price* treatment, in which subjects are asked to forecast the future DAX level. The initial assignment of the treatments remains unchanged over the course of the study. The exact wording of the two tasks is given as follows:

Task Price: *I expect the DAX in 1 month at ... points.*

Task Return: *Within 1 month I expect a DAX return (monthly percentage change) of ... percent.*

The exact wording of the questions for the two treatments are adopted from real-world surveys. The wording for the *task price* treatment is adopted from the ZEW Financial Market Survey itself. The exact wording for the *task return* treatment is adopted from the Duke/CFO Magazine Business Outlook Survey (see Ben-David, Graham, and Harvey, 2013).[3]

Attrition Rate and Internal Validity

In contrast to laboratory studies, online studies have the disadvantage that subjects can withhold their response or drop out (i.e. attrition) because of aspects of the treatment. Potential treatment-related attrition calls into question the internal validity in field studies. This is a relevant concern in Study 3 because the response rate to the main part is below 100%. In order to test for treatment-related attrition we compare subjects' responses to universal questions being part of the ZEW Financial Market Survey. Such questions are elicited outside the scope of the main part of our study. Hence they are the same for all subjects. Specifically, we compare the responses to three stock-market related questions - forecast of the DAX level 6 months ahead, directional forecast of the DAX 6 months ahead and the subjective perception of a current mispricing of the DAX.[4] We do not find any evidence of a difference between the responses of the two treatment groups to the three universal stock-market related question (see Appendix A.3 for a detailed description of the analysis).

[3] In order to specify that non-annualized monthly returns are required, the question for the monthly forecasting horizon includes a definition of monthly returns in brackets - "monthly percentage change". In addition, annual return expectations are included to Study 3 in September 2013 for the purposes of robustness checks.

[4] The exact wording of these questions is provided in the Online Appendix.

2.4 Data Preparation

In order to ensure that the subjects have read and understood the forecasting task, we perform consistency checks. Outliers in the experimental studies can arise from the subjects mistakenly responding in the format of the chart format instead of responding to the actual question which is asked. For instance, a subject who is asked to forecast returns and submits responses around 100, being in the range of the last price level, is likely to have only skimmed through the instructions. In order to prevent such outliers from affecting our regression coefficients, we exclude 5 (1) subjects from Study 1 (Study 2).[5] Overall, we exclude 3.6% of the observations in Study 1 and 2.5% of the observations in Study 2. In Study 3 we require that the return expectations are not unequivocally inconsistent with subjects' responses to other questions. For this purpose we use the subjects' responses to an additional question about the 90% confidence interval with respect to their DAX forecasts one month ahead. The subjective confidence intervals on the DAX one-month ahead forecasts were elicited in the format of the main forecasting question.[6] As a consistency check, we require that the subject's DAX forecast lies within the respective subjective 90% confidence interval. This criterion results in the exclusion of 17 observations submitted by 11 professionals, which make up 1.2% of the observations in Study 3.

3 Expectations Are Shaped by the Format of the Chart and the Question

In the following we examine whether different *tasks* and *stimuli* trigger different expectations. We show that asking subjects to forecast returns is not equivalent to asking them to forecast price levels. Similarly, showing subjects return bar charts induces different expectations compared to price line charts.

[Insert Figure 5 about here]

[5] For subjects, who initially respond in the wrong format and subsequently switch to the correct format, only the responses given in the wrong format are excluded.

[6] The exact wording of these questions is provided in the Online Appendix.

[Insert Table 2 about here]

These main effects are illustrated in Figure 5. The figure displays the average return expectations across treatments and studies. It shows that the *task return* treatment results in significantly higher expectations across stimuli and studies. In contrast, the *stimulus return* treatment induces lower expectations across tasks and studies.

Our empirical analysis of the main treatment effects is displayed in Table 2. We perform random effects panel regressions to account for subject-specific characteristics. Random effects are justified since the treatments were randomly assigned and are thus orthogonal to subject characteristics. Additionally, we control for diverse demographic characteristics as listed in Section 2.[7] Throughout the paper we report the results from the model specification with the largest number of control variables but all our results are qualitatively unchanged when demographics are fully or partly excluded. Table 2, Column 1 shows that asking subjects to forecast returns results in higher optimism by 2.4 percentage points per month (p-value < 0.000). Showing them return charts, however, results in higher pessimism by 1.7 percentage points per month (p-value < 0.000). Table 2, Column 2 shows that similar but slightly weaker treatment effects occur in the presence of performance-based incentives. The positive effect of the *task return* treatment amounts to 1.9 percentage points per month (p-value < 0.000). The negative effect of the *stimulus return* treatment amount to -1.0 percentage points per month (p-value $= 0.055$).[8]

As a robustness check, in Study 2 we make use of directional forecasts, which we elicit in addition to the numerical forecasts. It can be argued that subjects assigned to the *task return* treatments sometimes mean to submit a negative response but forget the minus sign, which could explain the positive effect of the *task return* treatment. In Table 2, Column 3 we exclude observations for which the directional forecast and the quantitative forecast are inconsistent. In particular, we require the sign of the quantitative forecast to be positive (negative) if the directional forecast points to a "positive" ("negative") return

[7] Appendix A.2 provides a list of the variables used in the empirical analysis throughout the paper.

[8] This paper focuses on the main treatment effects. An analysis of interactions between the *stimuli* and the *tasks* is also possible but we have not established any hypotheses in this regard.

or to an expected future price "higher" ("lower") than the current price. Following this restriction 3.9% of the observations are excluded. Table 2, Column 3 shows that the results from Study 2 are qualitatively unchanged in the restricted model, which means that the positive effect of the *task return* treatment cannot be explained by forgetting the minus sign. As a further robustness check, in Studies 1 and 2 we elicit quarterly return expectations in addition to the main forecasting task. Both treatment effects are robust to the alternative forecasting horizon.[9]

Table 2, Columns 4 and 5 show that the *task effect* also occurs in the real-world forecasts of professionals even though they have unrestricted access to real-world information sources and have experience with the forecasting task. We include treatment-wave fixed effects to account for the impact of time-series variables (such as the past DAX development and macroeconomic variables) as well as possible differences in the impact of time-series variables on the expectations in the different treatments. Table 2, Column 4 shows that the *task return* treatment induces higher return expectations by 1.1 percentage points per month (p-value < 0.000). Our results are qualitatively unchanged when the wave-fixed effects are excluded.

It can be argued that forecasting the short-term development of a real-world index is very similar to forecasting in the laboratory because investors in the real world might rely on past performance charts for lack of fundamental information available at high frequency (Menkhoff, 2010). Therefore, in September 2013 we extended the main part of Study 3 by a question on annual expectations which was repeated in 8 survey waves in the period from September 2013 until June 2015. Table 2, Column 5 analyzes the treatment effect in the annual return expectations. The positive effect of the *task return* treatment is as high as 1.8 percentage points p.a. (p-value $= 0.079$).

Is the difference between the format of price levels and the format of returns a survey artefact or is it relevant for the real world? Previous studies have shown that the survey-based expectations of professionals are related to their actual behavior. This has been

[9] The results of all robustness checks mentioned in the paper are displayed in Appendix A, unless stated otherwise.

shown by Gennaioli, Ma, and Shleifer (2016) and Ben-David, Graham, and Harvey (2013) for the survey-based expectations of Chief Financial Officers and a variety of corporate decisions. The majority of finance professionals in our sample in Study 3 do not make active decisions but rather act as opinion makers in their role as economists, stock market analysts, financial advisors to name a few. As opinion makers they have a high impact on the actual behavior of others through their participation in the ZEW survey, as indicated by the significant price reaction triggered by the release of the ZEW survey results on highly liquid financial markets (see, e.g., Entorf, Gross, and Steiner, 2012, for results of event studies). In the following we further argue that the *task effect* is unlikely to be restricted to the survey-based expectations of those professionals but is rather likely to occur also in their everyday forecasts outside the scope of Study 3. To this end we analyze the *task effect* in a subset of the professionals from Study 3, (i) who act as opinion makers, meaning that they conduct DAX forecasts on a regular basis as part of their occupational activity, and (ii) whose usual forecast format is close to the forecast format they are assigned to in Study 3.[10] The *task effect* is robust in both the monthly and the annual return expectations of this subset of professionals. Hence, we conclude that the *task effect* is likely to also occur in the real-world forecasts of those opinion makers.

Our results of a negative effect of the *stimulus return* treatment contradict the findings from previous experimental studies. Diacon and Hasseldine (2007) do not find the chart format to have any significant effect on the subjective return expectations. Stössel and Meier (2015) document significantly stronger optimism when subjects are shown return bar charts while simultaneously being asked to forecast returns. Glaser, Langer, Reynders, and Weber (2007) show that the *task effect* depends on the historical trend. According to this study, subjects report higher expectations in the *task return* treatment if and only if the displayed chart covers a period of positive past performance. In a robustness check, we analyze whether the main treatment effects depend on the sign of the annual past performance and we do not find a significant difference at a 10% level. The difference

[10] We elicit supplementary information on the usual forecast format and use it to categorize the professionals based on proximity of the *task format* to their usual forecast format. The exact wording of the question and our categorization are provided in the Online Appendix. The regression results based on the subset of professionals are shown in Table A4.

between our results and the previous studies could be explained by several differences in the study design. Firstly, previous studies have used real-world data and have partly displayed the real names of the assets (see, e.g., Glaser, Langer, Reynders, and Weber, 2007). Secondly, previous results can be affected by confounding factors such as a different data frequency in the different *stimuli* (see, e.g., Diacon and Hasseldine, 2007; Stössel and Meier, 2015).

Regarding the *stimulus effect* it can be hypothesized that the difference in expectations is driven by the difference between a *geometric average* and an *arithmetic average* of the past returns. In particular, one can argue that return bar charts enable a quick estimation of the arithmetic average whereas an estimation of the geometric average is facilitated by price line charts. However, the difference between the geometric and the arithmetic average past return cannot explain our evidence because it predicts an opposite sign of the *stimulus effect*. Since the *stimulus return* supposedly induces arithmetic averaging and the arithmetic average is equal or higher than the geometric average, *stimulus return* should be connected to higher expectations, which does not hold true in our results. Still, the different formats might induce different averaging but its small effect might be offset by other factors. Given the data generating process in Studies 1 and 2, the predicted positive *stimulus effect* is 0.16 percentage points per month.

In Studies 1 and 2 we focus explicitly on forecasting based on charts (i.e. judgmental forecasting).[11] Examining judgmental forecasting is important because investors rely heavily on charts of past performance in the real worlds, in spite of other information sources being available. This has been shown for both retail and institutional investors (see, e.g., Greenwood and Shleifer, 2014; Amromin and Sharpe, 2014). The consistency of our results from studies 1 and 2 compared with the results from Study 3, where subjects have real-world information at their disposal, supports this notion.

[11] Given that only past performance information is provided to the subjects the optimal strategy to forecast the median future return without any prior information would be to report the sample median return over the period covered by the chart.

In summary, we show that the chart format and the question format can induce heterogeneity of subjective expectations in financial markets in a predictable way. Our evidence suggests that having practical experience in financial markets is not a useful remedy since professionals with decades of experience in the finance industry are not immune to the effect of the question format either. In Section 4 we provide a possible explanation as to why this is the case.

4 Unveiling How Prices and Returns Affect Expectations

4.1 The Role of Intuitive Thinking

In Section 3 we show that having experience does not protect against one's susceptibility to format changes. Even professionals with decades of experience in the finance industry are not immune. One reason might be because experience misleads subjects in trusting their impulsive, intuitive judgment instead of deriving a deliberate, analytical solution to the task. In the following we show that a subject's tendency to reflect and deliberate mitigates both the *stimulus effect* and the *task effect*.

Psychological literature typically ascribes the impact of irrelevant aspects of the decision situation to the influence of impulsive, intuitive thinking (Kahneman and Frederick, 2007). A strand of literature on dual processing distinguishes between at least two processes of thought. On the one hand there is an effortless and impulsive intuitive thinking, labeled System 1 (see, e.g., Kahneman, 2011; Stanovich and West, 2002). It happens subconsciously, requires prior knowledge and experience with the task, and is slow to change in case it makes systematic mistakes. On the other hand there is a slower, deliberate, reflective and effortful reasoning, labeled System 2. The tendency to think intuitively or deliberately can be thought of as a cognitive style, in which case it is measured either by means of self-assessment (e.g., Faith-in-Intuition score by Epstein, Pacini, Denes-Raj, and Heier, 1996) or by tracking intuitive mistakes (see, e.g., Frederick, 2005). For the purposes of the laboratory experiments we use the 3-items Cognitive Reflection Test by Frederick (2005) and the additional 4 items recently proposed by Toplak, West, and

Stanovich (2014).

The Cognitive Reflection Test (hereafter referred to as CRT) constitutes an inventory of questions for which an intuitive but wrong answer quickly comes to mind. Subjects differ in their tendency to question the initial intuitive answer and as a result to deliberately find the correct solution - a trait which is called cognitive reflection. Previous studies relate high Cognitive Reflection to immunity against the valence framing effect (Frederick, 2005) and consistency in judgments and decisions (Cokely and Kelley, 2009). We choose the Cognitive Reflection Test as it is not malleable by social desirability - the subjects' willingness to look good in the test. As the measure is hardly applicable in studies with professionals, in Study 3 we elicit subjects' self-assessment on their reliance on intuitive thinking.[12] As a second-best we use a self-assessment proxy for intuitive thinking style, which reflects the question with the highest loading on the Faith-in-Intuition score (see, e.g., Epstein, Pacini, Denes-Raj, and Heier, 1996). We follow Sjöberg (2003) in adapting the original question to address the domain of short-term stock market forecasting. We compose a measure of the subjective Priority of Analytical Tools (PAT), which compares the subjects' tendency to rely on intuitive thinking when forecasting the DAX 1 month ahead with their tendency to rely on analytical forecasting tools, in particular econometric methods, simulations, technical analysis and fundamental analysis.[13]

[Insert Table 3 about here]

Table 3 shows that a stronger tendency to reflect and deliberately reach a solution, as measured by the CRT score and the PAT measure, diminishes the scope of both the *task effect* and the *stimulus effect*. This result holds for all studies and for all applied measures of intuitive thinking. In Table 3, Columns 1-4, we analyze the impact of intuitive thinking

[12] The CRT questions are designed for experiments with students. Similarly to Ben-David, Graham, and Harvey (2013) we were worried that asking professionals questions, that may appear too simplistic to them, may lead them to abandon the survey.

[13] The exact wording of the question was the following (translated from German):"How important are following factors for your short-term (1 month ahead) DAX forecasts?". The response categories were given as follows: technical analysis, fundamental analysis, econometric models, simulations, intuition. The individual categories were rated on a three-point scale from "low" to "high".

in Studies 1 and 2. We perform random effects regressions of the subjective return expectations as a dependent variable. The main independent variables are treatment dummies, a respective measure of intuitive thinking, centered around its median, and its interactions with the treatment dummies. Our results show that one additional correct response to the full 7-items CRT diminishes the positive effect of the *task return* by 0.6 (0.9) percentage points in Study 1 (Study 2). This constitutes a reduction by 26 percent (56 percent) compared to the scope of the *task effect* for a subject with a median CRT score. Analogously, one additional correct response to the full 7-items CRT diminishes the negative effect of the *stimulus return* treatment by 0.5 (0.9) percentage points. The results are similar for the new 4-items CRT. In Table 3, Column 5, we analyze the impact of intuitive thinking on the *task effect* in Study 3. We conduct subject-random effects regressions with treatment-wave fixed effects. The main independent variables are a *task return* dummy, the PAT measure and an interaction thereof. Our results show that a higher relative self-reported reliance on analytical tools as opposed to intuition diminishes the positive effect of the *task return* treatment (p-value = 0.095).

Since the CRT consists of innately mathematical tasks, it can be argued that the dampening effect of a high CRT score on the main treatment effects is due to better mathematical skills rather than a stronger tendency to rely on analytical thinking. Frederick (2005) reports a positive correlation between the CRT score and mathematical skills. In order to distinguish between analytical thinking and mathematical skills, in a robustness check we additionally control for the impact of mathematical skills on the scope of the two treatment effects. To this end, in Studies 1 and 2 we elicit a measure of advanced numeracy as given by the Berlin Numeracy Test described by Cokely, Galesic, Schulz, and Ghazal (2012). We extend the empirical analysis from Table 3 by including the measure of advanced numeracy and an interaction of numeracy with the two treatment dummies. Our results on the dampening effect of the CRT on the *task effect* and the *stimulus effect* are qualitatively unchanged when controlling for the effect of advanced numeracy. In contrast, advanced numeracy does not have any significant effect (at the 10% level) as a mediator of the two main effects.

4.2 Intuition and Numbers

From the results outlined in Section 3 it becomes apparent that the *task effect* and the *stimulus effect* are driven by the treatments in which the chart is displayed in a different format from the format of the question (hereafter referred to as mixed-format treatments). Mixed-format treatments are relevant in the real world because investors often have limited influence on the format of the charts. For instance a retail investor, who estimates the future value of an investment in a European fund based on the return bar charts displayed in the mandatory KIID, would find himself in a mixed-format forecasting task. A distinctive characteristic of the mixed-format treatments is that they require further mental calculations. Systematic mistakes in the mental calculations can serve as potential explanations of our results. In Section 4.1 we showed that our results are mediated by intuitive thinking, hence mental calculation mistakes which arise from the innate number intuition are particularly relevant for our setting (see Dehaene, 2011, for an overview). In the following we conduct an explorative analysis of several well-documented characteristics of the number intuition and discuss how they can explain the expectations in the mixed-format treatments and consequently - the *task effect* and the *stimulus effect*.

Negative Numbers Do Not Come Easily to the Intuitive Mind

Cognitive neuroscience documents that the human brain is evolutionary only equipped with an intuition for positive integers (see Dehaene, 2011, p.74). Our understanding of negative numbers relies on deliberately acquired mental models, in contrast to our intuitive understanding of positive numbers. The asymmetry in the way positive and negative numbers are processed affects each treatment differently and the difference might explain the *task effect*. In particular, subjects in the *task level* treatment can articulate negative expectations without having to use negative numbers, hence the burden to reporting pessimism is lower than in the *task return* treatment.

[Insert Table 4 about here]

Table 4 analyzes how the chart format and the question format are linked to the probability to report a negative return expectation. The table reports marginal fixed effects of probit regressions, where the dependent variable is a dummy variable which equals one for negative return expectations. Panel A shows that the likelihood of having negative expectations in Study 1 (Study 2) decreases by 18.5 (17.9) percentage points when subjects are asked to indicate it with a negative number (i.e. in *task return* treatment). In Panel B we calculate the marginal effects separately for subjects with below-median and above-median CRT score. Panel B shows that the effect is mainly driven by subjects who rely more strongly on their intuitive thinking (i.e. subjects with below-median CRT score). The decrease of the likelihood for reporting a negative return is as high as 28.9 (22.2) percentage points for intuitive subjects.

Table 4 further reports two robustness checks in order to isolate trivial reasons for misreporting of negative numbers. In particular, we need to make sure we do not underestimate the ratio of negative return expectations in the *task return* treatments for trivial reasons, for instance subjects not having understood that a negative return should be marked with a negative sign. Firstly, we exclude subjects, who never report a negative number in the *task return* treatment (see Columns 2 and 5 for Study 1 and 2 respectively). Secondly, in Study 2 we require that the sign of the numeric forecast is consistent with the sign of the additionally elicited directional forecast (see Column 4). Our results remain qualitatively unchanged across all specifications.

[Insert Table 5 about here]

To what extent can the lower likelihood to report negative numbers explain the positive effect of the *task return* treatment? The exact magnitude depends on the assumption about what numbers subjects report instead of reporting a negative number. Since the reluctance to state negative numbers is stronger for subjects who rely more strongly on intuitive thinking, it is plausible to assume those subjects replace negative numbers with intuitive numbers. We draw on a well-documented result from cognitive psychology and assume that the intuitive mind jumps to the intuitive small positive integers 1, 2 and 3 (see, e.g., Hyde and Spelke, 2008).[14] We calculate the hypothetical return expectations of the subject in the *task return* treatment assuming that they are reluctant to state negative numbers. In particular, we take the chart median return as a starting point of the hypothetical expectation formation process and we replace 30% of the initially negative return expectations with one of the intuitive numbers - 1,2 or 3 (with equal probability).[15] Table 5 shows that the reluctance to state negative numbers explains an optimism bias in the hypothetical return expectations of the *task return* treatment, which amounts to 0.6 percentage points (see *negative numbers reluctance*). This magnitude is lower than the magnitude of the *task effect*, but further mental calculation mistakes can play an amplifying role.

Naïve Compounding

Number intuition is limited in its ability to support more complex arithmetics. In order to assess the sample average past return from a price line chart subjects need to calculate the 12th root of the annual gross return, which is outside the power of the arithmetic operations supported by our innate number sense. Instead of calculating the 12th root of the annual gross return, subjects might simply divide the annual gross return by 12 (hereafter referred to as *naïve compounding*). Naïve compounding leads to a systematic overestimation of the average past returns, hence it might partly explain the positive effect of the *task return*.

In order to obtain a model-free first impression whether naïve compounding can affect the responses of subject who are asked to forecast returns from price charts, we explore the sample distribution of the return expectations. In our laboratory studies, the prices in the last period are positively skewed but the monthly returns are not. Hence a correct calculation of the average monthly return of the chart should result in a symmetric distribution of expected returns. In contrast, a naïve compounding predicts a positively

14 Consistent with the finding of Hyde and Spelke (2008), the small positive integers 1-3 make up 24.5% of the responses when subjects are asked to forecast returns while being shown price line charts.

15 Replacing 30% of the negative numbers with a positive number will result in a 15 percentage points lower ratio of negative numbers, which corresponds to the regression results from Table 4.

skewed sample distribution of expected returns. Figure 6 shows that the sample distribution of expected returns is positively skewed ($\upsilon = 2.177$ in Study 1 and $\upsilon = 3.307$ in Study 2) when subjects are asked to forecast returns from price line charts (p-value from D'Agostino-Pearson test < 0.000 for both Study 1 and Study 2). The positive skewness in the sample distribution of expected returns cannot be explained by the subjects' reluctance to state negative numbers (see Table 5).

[Insert Figure 6 about here]

In the next step, we look for traces of naïve compounding in subjects' perception of the average past return based on price charts. To this end, in Study 2 we elicit the perceived past average return from subjects in the *stimulus price* treatments. Table 6 analyzes whether the perceived past returns are affected by naïve compounding. As naïve compounding generates returns, which are highly correlated with the actual past average returns, we need to control for the latter. Table 6 shows that the hypothetical perceived returns, predicted by naïve compounding, explain subjects' perception of the average past returns even when controlling for the actual average past return. Our result is independent of the way we calculate the actual average return (i.e. arithmetic versus geometric average).

[Insert Table 6 about here]

To what extent can na ïve compounding explain the positive effect of the *task return* treatment? Table 5 shows that the distribution of hypothetical return expectations, predicted by naïve compounding, exhibits a positive skewness of 0.645. Hence, naïve compounding alone can explain 30% (20%) of the positive skewness in the distribution of expectations in the treatment *task return* & *stimulus price* in Study 1 (Study 2). Naïve compounding also predicts an optimism bias in the hypothetical return expectations of the *task return* treatment, although of very small magnitude (0.1 percentage points).

Other Potential Explanations

When thinking about numbers in any positive range, we overweight those numbers which are close to the lower bound of the range. Peters, Slovic, Västfjäll, and Mertz (2008) show that this bias towards small numbers is correlated with inconsistencies in economic decisions. In our setting, a bias towards small positive numbers might explain the lower return expectations when subjects are asked to forecast prices. The bias is more likely to occur in the mixed-format treatment, where subjects are shown return charts and asked to forecast prices, because in this case the mental urn is more important. Hence, it can potentially explain the *stimulus effect*. The bias further predicts negative outliers and consequently a negative skewness in the distribution of expected returns in this treatment. Figure 6 shows that when subjects are asked to forecast price levels from return charts the sample distribution of expected returns is negatively skewed ($\upsilon = -1.651$ in Study 1 and $\upsilon = -1.901$ in Study 2 with p-value from D'Agostino-Pearson test < 0.000 for both Study 1 and Study 2).

It can also be argued that in both mixed-format treatments subjects use the numbers they see on the chart as an *anchor* for their response, which they adjust insufficiently. The expectations in the mixed-format treatments are thus affected by the difference between the numerosity of the numbers on the y-axis of the chart and the numerosity of the required response. As an example, the presence of prices on the screen (i.e. large numbers) might induce too high expectations when subjects are asked to forecast returns (i.e. small numbers). The opposite effect can be expected for expectations submitted in the format of price levels (i.e. large numbers) in the presence of returns (i.e. small numbers) on the screen. Anchoring could thus explain both the *task effect* and the *stimulus effect*. In order to test the anchoring hypothesis, in Study 2 we introduce additional between-subject variation. The price data for half of the subjects is generated starting from a price of 100 monetary units and for the other half - starting from a price of 1000 monetary units. The anchoring hypothesis predicts that both the *task effect* and the *stimulus effect* are amplified if the prices are in a higher range. However, we do not find any evidence of a significant increase in the magnitude of the *task effect* nor the *stimulus effect* for subjects in the high-price-range condition.

5 The Format Affects Which Past Returns Matter

In Section 3 we show that the *task* and the *stimulus* systematically shift the subjective return expectations. In this section we analyze whether the format also changes the way subjects make sense of the available information on the past performance. In particular, we examine whether the treatments affect the focus of investors on recent performance (i.e. recency bias documented, e.g., by Greenwood and Shleifer, 2014). To this end, we examine the impact of past returns on the subjective return expectations.

[Insert Table 7 about here]

In Table 7 we show to what extent subjective return expectations depend on the past returns and how the different *stimuli* and the different *tasks* influence the link between past returns and expectations on the future. Table 7 shows the results of separate regressions for each between-subject treatment. The dependent variables are the logarithmized monthly expected returns in each treatment. The independent variables are the logarithmized average monthly returns as calculated over four non-overlapping periods. Since these regressions do not include the main treatment effects, it is possible to include subject-fixed effects instead of the demographic control variables. We report the results from fixed-effect regressions but our results are qualitatively unchanged when no control variables or only demographic control variables are included. Table 7 shows that subjects extrapolate from past returns across all treatments. This finding suggests that the empirically documented extrapolative behavior in the subjective return expectations of diverse groups of investors (see, e.g., Greenwood and Shleifer, 2014; Amromin and Sharpe, 2014) is robust to the exact format, in which the information is typically presented in the real world. This highlights the importance of taking extrapolative behavior into account when modeling subjective expectations (see, e.g., Barberis, Greenwood, Jin, and Shleifer, 2015, for a theoretical framework).

Table 7 further shows that the expectations in the different treatments also differ in the selection of past returns from which subjects extrapolate. Subjects who are asked to forecast price levels while being shown price line charts only extrapolate from the most recent quarter (see Panel A, left). The second-recent quarter of past returns does not have a significant influence on their return expectations at the 10% significance level. This result is consistent across both laboratory studies, which means that it is independent of the incentives scheme. In contrast, subjects who are shown return bar charts while being asked to forecast returns extrapolate from both recent and distant past returns. It should be noted that the past returns explain a larger part of the overall variation in this treatment compared to the other treatments, as indicated by the larger $R^2_{overall}$. Hence, the expectation formation process in this treatment can be reasonably well described as a simple calculation of the average return over the entire past period covered by the available chart. This evidence suggests that the way the information is presented can contribute to subjects' understanding of the data. A similar argument is made by Kaufmann, Weber, and Haisley (2013) who study the advantages of an alternative presentation format - experience sampling - for investors' understanding of financial information.

6 Conclusion

This paper compares the formats of prices and returns in the context of financial market expectations. We show that asking subjects to forecast returns as opposed to price levels results in higher expectations, whereas showing subjects return bar charts as opposed to price level line charts results in lower expectations. The effects of the question format and the chart format are mediated by the subjects' tendency to rely on intuitive thinking and can be partly explained by characteristics of the intuitive number sense. In addition, showing subjects price charts results in extrapolation of the recent past returns and ignorance of the more distance past returns. In contrast, showing subjects return bar charts and asking them to think about returns induces attention to the entire available data history.

Our results support the recent appeal of Camilleri and Larrick (2014) towards policy-makers to take into account format design as a choice architecture tool. It is particularly important for automated robo advisors, whose only way to "communicate" with their clients is by means of quantitative and graphical information. Regulation should also address the flexibility of financial advisors and information platforms to endogenously vary the format of the decision environment of their clients. Since the susceptibility to format changes is mediated by the involuntary impulses triggered by intuitive thinking and financial professionals are also not immune, having practical experience is unlikely to be a useful remedy. Therefore, future research should be devoted to examining potential remedies against subjects' susceptibility to changes of the format of the decision environment.

References

Amromin, G., and S. A. Sharpe (2014): "From the Horse's Mouth: Economic Conditions and Investor Epxectations of Risk and Return," *Management Science*, 60(4), 845–866.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar (2009): "Large Stakes and Big Mistakes," *Review of Economic Studies*, 76(2), 451–469.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015): "X-CAPM: An Extrapolative Capital Asset Pricing Model," *Journal of Financial Economics*, 115(1), 1–24.

Ben-David, I., J. R. Graham, and C. R. Harvey (2013): "Managerial Miscalibration," *Quarterly Journal of Economics*, 128(4), 1547–1584.

Camerer, C. F., and R. M. Hogarth (1999): "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty*, 19(1-3), 7–42.

Camilleri, A. R., and R. P. Larrick (2014): "Metric and Scale Design as Choice Architecture Tools," *Journal of Public Policy & Marketing*, 33(1), 108–125.

Cokely, E. T., M. Galesic, E. Schulz, and S. Ghazal (2012): "Measuring Risk Literacy: The Berlin Numeracy Test," *Judgment and Decision Making*, 7(1), 25–47.

Cokely, E. T., and C. M. Kelley (2009): "Cognitive Abilities and Superior Decision Making Under Risk: A Protocol Analysis and Process Model Evaluation," *Journal of Decision Making*, 4(1), 20–33.

Dehaene, S. (2011): *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press.

Diacon, S., and J. Hasseldine (2007): "Framing Effects and Risk Perception: The Effect of Prior Performance Presentation Format on Investment Fund Choice," *Journal of Economic Psychology*, 28(1), 31–52.

Entorf, H., A. Gross, and C. Steiner (2012): "Business Cycle Forecasts and Their Implications for High Frequency Stock Market Returns," *Journal of Forecasting*, 31(1), 1–14.

Epstein, S., R. Pacini, V. Denes-Raj, and H. Heier (1996): "Individual Differences in Intuitive-Experiential and Analytical-Rationall Thinking Styles," *Journal of Personality and Social Psychology*, 71(2), 390–405.

European Commission (2010): "Commission Regulation (EU) No 583/2010," *Official Journal of the European Union*, Available at: `http://eur-lex.europa.eu`.

Financial Conduct Authority (2016): "Financial Advice Market Review," *Quarterly Report of the Financial Conduct Authority*, March 2016, Available at: `https://www.fca.org.uk`.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 10, 171–178.

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19(4), 25–42.

Gennaioli, N., Y. Ma, and A. Shleifer (2016): "Expectations and Investment," *NBER Working Paper. 21260*, Available at: `http://www.nber.org/papers/w21260`.

Gilad, D., and D. Kliger (2008): "Priming the Risk Attitudes of Professionals in Financial Decision Making," *Review of Finance*, 12(3), 567–586.

Glaser, M., T. Langer, J. Reynders, and M. Weber (2007): "Framing Effects in Stock Market Forecasts: The Difference Between Asking for Prices and Asking for Returns," *Review of Finance*, 11(2), 325–357.

Goetzmann, W. N., D. Kim, A. Kumar, and Q. Wang (2015): "Weather-Induced Mood, Institutional Investors and Stock Returns," *Review of Financial Studies*, 28(1), 73–111.

Greenwood, R., and A. Shleifer (2014): "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 27(3), 714–746.

Haigh, M. S., and J. A. List (2005): "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis," *Journal of Finance*, 60(1), 523–534.

Hyde, D. C., and E. Spelke (2008): "All Numbers Are Not Equal: An Electrophysiological Investigation of Small and Large Number Representations," *Journal of Cognitive Neuroscience*, 21(6), 1039–1053.

Kahneman, D. (2011): *Thinking, Fast and Slow.* Farrar, Straus and Giroux, New York, 1st edn.

Kahneman, D., and S. Frederick (2007): "Frames and Brains: Elicitation and Control of Response Tendencies," *Trends in Cognitive Sciences*, 11(2), 45–46.

Kaufmann, C., M. Weber, and E. Haisley (2013): "The Role of Experience Sampling and Graphical Displays on One's Investment Risk Appetite," *Management Science*, 59(2), 323–340.

Kaustia, M., E. Alho, and V. Puttonen (2008): "How Much Does Expertise Reduce Behavioral Biases? The Case of Anchoring Effects in Stock Return Estimates," *Financial Management*, 37(3), 391–411.

Lawrence, M., and M. O'Connor (1993): "Scale, Variability, and the Calibration of Judgmental Prediction Intervals," *Organizational Behavior and Human Decision Processes*, 56(3), 441–458.

Menkhoff, L. (2010): "The Use of Technical Analysis by Fund Managers: International Evidence," *Journal of Banking and Finance*, 34(11), 2573–2586.

Meyler, A., and I. Rubene (2009): "Results of a Special Questionnaire for Participants on the ECB Survey of Professional Forecasters (SPF)," MPRA Archive.

Peters, E., P. Slovic, D. Västfjäll, and C. Mertz (2008): "Intuitive numbers guide decisions," *Judgment and Decision Making*, 3(8), 619–635.

Sarin, R. K., and M. Weber (1993): "The Effect of Ambiguity in Market Experiments," *Management Science*, 39(5), 602–615.

Schoorman, F., R. Mayer, C. Douglas, and C. Hetrick (1994): "Escalation of Commitment and the Framing Effect: An Empirical Investigation," *Journal of Applied Social Psychology*, 24(6), 509–528.

Simmons, J. P., and L. D. Nelson (2006): "Intuitive Confidence: Choosing Between Intuitive and Nonintuitive Alternatives," *Journal of Experimental Psychology: General*, 135(3), 409–428.

Sjöberg, L. (2003): "Intuitive vs. Analytical Decision Making: Which is Preferred?," *Scandinavian Journal of Management*, 19(1), 17–29.

Stanovich, K., and R. West (2002): "Individual Differences in Reasoning: Implications for the Rationality Debate?," in *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. by T.Gilovich, D. Griffin, and D. Kahneman, pp. 421–440. Cambridge University Press, Cambridge, United Kingdom.

Stössel, R., and A. Meier (2015): "Framing Effects and Risk Perception: Testing Graphical Representations of Risk for the KIID," *SSRN Working Paper*, Available at: http://papers.ssrn.com/abstract_id=2606615.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014): "Assessing Miserly Information Processing: An Expansion of the Cognitive Reflection Test," *Thinking & Reasoning*, 20(2), 147–168.

7 Tables and Figures

Figure 1: Study Design: Main Aspects

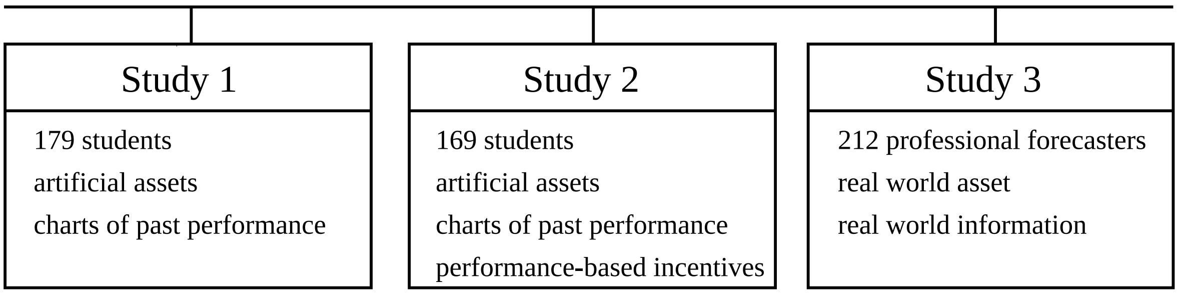


Figure 2: Stimulus Price Treatment (Exemplary Chart)

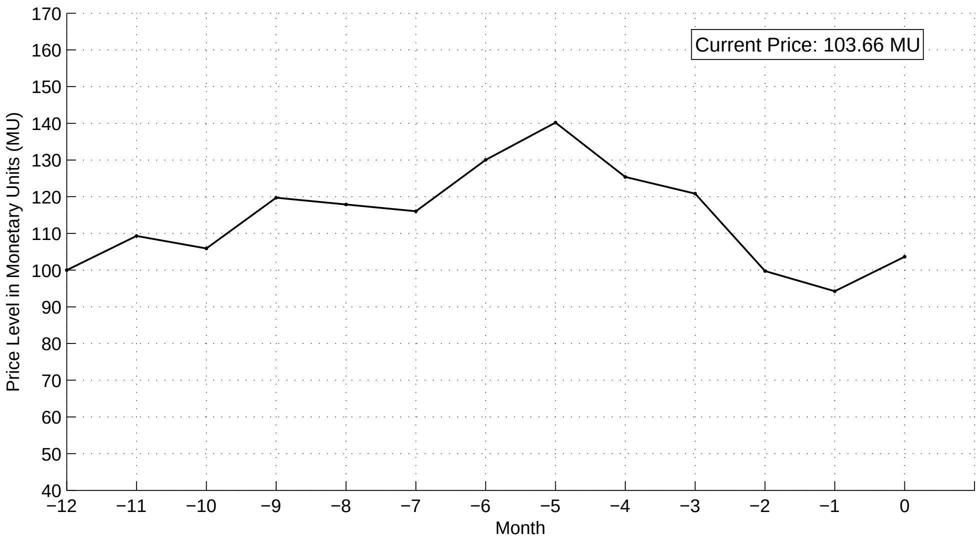


Figure 3: Stimulus Return Treatment (Exemplary Chart)

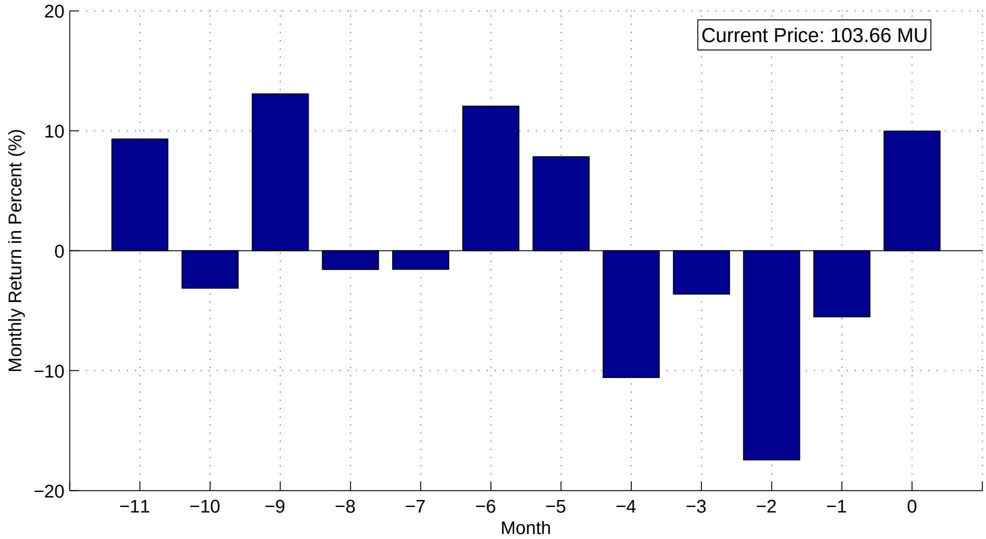


Figure 4: Timing and Duration of Study 3

This figure displays the development of the DAX over the period of Study 3. The study was repeated in 12 waves. In each wave the subjects had two weeks to respond. The shaded area illustrates the timing and the duration of the individual waves. Displayed are the respective deadlines for response submission in each individual wave.

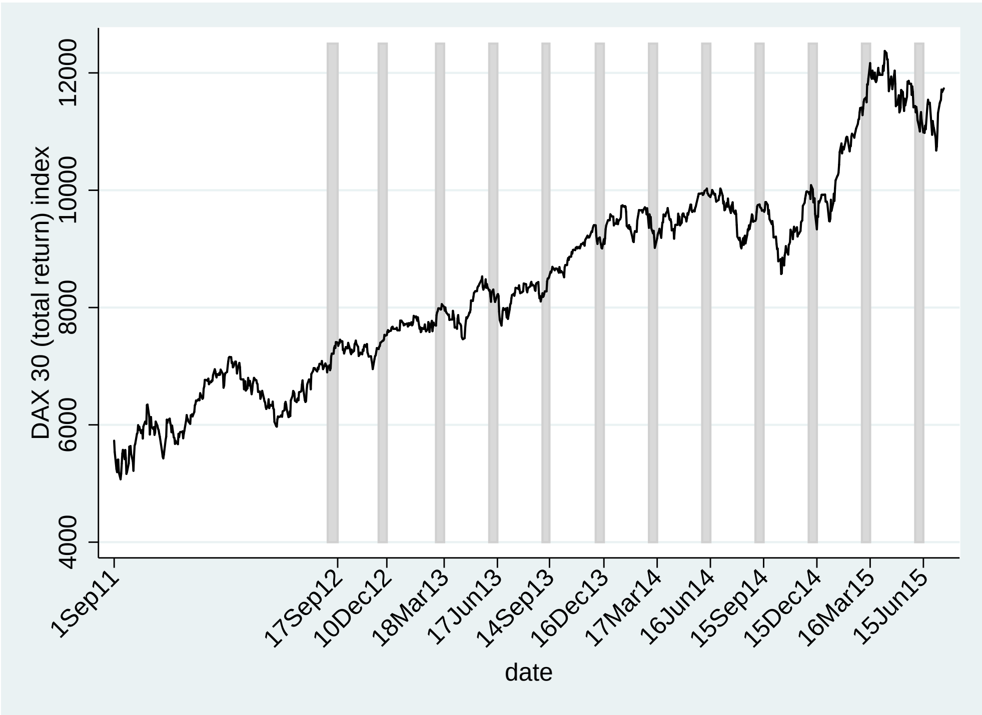
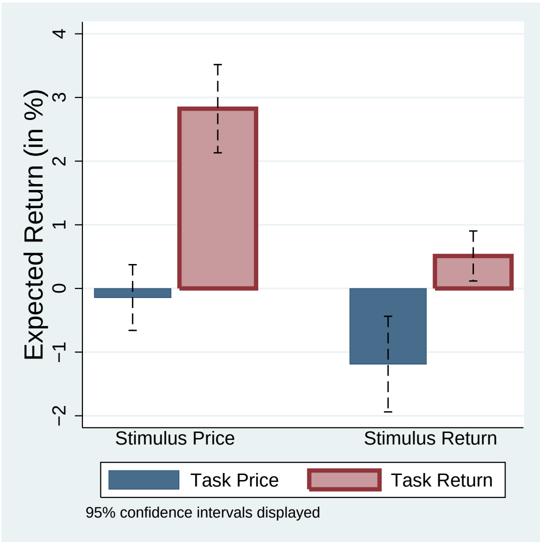
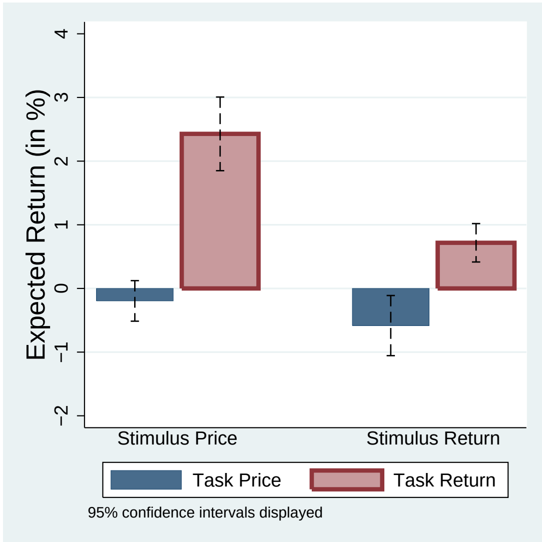


Figure 5: Main Effects of the Task and the Stimulus

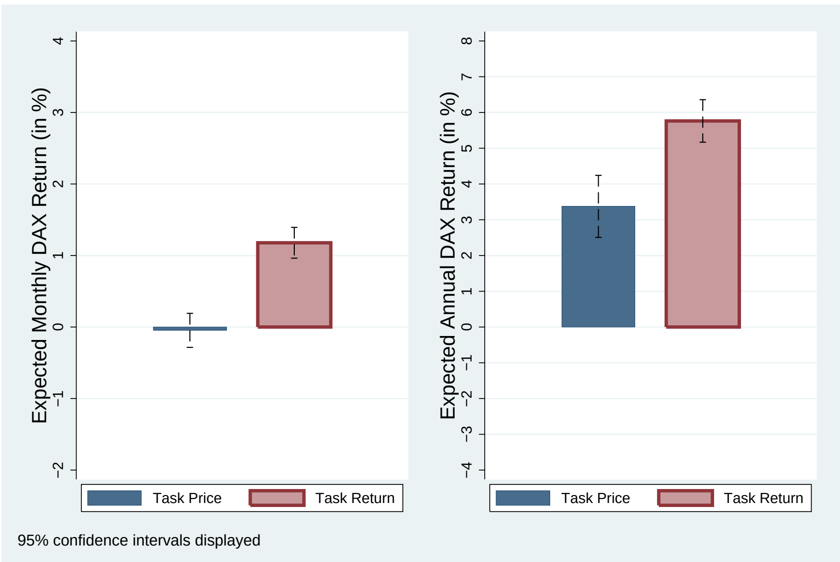
This figure displays the average return expectations (including 95% confidence intervals) across treatments and across studies. Expectations in the *task price* treatments in studies 1 and 2 are converted to return expectations using the last price level at the time of forecasting. DAX forecasts in the *task price* treatment in Study 3 are converted to return expectations using the last available DAX daily open level. For responses submitted on a bank holiday we use the last available DAX daily close level. Market data is downloaded from Datastream.



(a) Study 1



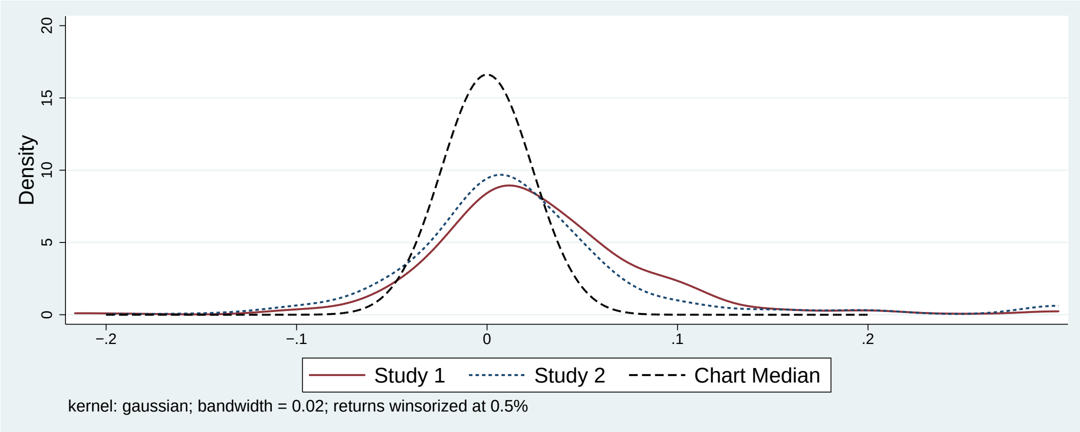
(b) Study 2



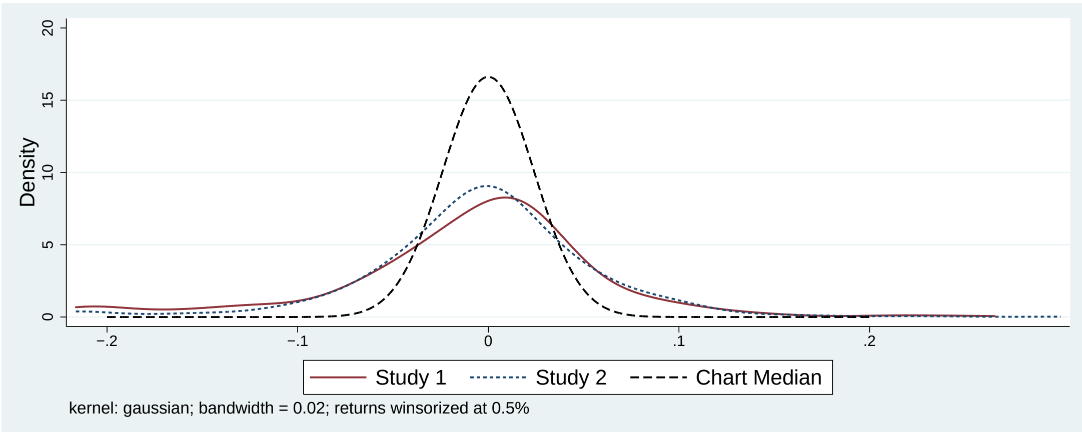
(c) Study 3

Figure 6: Distribution of Responses in Mixed-Format Treatments

This figure displays the sample distribution of the subjective expectations across the two mixed-format treatments. Each subfigure displays (i) a smoothed sample distribution of the expectations in Study 1, (ii) a smoothed sample distribution of the expectations in Study 2, (iii) the simulated distribution of the sample median returns across charts (i.e. chart median).



(a) Task Return & Stimulus Price



(b) Task Price & Stimulus Return

Table 1: Demographic Characteristics

This table reports an overview of demographic characteristics of the subjects in Studies 1, 2 and 3. Dummy variables are indicated in brackets as well as the number of categories of categorical variables. Columns (1)-(3) report sample averages. Column (4) provides the z-statistic of nonparametric Mann-Whitney tests comparing the two Laboratory Studies 1 and 2. Column (5) compares the average in Online Study 3 with the average among the two Laboratory Studies 1 and 2 based on Mann-Whitney tests.

Study:	1	2	3	Mann-Whitney Test (z)	
				$\mu_1 = \mu_2$	$\mu_3 = \bar{\mu}_{1,2}$
Number of Subjects	179	169	212	-	-
Age (years)	22.89	23.36	47.33	-0.508	-15.198
Female (D)	0.436	0.432	0.025	0.071	10.253
Career Experience (years)	-	-	23.09	-	-
PAT (-2-2)	-	-	0.833	-	-
Stock Market Interest (D)	0.525	0.527	-	-0.028	-
Stock Market Experience (D)	0.201	0.243	-	-0.931	-
Financial Markets Interest (D)	0.620	0.592	-	0.541	-
Financial Markets Experience (D)	0.134	0.148	-	-0.371	-
Statistics Course (D)	0.793	0.639	-	3.193	-
Behavioral Finance Course (D)	0.062	0.077	-	-0.568	-
Numeracy Score (0-4)	2.145	2.077	-	0.467	-
CRT4 Score (0-4)	2.704	2.704	-	-0.069	-
CRT7 Score (0-7)	4.760	4.675	-	0.155	-

Table 2: Effect of the Task and the Stimulus on Expectations

The dependent variables are the subjective return expectations from Studies 1, 2 and 3. The main independent variables are a dummy variable for the task return treatment ($D^{task=return}$) and a dummy variable for the stimulus return treatment ($D^{stimulus=return}$). A task treatment is included in all three studies whereas a stimulus treatment is included only in Studies 1 and 2 (laboratory). Columns (1)-(3) display the results of random-effects panel regressions. Since the treatments are randomly assigned and thus orthogonal to participant characteristics, random effects are justified. We control for demographic characteristics, given as follows: age, gender, experience and interest in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course, Numeracy score, Cognitive Reflection Test score (see Appendix B). Column (3) displays a robustness check: In Study 2 we elicit directional forecasts prior to the main forecasting task and we compare the sign of the directional forecast with the sign of the numerical forecast (main task). We exclude observations for which the directional forecast is inconsistent with the numerical forecast. Columns (4) and (5) display the results from Study 3 for monthly and annual expected DAX returns respectively. Included are treatment-wave fixed effects (FE) to account for a potentially different impact of time-series variables (such as the past DAX performance) on the expectations in the task price treatment and in the task return treatment. Standard errors are clustered at the subject level in all regressions.

	(1) $\mathbb{E}_t(R_{t+1})$	(2) $\mathbb{E}_t(R_{t+1})$	(3) $\mathbb{E}_t(R_{t+1})$	(4) $\mathbb{E}_t(R_{t+1}^{DAX})$	(5) $\mathbb{E}_t(R_{t+12}^{DAX})$
Study:	1	2	2	3	3
$D^{task=return}$	0.024***	0.019***	0.023***	0.011***	0.018*
	(5.972)	(3.664)	(3.653)	(3.943)	(1.753)
$D^{stimulus=return}$	-0.017***	-0.010*	-0.014**		
	(-3.908)	(-1.916)	(-2.253)		
Demographics	Yes	Yes	Yes	No	No
Treatment-Wave FE	-	-	-	Yes	Yes
$R^2_{overall}$	0.053	0.048	0.037	0.066	0.039
N	1773	3354	3222	1475	933

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 3: Impact of Intuitive Thinking on the Effects of the Task and the Stimulus

The dependent variables are the subjective monthly return expectations from Studies 1, 2 and 3 respectively. The main independent variables are a dummy variable for the task return treatment ($D^{task=return}$), a dummy variable for the stimulus return treatment ($D^{stimulus=return}$), a measure for analytical as opposed to intuitive thinking style, and an interaction between the latter and the treatment dummies. We use the following measures for analytical versus intuitive thinking style: In studies 1 and 2 we use the Cognitive Reflection Test score consisting of 4 items (i.e. CRT4) and 7 items (i.e. CRT7) items, respectively (see Toplak, West, and Stanovich, 2014) and centered around the median score. In Study 3 we use a measure of the Priority of Analytical Thinking (i.e. PAT), which captures the self-assessed relative importance of the most important deliberate forecasting tool compared to the importance of intuition for the purposes of solving the experimental task. In Studies 1 and 2 we control for the following demographic characteristics: age, gender, experience and interest in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course, Numeracy score. Standard errors are clustered at the subject level.

	(1) $\mathbb{E}_t(R_{t+1})$	(2) $\mathbb{E}_t(R_{t+1})$	(3) $\mathbb{E}_t(R_{t+1})$	(4) $\mathbb{E}_t(R_{t+1})$	(5) $\mathbb{E}_t(R_{t+1}^{DAX})$
Study:	1	1	2	2	3
$D^{task=return}$	0.023***	0.032***	0.016***	0.031***	0.016***
	(6.011)	(6.784)	(4.217)	(3.229)	(3.404)
$D^{stimulus=return}$	-0.017***	-0.024***	-0.007*	-0.018**	
	(-4.070)	(-4.700)	(-1.851)	(-2.182)	
$CRT7$	0.001		-0.003		
	(0.337)		(-1.360)		
$CRT7 \times D^{task=return}$	-0.006***		-0.009**		
	(-2.843)		(-2.377)		
$CRT7 \times D^{stimulus=return}$	0.005**		0.009**		
	(2.533)		(2.183)		
$CRT4$		0.000		-0.003	
		(0.180)		(-0.980)	
$CRT4 \times D^{task=return}$		-0.011***		-0.015**	
		(-3.718)		(-2.199)	
$CRT4 \times D^{stimulus=return}$		0.010***		0.012*	
		(3.107)		(1.890)	
PAT					0.004
					(1.303)
$PAT \times D^{task=return}$					-0.007*
					(-1.672)
Demographics	Yes	Yes	Yes	Yes	No
Treatment-Wave FE	-	-	-	-	Yes
$R^2_{overall}$	0.063	0.068	0.090	0.081	0.076
N	1773	1773	3354	3354	1043

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 4: Probability of a Negative Expected Return: Effect of the Task and the Stimulus

This table reports marginal fixed effects from probit regressions of the probability to forecast a negative return. The main independent variables are a dummy variable for the task return treatment ($D^{task=return}$) and a dummy variable for he stimulus return treatment ($D^{stimulus=return}$). We control for the following demographic characteristics: age, gender, experience and interest in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course, Numeracy score, a dummy variable for above-median Cognitive Reflection Test (i.e. CRT7) score. In Panel A we report the marginal fixed effects at the median level of the demographic controls. In Panel B we calculate the marginal fixed effects separately for above-median and below-median CRT7 score. Columns (1) and (2) report the regressions for Study 1. Column (1) includes all observations and Column (2) includes only subjects who have reported a negative number at least once during the study. Columns (3)-(5) report the results for Study 2. Column (3) includes all observations, column (4) requires that the sign of the numeric forecasts is consistent with the sign of the additionally elicited directional forecasts and column (5) additionally requires that the subject has reported a negative number at least once during the study. Standard errors are clustered at the subject level in all regressions.

Panel A	(1) $D^{\mathbb{E}_t(R_{t+1})<0}$	(2) $D^{\mathbb{E}_t(R_{t+1})<0}$	(3) $D^{\mathbb{E}_t(R_{t+1})<0}$	(4) $D^{\mathbb{E}_t(R_{t+1})<0}$	(5) $D^{\mathbb{E}_t(R_{t+1})<0}$
Study:	1	1	2	2	2
$D^{task=return}$	-0.185***	-0.118***	-0.179***	-0.158***	-0.105***
	(-7.017)	(-4.906)	(-6.877)	(-6.368)	(-5.141)
$D^{stimulus=return}$	0.089***	0.022	0.062**	0.053**	0.037*
	(3.031)	(0.812)	(2.285)	(2.111)	(1.769)
Demographics	Yes	Yes	Yes	Yes	Yes
N	1773	1603	3358	3222	3068

Panel B	(1)	(2)	(3)	(4)	(5)
			below median CRT7 score		
$D^{task=return}$	-0.289***	-0.199***	-0.222***	-0.194***	-0.109***
	(-8.46)	(-6.38)	(-5.34)	(-4.92)	(-3.58)
$D^{stimulus=return}$	0.105***	0.0216	0.104***	0.0889**	0.0505*
	(3.10)	(0.73)	(2.65)	(2.42)	(1.72)
			above median CRT7 score		
$D^{task=return}$	-0.0596	-0.0265	-0.139***	-0.125***	-0.105***
	(-1.44)	(-0.74)	(-3.98)	(-3.76)	(-3.62)
$D^{stimulus=return}$	0.0764*	0.0316	0.0335	0.0314	0.0235
	(1.71)	(0.80)	(0.96)	(0.95)	(0.81)
Demographics	Yes	Yes	Yes	Yes	Yes
N	1773	1603	3358	3222	3068

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 5: Comparison of Explanations

This table compares parameters of the hypothetical distribution of expected returns resulting from suggested response rules. The mean/standard deviation/skewness are denoted by $\mathbb{E}_t(R_{t+1})/\sigma(R_{t+1})/\upsilon(R_{t+1})$, respectively. *Negative numbers reluctance* displays the hypothetical return distribution if subjects base their responses on the chart median return, but they replace a negative response with one of the intuitive numbers - 1, 2 or 3 - with 30% probability. The assumption about the magnitude of the negative numbers reluctance (30%) is motivated by the average marginal fixed effect of the task return treatment on the tendency to respond with a negative number (see table 4). *Naïve compounding* displays the hypothetical distribution of expected returns assuming that subjects respond based on the following naïve proxy of the chart average return: $E_t(R_{t+1}) = (P_t - P_{t-12})/(12P_{t-12})$. The distributions are generated applying the suggested response rule on the data from all the simulated charts used in Study 1 and Study 2. The distribution of the chart median returns across all simulated charts is displayed by *Chart median return (benchmark)*.

Assumption	$\mathbb{E}_t(R_{t+1})$	$\sigma(R_{t+1})$	$\upsilon(R_{t+1})$
Negative numbers reluctance	0.006	0.023	-0.314
Naïve compounding	0.001	0.022	0.645
Chart median return (benchmark)	0.000	0.024	0.054

Table 6: Perceived Sample Average Returns

This table reports the results of fixed effect panel regressions with the subjective perception of the average past monthly return in the chart (elicited in study 2) as dependent variable. $\hat{R}^{naive}_{t,t-12}$ displays the hypothetical return expectations assuming that subjects respond based on following the naïve proxy of the chart average return: $E_t(R_{t+1}) = (P_t - P_{t-12})/(12P_{t-12})$. $\bar{R}^{arithm}_{t,t-12}$ and $\bar{R}^{geom}_{t,t-12}$ are the actual arithmetic and geometric average past returns in the sample covered by the respective chart. Standard errors are clustered at the subject level.

	(1) $\tilde{R}_{t,t-12}$	(2) $\tilde{R}_{t,t-12}$	(3) $\tilde{R}_{t,t-12}$	(4) $\tilde{R}_{t,t-12}$	(5) $\tilde{R}_{t,t-12}$
$\bar{R}^{geom}_{t,t-12}$	4.024***			-1.337	
	(7.761)			(-0.724)	
$\bar{R}^{arithm}_{t,t-12}$		4.037***			-1.162
		(7.762)			(-0.652)
$\hat{R}^{naive}_{t,t-12}$			4.013***	5.305***	5.131***
			(7.790)	(2.712)	(2.729)
Subject-FE	Yes	Yes	Yes	Yes	Yes
$R^2_{overall}$	0.279	0.279	0.289	0.290	0.290
N	843	843	843	843	843

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 7: Impact of Past Returns on Expectations

The dependent variables are the logarithmized return expectations from Studies 1 and 2 respectively. The independent variables are the logarithmized average monthly returns as calculated over four non-overlapping periods. Estimated are fixed effect (FE) models with standard errors clustered at the subject level. The individual columns report the results of separate regressions for each between-subject treatment in each lab experiment. Panel A displays the results for the two *stimulus price* treatments and Panel B shows the results for the two *stimulus return* treatments.

Panel A: Stimulus Price

Task Price	(1) $\mathbb{E}_t(r_{t+1})$	(2) $\mathbb{E}_t(r_{t+1})$	Task Return	(3) $\mathbb{E}_t(r_{t+1})$	(4) $\mathbb{E}_t(r_{t+1})$
Study:	1	2	Study:	1	2
$\bar{r}_{t-1,t-3}$	0.321***	0.260***	$\bar{r}_{t-1,t-3}$	0.394***	0.420***
	(2.984)	(3.872)		(2.759)	(5.950)
$\bar{r}_{t-4,t-6}$	-0.005	0.038	$\bar{r}_{t-4,t-6}$	-0.066	0.107
	(-0.054)	(0.685)		(-1.120)	(1.510)
$\bar{r}_{t-7,t-9}$	0.058	0.016	$\bar{r}_{t-7,t-9}$	0.137**	0.170***
	(0.989)	(0.369)		(2.317)	(2.781)
$\bar{r}_{t-10,t-12}$	0.092	0.031	$\bar{r}_{t-10,t-12}$	0.078	0.094**
	(1.072)	(0.672)		(0.786)	(2.040)
Subject-FE	Yes	Yes	Subject-FE	Yes	Yes
$R^2_{overall}$	0.051	0.047	$R^2_{overall}$	0.054	0.050
N	460	860	N	428	838

Panel B: Stimulus Return

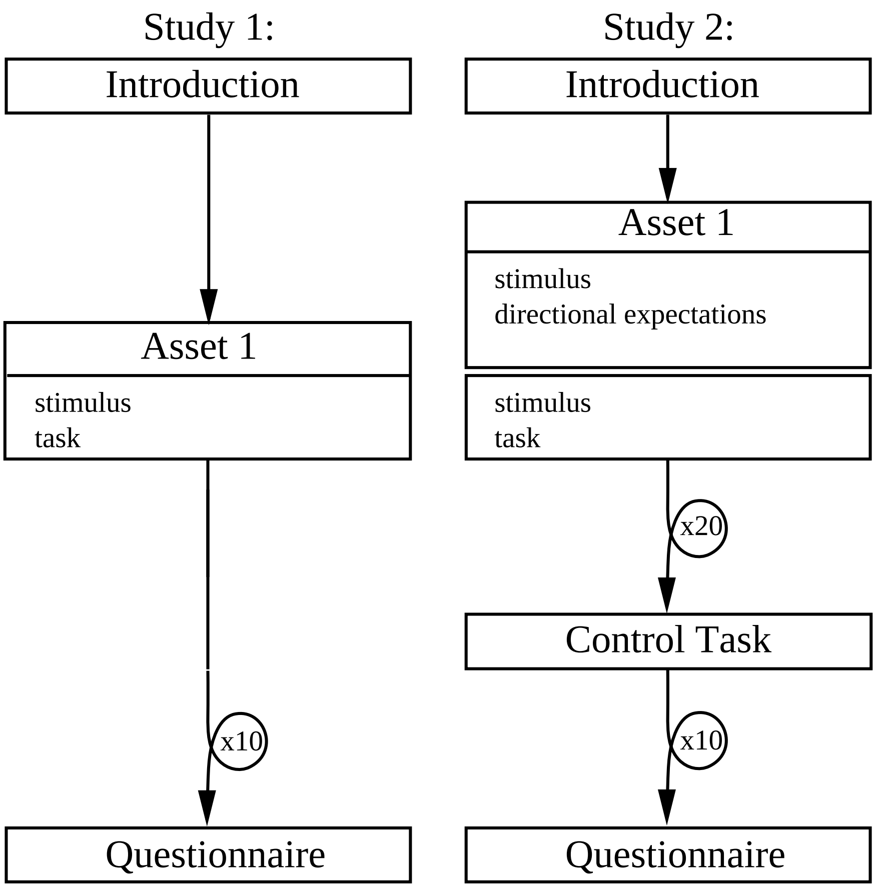
Task Price	(5) $\mathbb{E}_t(r_{t+1})$	(6) $\mathbb{E}_t(r_{t+1})$	Task Return	(7) $\mathbb{E}_t(r_{t+1})$	(8) $\mathbb{E}_t(r_{t+1})$
Study:	1	2	Study:	1	2
$\bar{r}_{t-1,t-3}$	0.466***	0.081	$\bar{r}_{t-1,t-3}$	0.295***	0.353***
	(3.531)	(0.650)		(4.454)	(6.203)
$\bar{r}_{t-4,t-6}$	0.362***	0.282***	$\bar{r}_{t-4,t-6}$	0.190***	0.199***
	(2.796)	(2.861)		(3.506)	(5.206)
$\bar{r}_{t-7,t-9}$	0.203*	0.172**	$\bar{r}_{t-7,t-9}$	0.216***	0.205***
	(1.686)	(2.620)		(3.574)	(5.099)
$\bar{r}_{t-10,t-12}$	0.129	0.177**	$\bar{r}_{t-10,t-12}$	0.085	0.107***
	(0.924)	(2.630)		(1.559)	(3.180)
Subject-FE	Yes	Yes	Subject-FE	Yes	Yes
$R^2_{overall}$	0.082	0.043	$R^2_{overall}$	0.181	0.203
N	425	796	N	460	860

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

A Appendix

A.1 Structure of Studies 1 and 2

Figure A1: Structure of Laboratory Studies 1 and 2



A.2 List of Variables

Variable	Study	Description
$\mathbb{E}_t(R_{t+n})$	1&2	Subjective expected return n periods ahead. In treatment *task price* the expected return is calculated based on the most recent available price level.
$\mathbb{E}_t(r_{t+1})$	1&2	Logarithmized subjective expected monthly return.
$\mathbb{E}_t(R^{DAX}_{t+m})$	3	Subjective expected monthly ($n = 1$) and annual ($n = 12$) return calculated from the responses in each experimental treatment. Expected return in task price is calculated based on the opening level of the DAX on the day of the submission of the response. If the response is submitted on the weekend or on a public holiday we use the last available closing level of the DAX.
$\bar{r}_{t-j,t-k}$	1&2	Logarithmized average monthly return as calculated over the period from period $t-k$ until period $t-j$.
$\tilde{R}_{t,t-12}$	2	Logarithmized self-assessed average monthly return over the past year. This variable was elicited for all subjects in the *stimulus price* treatments in Study 2. The self-assessed average monthly return was elicited within the scope of a control task subsequent to the main experimental task.
$CRT4$ & $CRT7$	1&2	The 4-item and 7-item Cognitive Reflection Test score measure the number of correct responses to the 4 new items introduced by Toplak, West, and Stanovich (2014) as an extension of the 3-item Cognitive Reflection Test by Frederick (2005) and the total number of correct responses to all items. Both measures are median-centered.
PAT	3	Priority of Analytical Tools measures the self-reported relative importance of deliberate forecasting tools compared to the importance of intuitive thinking. The subjects rate the importance of the forecasting approaches regarding short-term stock market forecasting. The importance of deliberate forecasting tools is given by the maximum importance among following deliberate approaches: technical analysis, fundamental analysis, econometric models, simulations.
Num	1&2	The Numeracy score is based on the number of correct responses on the 4-item Berlin Numeracy Test introduced by Cokely, Galesic, Schulz, and Ghazal (2012), centered around the median.

A.3 Addressing Issues regarding Internal and External Validity in Study 3

Unlike laboratory experiments, subjects in online studies cannot be obliged to participate and they can exit the survey at any point of time (i.e. attrition), which might potentially depend on aspects of the treatment. Thus potential treatment-dependent attrition might be a concern with respect to the internal validity of online experiments. For instance, participants who are reluctant to state negative numbers may refrain from responding in the *task return* but not in the *task price*. In the following, we address the issue of treatment-related attrition and provide evidence regarding the internal and external validity of our results.

Table A2: Validation of Between-Subject Randomization: Responses to Stock Market Related Tasks Outside the Scope of the Experiment

Table A2 displays the difference between the two between-subject treatments - *treatment return* and *treatment level*) - in stock market related estimates/assessments, which are outside the scope of the online experiment. Dependent variable in Column (1) is the subjective expectation regarding the level of the DAX 30 index 6 months ahead. All participants are asked to forecast the level of the DAX in points. Subjective point forecasts are converted into return forecasts by means of DAX daily open level on the day of the response. Dependent variable in Column (2) is the directional expectation on the DAX 30 index 6 months ahead elicited on a three-category Likery-type scale (i.e. Increase/Stay the same/Decrease). Displayed is the marginal effect of a *treatment return* on the probability to choose the response category *Increase*. Dependent variable in Column (3) is the subjective assessment of a current mispricing in the DAX 30 index, which is elicited on a three-category Likert-type scale (i.e. Overpriced/Fair priced/Underpriced). Displayed is the marginal effect of a *treatment return* on the probability to choose the response category *Overpriced*. Independent variable is a dummy variable which equals 1 if the subject is assigned to the *treatment return* and 0 if the subject is assigned to the *treatment level*. Daily data on DAX is downloaded from Datastream. Reported are coefficients from panel regressions (Column 1) and panel ordered-probit regressions with cluster-robust standard errors.

	(1) $E_t(R^{DAX}_{t+6})$ 6m	(2) Direction within 6m D(*Increase*)	(3) Current Mispricing D(*Overpriced*)
$D^{task=return}$	0.001	0.055	0.028
	(0.124)	(0.770)	(0.787)
N	1468	1465	1438

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

In order to address the hypothesis of treatment-related attrition, we use subjects' stock market (i.e. DAX) related expectations and assessments, which were elicited prior to the main part of our study and are identical for both treatment groups. Within the scope of the regular part of the ZEW Financial Market Survey questionnaire there are three questions preceding the main experimental task, which relate to the German stock market: (i) subjective expectations regarding the level of the DAX 6 months ahead; (ii) directional forecasts regarding the DAX elicited on a three-category Likert-type scale with the response categories "*increase*", "*stay same*" and "*decrease*"; (iii) subjective assessment of a current mispricing of the DAX elicited on a three-category Likert-type scale with the response categories "*overpriced*", "*fair priced*" and "*underpriced*". The exact wording of these questions is included in Appendix B.3, Panel B. We compare the responses of the two treatment groups to the three stock market related questions outside the scope of the main experimental task. We use linear regressions for the numerical dependent variable and ordered-probit regressions for the two categorical dependent variables. Table A2 shows that there is no statistically significant difference at the 10% level between the responses in the two treatments to any of the three stock market related variables. Therefore, we conclude that the main treatment effects discussed in Section 3 cannot be explained by systematic differences between the two treatment groups driven by treatment-related attrition.

Table A3: Self-Selection in Online Experiment: Stock Market Expertise and Mental Capacity

Dependent (binary) variable is the subject's choice to respond to the experimental task conditional on having participated in the ZEW survey in the respective wave. Included are all participants from the ZEW survey panel who faced the choice to participate in the experiment task - participants who opened the survey link in the respective wave (instead of responding to the survey via email or regular mail) and who participated in the survey in German language. Independent variable in Column (1) is a dummy variable indicating whether the participant conducts stock market forecasts outside the scope of the ZEW survey as a part of his professional occupation (self-reported). Independent variable in Column (2) is a dummy variable for the response being submitted on a Friday. Displayed are regression coefficients from panel regressions with cluster-robust standard errors.

	(1) $D(Participation)$	(2) $D(Participation)$
Not a Stock Market Forecaster (Dummy)	-0.137***	
	(-2.799)	
Friday (Dummy)		0.044**
		(1.978)
N	1507	2182

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Giving participants the choice to refrain from responding to particular questions has the purpose of reducing noise from participants who do not have sufficient expertise in a particular field. Since stock market forecasting is not the primary aim of the ZEW Financial Market Survey it is likely the case that some participants in the survey do not have any stock market expertise as measured by their usual professional activity. It is important for the implications of our results for real-world forecasting to examine the correlates of response behavior. Table A3 displays the results of linear panel regressions. The dependent variable is a dummy variable indicating whether the subject has responded to the experimental task conditional on responding to any other question in a given wave. Subjects who report that they also conduct stock market forecasts outside the scope of the ZEW survey (57% of the subjects), have a 13% higher probability of responding to the experimental task. The coefficient is significant at the 1% significance level. Furthermore, the lower response rate to the experimental tasks is likely to be determined by busyness during the week. The tendency to respond to the experimental tasks instead of only filling out the regular part of the ZEW survey is significantly higher on a Friday (at the 5% significance level), when participants are presumably less busy with regular activity and have more mental capacity for the experimental task.

The experiment subjects appear rather representative of the German financial industry in light of the methods they prefer when carrying out stock market forecasting. In September 2012 we collected background information on the methods used by experiment participants when conducting short-term DAX forecasts - the main focus of our experiment. Technical analysis is by far the most intensively used forecasting tool - 65% of the participants indicate that it is of great importance for their short-term forecasts. This result is in line with recent evidence by Menkhoff (2010) on the wide usage of technical analysis of fund managers in Germany, especially for an investment horizon of several weeks. Further factors which play a role for the short-term DAX forecasts are fundamental analysis and intuition with 31% and 22% of the participants, respectively, ranking them as highly important. In contrast, a respective majority of 57% and 64% of the participants consider econometric models and simulations of minor importance for their short-term DAX forecasts. This evidence is in line with a study by Meyler and Rubene (2009) who show that professional forecasters from the ECB Survey of Professionals Forecasters admit that they use their own judgment more often than econometric or fundamental analysis. The consistency with those studies indicates a representativeness of the ZEW panel of finance professionals.

A.4 Further Results

Table A4: Robustness Checks: Effect of the Task and the Stimulus on Expectations

This table presents robustness checks for the main treatment effects. The dependent variables are the subjective return expectations from Studies 1, 2 and 3. In Columns (1)-(3), we use the quarterly return expectations elicited subsequent to the main forecasting task in studies 1 and 2. In Columns (4) and (5) we use a subsample of the subject pool in Study 3 in order to test the external validity of our evidence on the effect of task return outside the scope of the ZEW survey. The subsample includes only subjects who are highly familiar with the experimental task - subjects who perform DAX forecasts as part of their occupational activity and whose usual forecast format is similar to the forecast format they are randomly assigned to (see Appendix B.3, Panel C). The main independent variables are a dummy variable for the task return treatment ($D^{task=return}$) and a dummy variable for the stimulus return treatment ($D^{stimulus=return}$). A task treatment is included in all three studies whereas a stimulus treatment is included only in Studies 1 and 2 (laboratory). Columns (1)-(3) display the results of random-effects panel regressions. Since the treatments are randomly assigned and thus orthogonal to personal characteristics, random effects are justified. We control for demographic characteristics, given as follows: age, gender, experience and interest in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course, Numeracy score, Cognitive Reflection Test score (see Appendix B). Column (3) displays a robustness check: In Study 2 we elicit directional forecasts prior to the main forecasting task and we compare the sign of the directional quarterly forecast with the sign of the numerical quarterly forecast. We exclude observations for which the directional forecast is inconsistent with the numerical forecast. Columns (4) and (5) display the results for the above-mentioned subsample from Study 3 for monthly and annual expected DAX returns respectively. Included are treatment-wave fixed effects to account for potentially different impact of time-series variables (such as the past DAX performance) on the expectations in the task price treatment and in the task return treatment. Standard errors are clustered at the subject level in all regressions.

	(1) $\mathbb{E}_t(R_{t+3})$	(2) $\mathbb{E}_t(R_{t+3})$	(3) $\mathbb{E}_t(R_{t+3})$	(4) $\mathbb{E}_t(R^{DAX}_{t+1})$	(5) $\mathbb{E}_t(R^{DAX}_{t+12})$
Study:	1	2	2	3	3
$D^{task=return}$	0.030***	0.019***	0.021***	0.018***	0.028*
	(4.192)	(2.963)	(2.983)	(3.357)	(1.670)
$D^{stimulus=return}$	-0.028***	-0.015**	-0.018***		
	(-3.578)	(-2.394)	(-2.577)		
Demographics	Yes	Yes	Yes	No	No
Treatment-Wave FE	-	-	-	Yes	Yes
$R^2_{overall}$	0.053	0.030	0.028	0.147	0.086
N	1773	3354	3206	579	371

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A5: Impact of Positive and Negative Past Returns on the Treatment Effects

The dependent variables are the subjective return expectations from Studies 1 and 2 respectively. The main independent variables are a dummy variable for task return ($D^{task=return}$), a dummy variable for stimulus return ($D^{stimulus=return}$), a dummy variable for positive past annual return ($D^{R^{geom}_{t,t-12}>0}$) and interactions between the latter and the treatment dummies. We control for following demographic characteristics: age, gender, experience and interest in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course, Cognitive Reflection Test score. Standard errors are clustered at the subject level.

	(1) $\mathbb{E}_t(R_{t+1})$	(2) $\mathbb{E}_t(R_{t+1})$
Study:	1	2
$D^{task=return}$	0.021***	0.015***
	(4.129)	(2.829)
$D^{stimulus=return}$	-0.022***	-0.013**
	(-3.720)	(-2.484)
$D^{R^{geom}_{t,t-12}>0} \times D^{task=return}$	0.005	0.010*
	(0.783)	(1.908)
$D^{R^{geom}_{t,t-12}>0} \times D^{stimulus=return}$	0.010	0.007
	(1.513)	(1.457)
$D^{R^{geom}_{t,t-12}>0}$	0.017***	0.013***
	(3.342)	(3.269)
Demographics	Yes	Yes
$R^2_{overall}$	0.093	0.076
N	1773	3354

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A6: Robustness Check: Impact of Intuitive Thinking on the Effects of the Task and the Stimulus

The dependent variables are the subjective monthly return expectations from Studies 1 and 2, respectively. The main independent variables are a dummy variable for the task return treatment ($D^{task=return}$), a dummy variable for the stimulus return treatment ($D^{stimulus=return}$), a measure of the subjects' Cognitive Reflection and numeracy and an interaction between the latter and the treatment dummies. We measure Cognitive Reflection by means of the Cognitive Reflection Test score comprising of 4 items (i.e. CRT4) and 7 items (i.e. CRT7), respectively (see Toplak, West, and Stanovich, 2014), centered around the median score. We measure Numeracy by means of the Berlin Numeracy Test score (see Cokely, Galesic, Schulz, and Ghazal, 2012), centered around the median score. We control for the following demographic characteristics: age, gender, experience and interest in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course. Standard errors are clustered at the subject level.

	(1) $\mathbb{E}_t(R_{t+1})$	(2) $\mathbb{E}_t(R_{t+1})$	(3) $\mathbb{E}_t(R_{t+1})$	(4) $\mathbb{E}_t(R_{t+1})$
Study:	1	1	2	2
$D^{task=return}$	0.023***	0.031***	0.015***	0.030***
	(5.877)	(6.859)	(4.222)	(3.107)
$D^{stimulus=return}$	-0.017***	-0.024***	-0.005	-0.019**
	(-3.951)	(-4.698)	(-1.567)	(-2.268)
$CRT7 \times D^{task=return}$	-0.005**		-0.011**	
	(-2.169)		(-2.235)	
$CRT7 \times D^{stimulus=return}$	0.005*		0.012**	
	(1.857)		(2.525)	
$CRT4 \times D^{task=return}$		-0.010***		-0.015**
		(-3.327)		(-2.027)
$CRT4 \times D^{stimulus=return}$		0.009***		0.013**
		(2.659)		(2.119)
$Num \times D^{task=return}$	-0.002	-0.002	0.003	-0.001
	(-0.627)	(-0.605)	(0.514)	(-0.166)
$Num \times D^{stimulus=return}$	0.001	0.001	-0.006	-0.001
	(0.276)	(0.345)	(-1.096)	(-0.134)
Demographics	Yes	Yes	Yes	Yes
$R^2_{overall}$	0.063	0.068	0.092	0.081
N	1773	1773	3354	3354

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table A7: Descriptive Statistics

This table displays descriptive statistics of the distributions of subjective expectations across treatments and studies. In each table, "P" indicates price level treatments and "R" denotes return treatments. We provide the following parameters of the sample distributions: sample average (i.e. μ), between-subject and within-subject standard deviation (σ_b and σ_w respectively), skewness (υ). Significance levels indicate the results of Wald tests (with respect to μ) and D'Agostino-Pearson tests (with respect to υ).

Study 1			Task: P	Task: R
Stimulus	P	μ	-0.001	0.028***
		$\sigma_w(\sigma_b)$	0.016 (0.054)	0.038 (0.062)
		υ	-0.392***	2.177***
	R	μ	-0.012***	0.005**
		$\sigma_b(\sigma_w)$	0.033 (0.072)	0.012 (0.041)
		υ	-1.651***	-0.681***

Study 2			Task: P	Task: R
Stimulus	P	μ	-0.002	0.024***
		$\sigma_w(\sigma_b)$	0.012 (0.046)	0.066 (0.054)
		υ	-0.079	3.307***
	R	μ	-0.006**	0.007***
		$\sigma_b(\sigma_w)$	0.018 (0.065)	0.017 (0.042)
		υ	-1.901***	0.254***

* p<0.1, ** p<0.05, *** p<0.01

Table A8: Impact of the Price Level on the Main Treatment Effects

This table reports the results of random effect panel regressions with the subjective return expectations from Study 2 as dependent variable. The main independent variables are a dummy variable which equals 1 for task return ($D^{task=return}$) and a dummy variable which equals 1 for stimulus return ($D^{stimulus=return}$), a dummy variable for the price level sequence starting from 1000 monetary units as opposed to 100 monetary units ($D^{P_0=1000}$) and an interaction of the latter with the task dummy and the stimulus dummy. We test random-effect models and control for demographic characteristics, given as follows: age, gender, experience and interest in stock markets and financial markets in general, attendance of a statistics course and a Behavioral Finance course, Numeracy score, Cognitive Reflection Test score. Standard errors are clustered at the subject level in all regressions.

	(1) $\mathbb{E}_t(R_{t+1})$	(2) $\mathbb{E}_t(R_{t+1})$
$D^{task=return}$	0.022***	0.024***
	(3.636)	(3.546)
$D^{stimulus=return}$	-0.009	-0.013*
	(-1.473)	(-1.886)
$D^{P_0=1000}$	0.005	0.008
	(0.907)	(1.248)
$D^{P_0=1000} \times D^{task=return}$	-0.004	-0.003
	(-0.417)	(-0.226)
$D^{P_0=1000} \times D^{stimulus=return}$	-0.001	-0.001
	(-0.117)	(-0.066)
Demographics	Yes	Yes
$R^2_{overall}$	0.048	0.039
N	3354	3222

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

B Online Appendix

All experimental studies and supplementary questionnaires described in this paper were conducted in German. In the following we display the translated versions. The original versions in German language are available upon request.

B.1 Instructions - Studies 1 and 2

In the following text parts included only in the instructions in Study 1 are highlighted in *red*, whereas the text parts included only in Study 2 are highlighted in *blue*.

Treatment PP: Stimulus Price and Task Price

`Screen 1:`

In the following you will see overall 10 charts. Each chart displays the simulated historical price development of one financial market instrument. After each chart we ask you to forecast the *future price level* of the respective financial market instrument. The actual realization of the future price level should lie with equal probability above or below your forecast. For each financial market instrument we will ask you to forecast the future price level in one month and in three months. For this part of the experiment you will receive a fixed remuneration of 4 euros.

The future price level will be simulated subsequently. The characteristics of the financial market instrument will remain unchanged and the future price level will not be affected by your forecast.

Your remuneration depends on the accuracy of your forecast. At the end of the experiment we will randomly draw one instrument and one forecast horizon, which will then be relevant for your remuneration. Your remuneration depends on the absolute deviation of your forecast from the realized future price level. If your forecast is sufficiently close to the realized future price level, your remuneration will amount to 21 euros. Otherwise your remuneration will amount to 6 euros.

Click "Next" to proceed to the chart of the first financial market instrument.

`Screen 2:`

The chart displays the historical price development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Compared to the current price level you forecast the future price level of the financial market instrument one month from now to be rather... [higher/equal/lower]

Compared to the current price level you forecast the future price level of the financial market instrument three months from now to be rather... [higher/equal/lower]

`Screen 3:`

The chart displays the historical price development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Please provide a forecast of the future price level (in monetary units) of this financial market instrument one month from now. The actual realization of the future price level should lie with equal probability above or below your forecast.

How certain are you about your response? [1 "not at all" - 6 "very certain"]

Please provide a forecast of the future price level (in monetary units) of this financial market instrument in three months from now. The actual realization of the future price level should lie with equal probability above or below your forecast.

How certain are you about your response? [1 "not at all" - 6 "very certain"]

...

`Screen 42`:

You have completed the first part of the experiment. In the next minutes we will draw the instrument and the forecast horizon which is relevant for your remuneration. While you are waiting we would ask you to look once again at some of the charts your just saw and respond to some additional questions.

`Screen 43`:

The chart displays the historical price development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU). What is the average monthly return of the financial market instrument over the past year? A return is defined as the percentage change of the price of the instrument. [Please mark a negative return with a minus.]

Treatment PR: Stimulus Price and Task Return

`Screen 1`:

In the following you will see overall 10 charts. Each chart displays the simulated historical price development of one financial market instrument. After each chart we ask you to forecast the *future return* of the respective financial market instrument. The actual realization of the future return should lie with equal probability above or below your forecast. Return is defined as the percentage change of the price of the financial market instrument. For each financial market instrument we will ask you to forecast the future return over the next month and over the next three months. For this part of the experiment you will receive a fixed remuneration of 4 euros.

The future return will be simulated subsequently. The characteristics of the financial market instrument will remain unchanged and the future return will not be affected by your forecast.

Your remuneration depends on the accuracy of your forecast. At the end of the experiment we will randomly draw one instrument and one forecast horizon, which will then be relevant for your remuneration. Your remuneration depends on the absolute deviation of your forecast from the realized future return. If your forecast is sufficiently close to the realized future return, your remuneration will amount to 21 euros. Otherwise your remuneration will amount to 6 euros.

Click ”Next” to proceed to the chart of the first financial market instrument.

`Screen 2`:

The chart displays the historical price development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Your forecast of the future return of the financial market instrument over the next month is rather... [positive/zero/negative]

Your forecast of the future return of the financial market instrument over the next three months is rather... [positive/zero/negative]

`Screen 3`:

The chart displays the historical price development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Please provide a forecast of the future return (in percent) of this financial market instrument over the next month. The actual realization of the future return should lie with equal probability above or below your forecast.

How certain are you about your response? [1 ”not at all” - 6 ”very certain”]

Please provide a forecast of the future return (in percent) of this financial market instrument over the next three months. The actual realization of the future return should lie with equal probability above or below your forecast.

How certain are you about your response? [1 ”not at all” - 6 ”very certain”]

*Return is defined as the percentage change of the price of the financial market instrument.

. . .

`Screen 42`:

You have completed the first part of the experiment. In the next minutes we will draw the instrument and the forecast horizon which is relevant for your remuneration. While you are waiting we would ask you to look once again at some of the charts your just saw and respond to some additional questions.

`Screen 43`:

The chart displays the historical price development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU). What is the average monthly return of the financial market instrument over the past year? A return is defined as the percentage change of the price of the instrument. [Please mark a negative return with a minus.]

Treatment RP: Stimulus Return and Task Price

`Screen 1`:

In the following you will see overall 10 charts. Each chart displays the simulated historical return development of one financial market instrument. Return is defined as the percentage change of the price of the financial market instrument. After each chart we ask you to forecast the *future price level* of the respective financial market instrument. The actual realization of the future price level should lie with equal probability above or below your forecast. For each financial market instrument we will ask you to forecast the future price level in one month and in three months. For this part of the experiment you will receive a fixed remuneration of 4 euros.

The future price level will be simulated subsequently. The characteristics of the financial market instrument will remain unchanged and the future price level will not be affected by your forecast.

Your remuneration depends on the accuracy of your forecast. At the end of the experiment we will randomly draw one instrument and one forecast horizon, which will then be relevant for your remuneration. Your remuneration depends on the absolute deviation of your forecast from the realized future price level. If your forecast is sufficiently close to the realized future price level, your remuneration will amount to 21 euros. Otherwise your remuneration will amount to 6 euros.

Click "Next" to proceed to the chart of the first financial market instrument.

`Screen 2`:

The chart displays the historical return development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Compared to the current price level you forecast the future price level of the financial market instrument one month from now to be rather... [higher/equal/lower]

Compared to the current price level you forecast the future price level of the financial market instrument three months from now to be rather... [higher/equal/lower]

`Screen 3`:

The chart displays the historical return development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Please provide a forecast of the future price level (in monetary units) of this financial market instrument one month from now. The actual realization of the future price level should lie with equal probability above or below your forecast.

How certain are you about your response? [1 ”not at all” - 6 ”very certain”]

Please provide a forecast of the future price level (in monetary units) of this financial market instrument in three months from now. The actual realization of the future price level should lie with equal probability above or below your forecast.

How certain are you about your response? [1 ”not at all” - 6 ”very certain”]

*Return is defined as the percentage change of the price of the financial market instrument.

. . .

Screen 42:

You have completed the first part of the experiment. In the next minutes we will draw the instrument and the forecast horizon which is relevant for your remuneration. While you are waiting we would ask you to look once again at some of the charts your just saw and respond to some additional questions.

Screen 43:

The chart displays the historical return development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU). Imagine you had invested 100 monetary units in this instrument one year ago. What would be the value of your investment today?

Treatment RR: Stimulus Return and Task Return

Screen 1:

In the following you will see overall 10 charts. Each chart displays the simulated historical return development of one financial market instrument. Return is defined as the percentage change of the price of the financial market instrument. After each chart we ask you to forecast the *future return* of the respective financial market instrument. The actual realization of the future return should lie with equal probability above or below your forecast. For each financial market instrument we will ask you to forecast the future return over the next month and over the next three months. For this part of the experiment you will receive a fixed remuneration of 4 euros.

The future return will be simulated subsequently. The characteristics of the financial market instrument will remain unchanged and the future return will not be affected by your forecast.

Your remuneration depends on the accuracy of your forecast. At the end of the experiment we will randomly draw one instrument and one forecast horizon, which will then be relevant for your remuneration. Your remuneration depends on the absolute deviation of your forecast from the realized future return. If your forecast is sufficiently close to the realized future return, your remuneration will amount to 21 euros. Otherwise your remuneration will amount to 6 euros.

Click "Next" to proceed to the chart of the first financial market instrument.

`Screen 2:`

The chart displays the historical return development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Your forecast of the future return of the financial market instrument over the next month is rather... [positive/zero/negative]

Your forecast of the future return of the financial market instrument over the next three months is rather... [positive/zero/negative]

`Screen 3:`

The chart displays the historical return development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU).

Please provide a forecast of the future return (in percent) of this financial market instrument over the next month. The actual realization of the future return should lie with equal probability above or below your forecast.

How certain are you about your response? [1 ”not at all” - 6 ”very certain”]

Please provide a forecast of the future return (in percent) of this financial market instrument over the next three months. The actual realization of the future return should lie with equal probability above or below your forecast.

How certain are you about your response? [1 ”not at all” - 6 ”very certain”]

*Return is defined as the percentage change of the price of the financial market instrument.

...

`Screen 42`:

You have completed the first part of the experiment. In the next minutes we will draw the instrument and the forecast horizon which is relevant for your remuneration. While you are waiting we would ask you to look once again at some of the charts your just saw and respond to some additional questions.

`Screen 43`:

The chart displays the historical return development of one financial market instrument over one year as well as the current price level of the financial market instrument at the time of forecasting (month 0) in monetary units (MU). Imagine you had invested 100 monetary units in this instrument one year ago. What would be the value of your investment today?

B.2 Questionnaire: Studies 1 and 2

A. Demographics
D-1. Your age: ....
D-2. Your gender: ...
D-3. Do you have any interest in stock markets?
D-4. Do you have any interest in financial markets in general?
D-5. Do you have any experience with investing in stocks or equity funds?
D-6. Do you have any experience with investing in other financial market instruments?
D-7. Have you attended a statistics course at the University?
D-8. Have you attended a Behavioral Finance course at the University?

B. Cognitive Reflection Test
**CRT7 score comprises of questions 1-7; CRT 4 score comprises of questions 4-7*
CRT-1. A bat and a ball cost 1.10 euros in total. The bat costs a dollar more than the ball. How much does the ball cost? *Correct answer: 5 cents; Intuitive answer: 10 cents*
CRT-2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

Continued on next page

Continued from previous page

Correct answer: 5 minutes; Intuitive answer: 100 minutes

CRT-3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Correct answer: 47 days; Intuitive answer: 24 days

CRT-4. 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?

Correct answer: 4 days; Intuitive answer: 9 days

CRT-5. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?

Correct answer: 29 students; Intuitive answer: 30 students

CRT-6. A man buys a pig for 60 euros, sells it for 70 euros, buys it back for 80 euros, and sells it finally for 90 euros. How much has he made?

Correct answer: 20 euros; Intuitive answer: 10 euros

CRT-7. Simon decided to invest 8.000 euros in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: (a) broken even in the stock market, (b) is ahead of where he began, (c) has lost money.

Correct answer: c; Intuitive answer: b

Sources: Frederick (2005), Toplak, West, and Stanovich (2014)

Continued on next page

Continued from previous page

C. Numeracy: Berlin Numeracy Test

N-1. Out of 1.000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.
Correct answer: 25%

N-2. Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?
Correct answer: 30 out of 50 throws

N-3. Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?
Correct answer: 20 out of 70 throws

N-4. In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?
Correct answer: 50%

Source: Cokely, Galesic, Schulz, and Ghazal (2012)

D. Other Variables

Study 1: Faith-in-Intuition Score (Epstein, Pacini, Denes-Raj, and Heier, 1996)
Study 2: field of study

B.3 Instructions and Questionnaire: Study 3

A. Main Task	
MONTHLY RETURN EXPECTATIONS	
- Treatment:	Between-subject randomized
- Timing:	Sep 2012 - Jun 2015; quarterly repeated
- Question wording:	Task Return
	Within 1 month I expect a DAX return (monthly percentage change) of ...% percent.
	With 90% probability the DAX return will then lie between ... percent and ... percent.
	Task Price
	I expect the DAX in 1 month at ... points.
	With 90% probability the DAX will then lie between ... and ... points.
ANNUAL RETURN EXPECTATIONS	
- Treatment:	Between-subject randomized
- Timing:	Sep 2013 - Jun 2015; quarterly repeated
- Question wording:	Task Return
	Within 1 year I expect a DAX return of ...% percent.
	With 90% probability the DAX return will then lie between ... percent and ... percent.
	Task Price
	I expect the DAX in 1 year at ... points.
	With 90% probability the DAX will then lie between ... and ... points.
B. Stock Market Related Tasks Outside the Scope of the Main Task	
MID-TERM RETURN EXPECTATIONS (POINT FORECAST)	
- Note:	All subjects receive the same questions

Continued on next page

continued from previous page

- Timing:	Included in all survey waves in the period Sep 2012 - Jun 2015
- Question wording:	*I expect the DAX in 6 months at ... points.*
Perception of Mispricing of the DAX (Likert-type Scale)	
- Timing:	Included in all survey waves since 2011
- Question wording:	*Taking into account the fundamental data regarding the DAX companies, I assess the DAX as currently ...*
- Response categories:	*overpriced/fair priced/underpriced*
Mid-Term Return Expectations (Likert-type Scale)	
- Timing:	Included in all survey waves since 1995
- Question wording:	*DAX (Germany) will ... in the mid-term (6 months).*
- Response categories:	*increase/stay the same/decrease*
C. Questionnaire (Stable Personal Traits)	
Importance of Diverse Forecasting Tools	
- Note:	All subjects receive the same questions
- Timing:	Supplementary Question in September 2012
- Question wording:	*How important are following factors for your short-term (1 month ahead) DAX forecasts?* technical analysis (TA), fundamental analysis (FA), experience econometric models (EM), simulations (SIM), intuition (INT)
- Response categories:	low/medium/high
- Measure:	$PAT = max_j(importance^j) - importance^{INT}$, where $j = TA, FA, EM, SIM$
Proximity between Forecaster's Usual Forecast Format and Format of Randomly Assigned Main Task	
- Timing:	Supplementary Question in June 2013
- Question wording:	*What is the usual type of your regular forecasts outside the scope of the ZEW Financial Market Survey?*
- Assumption:	*[P] indicates that we categorize the format as "close to task price";*

Continued on next page

continued from previous page

	[R] indicates that we categorize the format as "close to task return"
- Response categories:	[P] level forecasts (e.g. I expect the DAX at ... points.)
	[R] return forecasts (e.g. I expect a DAX return of ... %.)
	[P] range forecasts (e.g. I expect the DAX between ... and ... points.)
	[R] directional forecasts (e.g. The DAX will increase/stay same/decrease.)
	[R] probability estimate
	other
	I do not conduct any explicit stock market forecasts.