Overconfidence and the Political and Financial Behavior of a Representative Sample *

Ciril Bosch-Rosa[1], Bernhard Kassner[2], and Steffen Ahrens[3]

[1]Chair of Macroeconomics, Technische Universität Berlin
[2]Seminar for Economic Policy, Ludwig-Maximilians-Universität München
[3]Chair of Macroeconomics, Freie Universität Berlin

July 14, 2023

Abstract

We study the relationship between overconfidence and the political and financial behavior of a nationally representative sample. To do so, we introduce a new method to directly elicit individual overconfidence which is simple to understand, fast to answer, and captures respondents' excess confidence in their own judgment. Our results show that, in line with theoretical predictions, an excessive degree of confidence in one's judgment is correlated with lower portfolio diversification, larger stock-price forecasting errors, and more extreme political views. We also find that overconfidence is correlated with voting absenteeism. Finally, using a companion survey, we show that overconfidence is consistent across different domains. All of these findings suggest that overconfidence is a personality trait that permeates most aspects of people's lives.

Keywords Overconfidence, SOEP, Survey

JEL Classification C83 · D91 · G41

*Corresponding author: cirilbosch@gmail.com. We would like to thank Agustin Casas, Antoni-Italo de Moragas, Benjamin Enke, Tilman Fries, Thomas Gräber, Andreas Haufler, Frank Heinemann, Don Moore, Guillem Riambau, Uwe Sunde, the participants at the CESifo Area Conference in Behavioral Economics as well as the participants of seminars at CUNEF, EMLyon, and FU Berlin for their insightful comments. We acknowledge the financial support from the Deutsche Forschungsgemeinschaft (DFG) through CRC TRR 190 "Rationality and Competition." Ethical permission was provided by the Scientific Advisory Board of DIW Berlin and the Ethics Committee of the Faculty of Economics at Ludwig-Maximilians-Universität München and informed consent was obtained from all participants included in the analyses. The authors are not presented in alphabetical order, as the first two authors contributed more to the paper.

What would I eliminate if I had a magic wand? Overconfidence.

—Daniel Kahneman, *The Guardian*, 18 July 2015

1 Introduction

Overconfidence is a pervasive and potent bias in human judgment (Mannes and Moore, 2013; Kahneman, 2011). It leads to wars (Johnson, 2004), to excessive entry into markets (Camerer and Lovallo, 1999), or to 80% of the population thinking that they are above-average drivers (Svenson, 1981). However, overconfidence is a general term that encompasses three different phenomena: overestimation, overplacement, and overprecision (Moore and Healy, 2008; Moore and Schatz, 2017). Overestimation has to do with absolute values—you think that you are better than you really are. Overplacement has to do with relative values—you think that your performance is better than that of others, when it is not. In this paper, we focus on overprecision. Overprecision has to do with the degree of certainty with which a person judges their own knowledge—you think that your knowledge is more accurate than it really is. In other words, overprecision relates to the second moment of the distribution, such that a person may hold accurate beliefs on average but underestimate the variance of the possible outcomes (Malmendier and Taylor, 2015).

Overprecision has important consequences. From an economic point of view, overprecision may lead consumers to buy less insurance than they should (Grubb, 2015) or to large distortions in corporate investment decisions (Ben-David et al., 2013; Moore et al., 2015). In finance, overprecision is responsible for an under-diversification of portfolios (Goetzmann and Kumar, 2008), for excessive trading (Barber and Odean, 2001), and for systematic forecasting errors (Deaves et al., 2019). In a political context, overprecision leads to ideological extremism, strong partisan identification (Ortoleva and Snowberg, 2015a,b; Stone, 2019), and increased susceptibility to “fake news” (Thaler, 2023). However, the existing evidence either uses indirect measures of overprecision (such as gender or the tendency to make extreme predictions), estimates derived from econometric models, or elicited confidence intervals – a method that has been shown to be problematic (Teigen and Jørgensen, 2005; Bazerman and Moore, 2013; Moore et al., 2015).

In this paper, we study whether overprecision is a stable personality trait and how it relates to the political and financial behavior of a nationally representative sample of the German population. To do so, we introduce a new method to elicit overprecision, which we call the "Subjective Error Method." This method consists of a two-step process. In the first step, participants answer a numerical question (e.g., In what year was Saddam Hussein captured by the US army? or How many meters tall is the Eiffel Tower?). In the second step, they are asked to estimate the "distance" (in the units of the question) between their response to the first question and the correct answer. In other words, in the second step, respondents are asked to report their expected *absolute error* to the first question (henceforth, subjective error). By comparing the realized error to their subjective error, we can determine respondents' overprecision in a simple and direct way. Ultimately, by aggregating over multiple questions, we can derive a reliable measure of overprecision for each respondent.

The richness of our data allows us to correlate the overprecision of respondents with their socio-demographic characteristics, as well as their financial and political behavior. The results show that overprecision is positively correlated with narcissism, negatively correlated with age, years of education, gross income, and financial literacy, but does not differ across genders. We also find that overprecision aligns well with several theoretical conjectures regarding political and financial behavior. Specifically, our measure is positively correlated with larger forecasting errors in respondents' stock price predictions and with lower portfolio diversification, as suggested by Odean (1998) and Barber and Odean (2000). Regarding subjects' political views and behavior, our measure of overprecision predicts a tendency to hold extreme political ideologies, as suggested by Ortoleva and Snowberg (2015b). Yet, in contrast to Ortoleva and Snowberg (2015b), our measure of overprecision is associated with voting absenteeism rather than an increased likelihood to vote. We surmise that the difference could be attributed to the different electoral systems in Germany and the United States. The alignment of the results with the theoretical predictions suggests that the Subjective Error Method reliably captures behavior that is consistent with overprecision.

We also test whether overprecision is a stable personality trait—that is, a cognitive bias that is robust across different domains. To do so, we employ the Subjective Error Method in a companion online survey administered to a representative sample of the

German population and covering five different domains (contemporary history, general knowledge, economics, four-week ahead stock price predictions, and a neutral counting task).[1] Using different statistical approaches to test the correlation of overprecision across the different domains, we show that overprecision is robust *within* individuals across all domains. This suggests that overprecision is a persistent personality trait in an instant of time.

Our paper contributes to the existing literature on overprecision in four dimensions: first, we directly elicit overprecision by introducing the Subjective Error Method, a novel technique that is easy to understand, quick to implement, and captures respondents' excess confidence in their own judgment. Second, applying our new measure of overprecision, we can confirm distinct theoretical predictions regarding the financial and political behavior of respondents. Specifically, we show that a higher degree of overprecision results in lower portfolio diversification, larger stock price forecasting errors, and ideological extremism. Third, using an online survey, we show that overprecision is a personality trait that is robust within individuals across different domains. Finally, while most of the existing literature on overprecision uses university students (e.g., Alpert and Raiffa, 1982), or special pools of subjects (e.g, Glaser and Weber (2007) use finance professionals and McKenzie et al. (2008) IT professionals), we test theoretical predictions across different domains on a representative sample of the German population. Taken together, our paper delivers further empirical evidence that overprecision impacts different aspects of individuals' lives.

The remainder of the paper proceeds as follows: Section 2 discusses the notion of overprecision and introduces our measure of overprecision, the Subjective Error Method. In Section 3, we present the SOEP-IS data set, correlate overprecision with various sociodemographic measures, and use our measure of overprecision to predict the behavior of respondents on various domains in finance and politics such as predicting asset market returns, portfolio diversification, or voting behavior. In Section 4, we present the com-

[1] These five domains are distinctive in terms of the potential to know the true answer. For the contemporary history, general knowledge, and economics domain, a correct answer exists on the day of the survey. This correct answer can be known by the individual, depending on the degree of prior knowledge of this topic. For the neutral counting task, although a correct answer exists, it can only be estimated. For the stock price predictions, the correct answer is unknown on the day of the survey. Since the latter two are independent of previous knowledge, we can test whether overprecision is robust even across domains where the influence of prior knowledge is limited.

panion online survey and test the robustness of overprecision within individuals across domains. The last section concludes.

2 Overprecision and the Subjective Error Method

2.1 Overprecision Measurements

Overprecision (also known as miscalibration) is a type of overconfidence that results from an excess of confidence in one's own information (Moore et al., 2015). It relates to the second moment of the belief distribution and thereby directly affects how information is processed. For this reason, it is widely used in finance and political science to model overconfident agents. For example, Odean (1998) finds that overconfident traders trade excessively and hold underdiversified portfolios because they believe that their private signals are more precise than they really are. Scheinkman and Xiong (2003) combine a constraint on short sales and overprecise traders to explain the formation of asset market bubbles.[2] In the political science literature, Ortoleva and Snowberg (2015b) show that more overprecise people tend to vote more, hold more extreme political views, and show stronger partisan identification. Consistent with this, Stone (2019) suggests that overprecision increases partisanship through excessively strong inferences from (biased) information sources. More recently, the literature has begun to study the role that overprecision plays in disseminating fake news (Pennycook et al., 2021; Thaler, 2023).

Yet, precisely because overprecision deals with the second moment of the belief distribution, it is difficult to measure (Moore et al., 2015). The most common way to measure overprecision, introduced by Alpert and Raiffa (1982), is to elicit the respondents' 90% confidence intervals (CI) for a series of numerical questions (e.g., How long is the Nile River?). Using this paradigm, a perfectly calibrated respondent would not capture the correct answer within the CI in one out of every ten questions. However, the literature has shown that this method is problematic since respondents are not familiar with CIs and do not fully grasp what they are being asked (Moore et al., 2015). This is corroborated by the finding that the elicited intervals resulting from asking 90% CIs are practically identical to those resulting from asking for 50% CIs (Teigen and Jørgensen, 2005) and

[2] For a longer discussion on the different models of overprecision used in the finance literature see Daniel and Hirshleifer (2015).

that oftentimes the purported 90% CIs only contain the correct answer between 30% to 60% of the time (e.g., Russo and Schoemaker, 1992; Bazerman and Moore, 2013; Moore et al., 2015).

While there are some alternatives to CIs when measuring overprecision, these tend to be either time-consuming or limited in the information they provide. For example, the two-alternative forced-choice (2AFC) method developed by Griffin and Brenner (2004) asks respondents to choose between two possible answers to a question and then indicate how confident they are that their answer is correct. By comparing the number of correct answers to the stated confidence, one can measure whether, on average, respondents are overconfident. However, this method has several drawbacks as it cannot distinguish between overprecision and overestimation of one's own knowledge (Moore et al., 2015) and cannot capture continuous distributions (see Moore et al. (2015) and Griffin and Brenner (2004) for a further discussion of the 2AFC method and its statistical limitations). Another approach to measuring overprecision is the Subjective Probability Interval Estimates (SPIES) method by Haran et al. (2010). The SPIES method elicits complete probability distributions from respondents. Although the SPIES method appears to measure overprecision more accurately than CIs (Moore et al., 2015), it is time-consuming, and it requires respondents to understand the concept of probability distributions. Additionally, because distributions can only be elicited by partitioning the support into discrete bins, researchers need to make a series of *ad hoc* decisions to implement and define the desired 90% boundaries of the distribution. Finally, Ortoleva and Snowberg (2015b) use an estimation method where they regress a self-reported measure of confidence in the accuracy of their answers on a six-point scale on a polynomial of the realized error. The drawback of this approach is that the individual measure of overprecision is dependent on the relationship between confidence and accuracy for the entire population of respondents, which might incorrectly classify subjects as over- or underprecise.[3]

To address these caveats, we introduce the Subjective Error Method, a novel method, which is easy to understand, quick and simple to implement, and does not depend on subjects' knowledge of statistical concepts or the properties of the entire sample.

[3] See Appendix C for a more detailed discussion with examples.

2.2 The Subjective Error Method

The Subjective Error Method consists of asking two consecutive questions to respondents. The first question (a) can be on any topic but needs to have a numerical answer.[4] The second question (b) asks respondents how far away they expect their answer to question (a) to be from the true answer. In other words, the second question asks respondents to report their absolute subjective error.[5] An example would be:

(a) *How long (in kilometers) is the Nile River?*

(b) *How far away (in kilometers) do you think your answer to (a) is from the true answer?*

By comparing the *subjective error* of respondents stated in (b) to the absolute *realized error* from question (a), we get a measure of how over- or underprecise a respondent is about their knowledge.

To fix ideas, assume that a respondent's true error is normally distributed, with mean 0 and variance σ^2 as shown by the solid curve in Figure 1. A perfectly calibrated individual would, on average, correctly assess the distribution of the true error when answering questions using the Subjective Error Method. However, the perceived distribution for most respondents might not necessarily coincide with the true distribution. If the respondent is overprecise, then their perceived variance of the error $\hat{\sigma}^2$ is smaller than the true variance, i.e., the precision $\rho = 1/\hat{\sigma}^2$ is larger (dashed curve in Figure 1). In this case, the subjective error would, on average, consistently deviate from the true error, resulting in a systematic deviation across all questions.[6]

[4] Some examples of numerical questions are the result of multiplying 385 by 67, the length of the Nile River, or the year of Lady Diana's death. Some examples of questions that do not work are the name of the oldest son of Lady Diana, the color of the Batmobile, or the gender of the current prime minister of the United Kingdom.

[5] Respondents could have different concepts of the subjective error in mind when answering this question, which creates additional noise. In a companion online experiment in Appendix G, we show that respondents majoritarily are consistent in the way they interpret the subjective error question. The mode of the distribution is at a 50% confidence level with symmetric deviation in both directions. We are willing to accept this additional noise in the trade-off of having a simpler measure of overprecision.

[6] Note that the difference between the true error and the subjective error that would realize with the same cumulative probability is directly proportional to the difference in the precision of the underlying distributions.

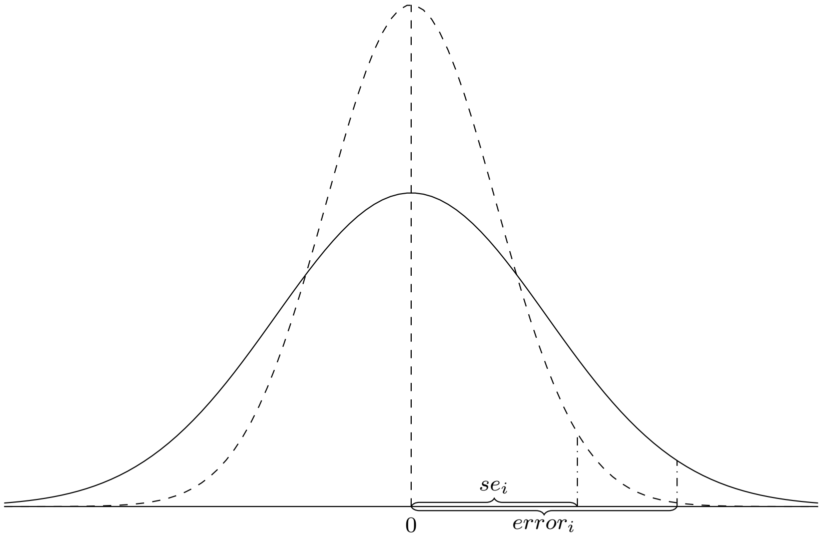


Figure 1: The figure shows two hypothetical normal distributions of the (subjective) error. The solid curve shows the true distribution of the error with a standard deviation of 2 (precision of .25). The dashed curve shows the perceived distribution by an overprecise respondent with a standard deviation of 1.25 (precision of .64). The dash-dotted vertical lines indicate the subjective error se_i and the absolute true error $error_i$ resulting with the same cumulative probability.

Denote the answer of respondent i to question j as $a_{i,j}$, their subjective error for question j as $se_{i,j}$, and the true answer to the question as ta_j, then our measure of overprecision for respondent i for question j is:

$$error_{i,j} = |a_{i,j} - ta_j|, \tag{1}$$

$$op_{i,j} = error_{i,j} - se_{i,j}, \tag{2}$$

where equation (1) measures the absolute realized error ($error_{i,j}$) of respondent i to question j. Note that this equation calculates the *absolute error*; that means that we do not care about the direction of the error but rather about its size. In equation (2), we calculate the difference between the subjective error ($se_{i,j}$) and the realized error ($error_{i,j}$) of respondents i to question j. In this case, we do care about the direction of the error, as a respondent who underestimates their subjective error (i.e., $error_{i,j} > se_{i,j}$) is considered *overprecise*, while a respondent who overestimates their subjective error (i.e., $error_{i,j} < se_{i,j}$) is considered *underprecise*. Finally, those respondents who correctly guess their subjective error (i.e., $error_{i,j} = se_{i,j}$) are considered perfectly calibrated for that question.

As it is clear from the setup, the biggest advantage of eliciting overprecision using the Subejctive Error Method over other alternatives is its simplicity. As opposed to most other alternative methods, the SEM does not require respondents to have any statistical knowledge to answer the questions. This adds to making the setup easy to explain, and fast to answer. Moreover, the dual question approach allows us to simultaneously capture a respondent's knowledge on a topic and their confidence in that knowledge. Therefore, this measure of overprecision should not be affected by knowledge, as any reduction of errors in the first question is likely offset by a symmetric reduction of the subjective error.[7]

In a related paper, Enke and Graeber (2021) study the "subjective uncertainty about the optimal action" that experimental subjects have when confronted with choices across different economic domains. To measure such uncertainty, they take an approach very similar to the Subjective Error Method—they allow subjects to provide a symmetric interval of "uncertainty" around the answers provided to each question. Their results show that such symmetric bounds are robust within and across subjects and have strong predictive power across the different domains they study. Overall, while the setup proposed by Enke and Graeber (2021) is not designed to measure overprecision, it lends support to the Subjective Error Method as a robust tool to elicit the degree of uncertainty of respondents for a given answer.

3 The Subjective Error Method and the Behavior of Individuals

In this section, we apply the Subjective Error Method to study how overprecision correlates with socio-demographic characteristics (Section 3.3) and the political and financial behavior (Section 3.4) of a nationally representative sample of the German population.

3.1 Data

We use data from the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). The Innovation Sample is a companion panel of the larger SOEP-Core and is designed to host and test novel survey items (see, Richter and Schupp, 2015). We use the 2018 wave of the SOEP-IS, which had 4,860 individual respondents distributed across 3,232

[7] In principle, knowledge should not affect the measure of overprecision. This is because any reduction of errors in the first question is likely offset by a symmetric reduction of the subjective error, resulting in a "neutral" effect of knowledge. Such neutral effect is corroborated by the literature (e.g., Önkal et al., 2003; McKenzie et al., 2008) and by the results of our companion survey reported in Appendix F.1. In Appendix F.1 we confirm that the SEM is not affected by expertise.

different households. As the SOEP-Core, the SOEP-IS is a high-quality survey with all interviews conducted face-to-face by a professional interviewer.

To construct our measure of overprecision, we use data from seven different questions. In each question, we ask respondents to answer two things, (a) the year of a specific historical event that occurred not more than 100 years ago, and (b) the distance (in years) between their answer to (a) and the correct answer to (a).[8] In other words, we ask respondents to answer a contemporary history question and then we ask them to report the absolute error they expect to make, i.e., their subjective error (see Section 2.2).

We ask seven different questions about events taking place between 1938 and 2003. The questions are designed to vary in difficulty and to cover different decades. The content of the questions ranges from the year in which the Volkswagen Beetle was introduced (1938) to the year in which Saddam Hussein was captured by the US Army (2003) (see Table B.1 and Table B.2 in the appendix for all questions and their correct answers in English and German, respectively).[9] These questions were asked to those respondents who joined the panel in 2016 ($N = 902$). We supplement the data with additional information on personal characteristics from the survey years 2016–2018. We drop 55 respondents who did not answer any of the overprecision questions, since this is our main variable of interest, and 42 respondents with incomplete information. In total, we end up with a sample of 805 respondents across 584 different households.[10]

3.2 Individual Overprecision

In Figure 2, we plot the density of answer $a_{i,j}$ for each question j. The red vertical line marks the correct answer. The dispersion of the densities shows that some questions were easier for respondents than others. In Figure 3, we plot the realized error ($error_{i,j}$) in

[8] The questions are formulated in German. For the example in which we ask about the year of the death of Lady Diana we ask: (a) *In welchem Jahr starb Lady Diana, die erste Frau von Prinz Charles?* and then (b) *Was schätzen Sie, wie viele Jahre Ihre Antwort von der richtigen Antwort entfernt ist?*.

[9] Subjects could answer using any integer between 1900 and 2019 for question (a) and between 0 and 119 for question (b).

[10] To test whether our estimation sample is still representative of the German population, we compare the unweighted means of personal characteristics in our sample with the weighted means according to the sampling weights in the larger SOEP-Core, which is representative of the German population. The results in Table B.4 in the appendix show that our subsample is still broadly representative of the larger SOEP-Core, with only some significant but small and nonmeaningful differences. When applying the sampling weights to our estimation sample, the differences disappear.

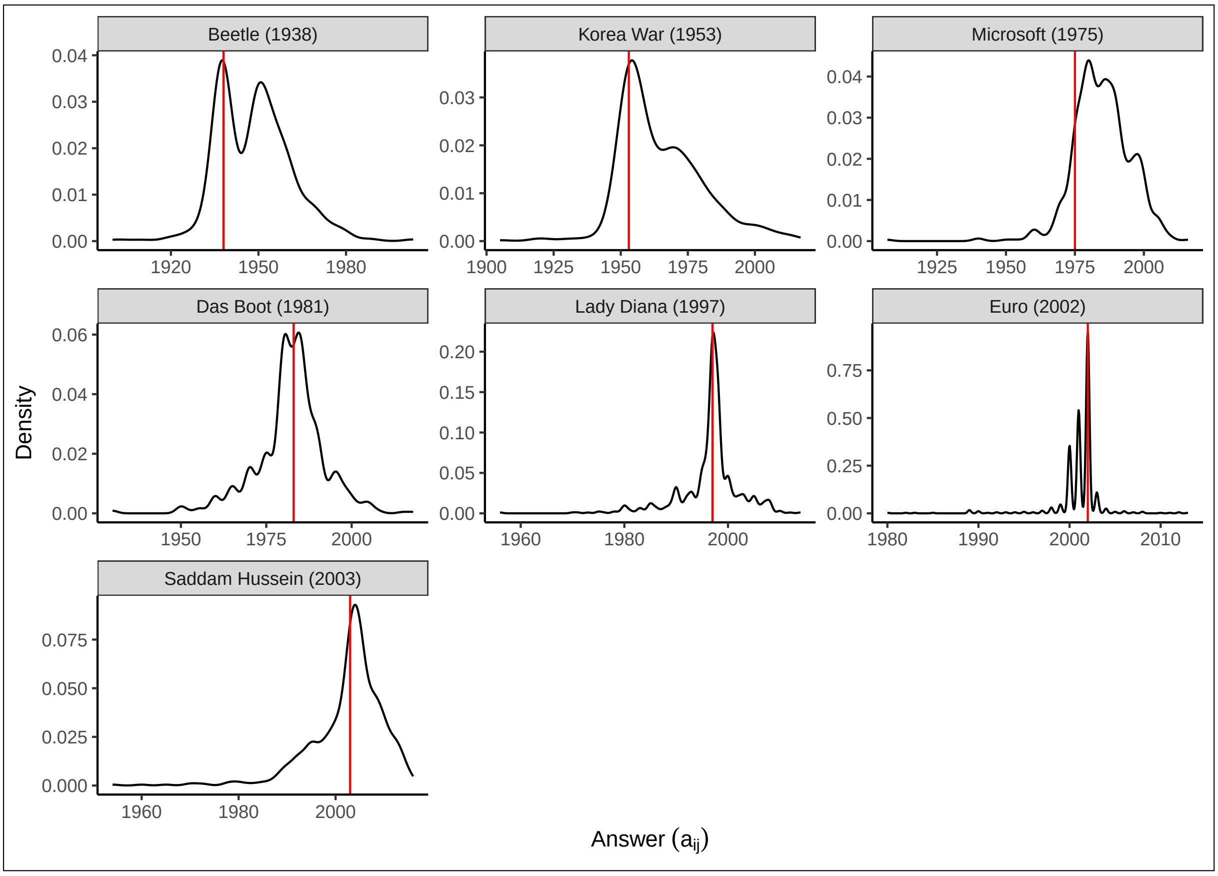


Figure 2: Density of the answers ($a_{i,j}$) for each question. The red vertical line marks the correct answer. Note that the vertical axis is different for each question.

the vertical axis and subjective error ($se_{i,j}$) in the horizontal axis for each of the seven questions. Additionally, we plot a 45-degree red line, so that any dot above is a respondent who is overprecise ($error_{i,j} > se_{i,j}$) in their answer to the question, and any point below corresponds to a respondent who is underprecise ($error_{i,j} < se_{i,j}$). It is clear from the figure that respondents are, on average, overprecise in their answers across all questions independent of their difficulty.

We construct overprecision for each question ($op_{i,j}$) following the outline in Section 2.2. To measure consistency in the overprecision measure across the seven questions for each subject, we use congeneric reliability, which is commonly referred to as coefficient omega (e.g., Cho, 2016). Congeneric reliability indicates the share of variation (variance and covariance) among a set of variables that can be explained by an unobserved factor.[11]

[11] Consider a model in which each observed outcome i of item j can be expressed as $T_{i,j} = \mu_j + \lambda_j F_i + e_{i,j}$, where $T_{i,j}$ is the i^{th} outcome of item j, μ_j is a constant term, $e_{i,j}$ is the individual score error, and λ_j is the factor loading on the latent common factor F. To construct congeneric reliability, we estimate

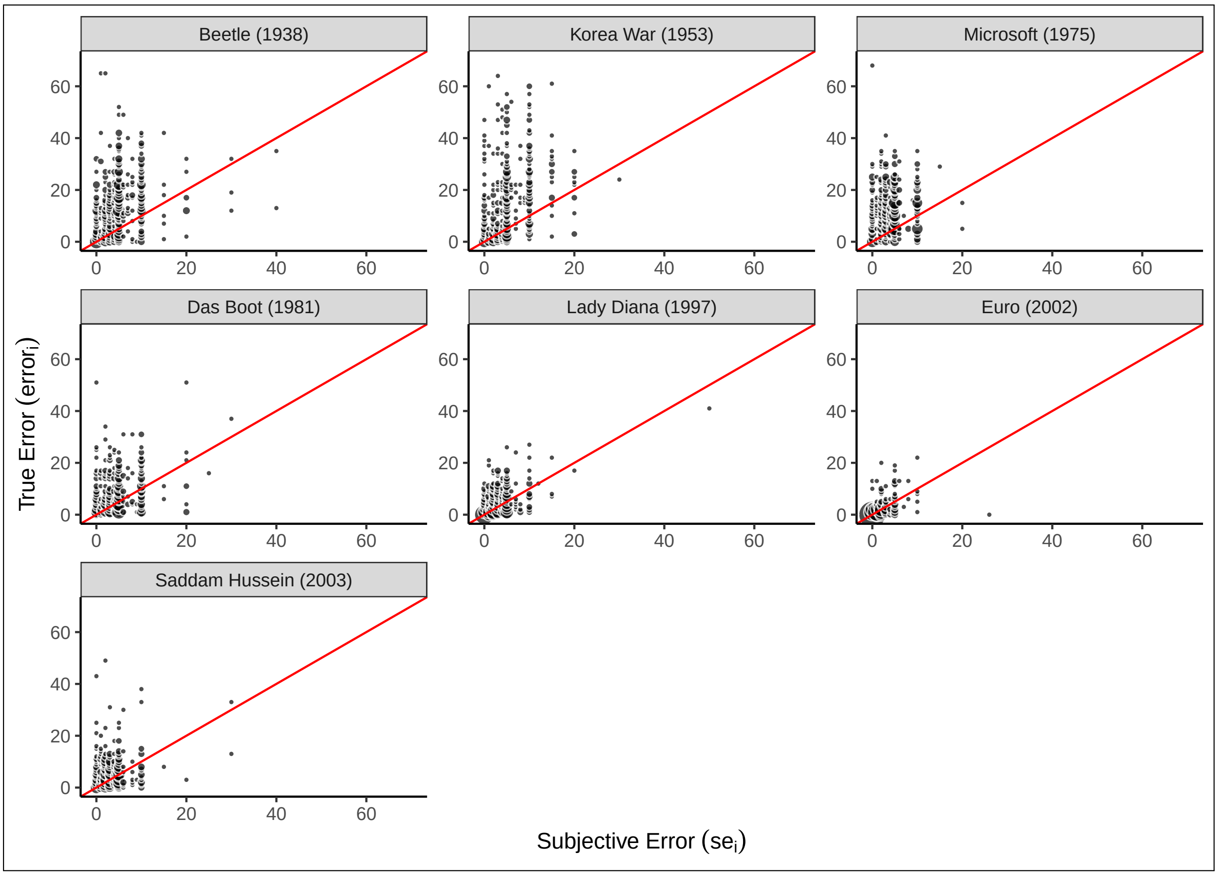


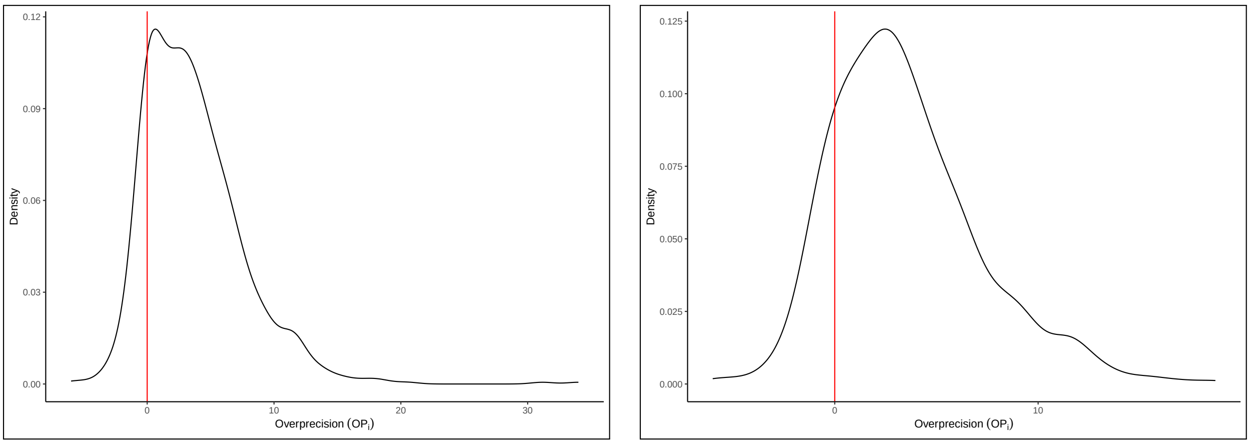
Figure 3: Relation between the realized error ($error_{i,j}$) in the vertical axis and the subjective error ($se_{i,j}$) in the horizontal axis. Any dot above (below) the 45-degree red line is an overprecise (underprecise) answer by the respondent.

The results show that 49% of the variation among the seven items can be explained by a common factor, which we interpret as overprecision.

To create a unique measure of overprecision for each respondent i (op_i), we take the average overprecision across all seven questions j.[12] We plot the density of op_i in Figure 4a.

the factor loadings, $\hat{\lambda}_j$, for the overprecision measure of each question with respect to one common factor. We interpret this common factor as overprecision. Congeneric reliability is calculated according to the formula $\frac{(\sum \hat{\lambda}_j)^2}{(\sum \hat{\lambda}_j)^2 + \sum \hat{\sigma}^2_{e_j}}$, where $\hat{\sigma}^2_{e_j}$ is the estimated variance of the error. This is a generalized version of Cronbach's alpha (Cronbach, 1951) and measures the share of variation among the set of items j that can be explained by the latent factor. While congeneric reliability allows for different factor loadings of the latent common factor, Cronbach's alpha assumes that the latent factor equally loads on all items and is, thus, a lower bound for reliability. For the case of τ-equivalence, i.e., $\lambda_j = \lambda_k \ \forall k$, all factor loadings are equal and both measures coincide.

[12] Given that a principal component analysis of the seven items yields a strong first factor, an alternative would be to construct the composite measure op_i using the principal component as in Ortoleva and Snowberg (2015b). The composite overprecision measure resulting from using the first component is very similar to using the simple average ($\rho^{Pearson} = .88; \rho^{Spearman} = .84; N = 805$). Since the results only marginally change when using the principal component, we do not report the results, which are available



(a) Density of op_i for all respondents (N=805)

(b) Density of op_i for subset (N=410)

Figure 4: Density of Overprecision (op_i). In the left panel we plot the density of op_i, which is the average overprecision for each respondent i across all questions j. In the right panel we plot the density of op_i only for those respondents who answered all questions in the survey.

Consistent with Figure 3, Figure 4a shows that the great majority of respondents (82%) are overprecise. On the other hand, and in contrast with most of the literature using CIs to measure overprecision, we find a relatively large number of respondents that are underprecise (approximately 11%).

Moreover, 7% of the respondents seem to be perfectly calibrated (vertical red line in Figure 4a) in the aggregate measure. Of these 52 respondents, 83% are perfectly calibrated across all the questions they answer. However, note that in the SOEP-IS respondents can decide not to answer a question; 51% of the respondents answered all questions, with 5% answering only one (see Figure A.1 in the appendix for a detailed breakdown). Of those perfectly calibrated respondents, 40% answered only one question, and only 12% answered all seven. This means that what we see in Figure 4a is an “upper bound” of perfectly calibrated respondents. As can be seen in Figure 4b, once we plot the density function for the subset of respondents that answered *all questions*, we find that respondents are substantially less calibrated, with the mode of op_i shifting to the right and leaving only 1% of the respondents perfectly calibrated; at the same time, there is an increase in the proportion of underprecise respondents (15%).

For ease of interpretation in the subsequent analysis, we standardize the aggregate

upon request. Additionally, to alleviate concerns about different scales, we also construct a standardized measure of overprecision by standardizing each measure of overprecision ($op_{i,j}$) before aggregating in Appendix C.

score (op_i) to be mean zero and standard deviation one (sop_i).[13]

3.3 Socio-Demographic Determinants of Overprecision

In Table 1, we regress sop_i on a series of socio-demographic variables using five different OLS models. In all models, we control for age, gender, and years of education. In Column (2) we add the number of overprecision questions answered. In Column (3), we add the monthly gross individual income (*gross income*) measured in thousands of euros as well as dummies for labor force status (e.g., employed, unemployed, maternity leave, etc.) and a dummy for those respondents who were living in East Germany in 1989.[14] In Column (4), we add further personal characteristics, which are financial literacy, risk aversion, impulsivity, patience, and narcissism. Finally, in Column (5), we add federal state (*Bundesland*) and month-of-interview fixed effects.[15]

The results show a negative correlation between overprecision and age, education, or income. For example, an increase in the gross income of 2,000 euros is associated with a reduction in overprecision by almost one-tenth of a standard deviation, and every 2 years of education are associated with a reduction in overprecision by about one-tenth of a standard deviation. Furthermore, overprecision is negatively correlated with our measure of financial literacy and positively correlated with our measure of narcissism. It is also important to note that the number of questions answered by respondents, which we include in Column (2), is not random, with overprecision increasing as subjects answer more questions (see Figures A.2 and A.1 in the appendix for a graphical overview of these results). In all subsequent analyses, we use the above-mentioned variables as controls.

The results in Table 1 largely align with the existing literature. For example, the positive correlation with narcissism aligns with previous results (e.g., Campbell et al.,

[13] In Appendix C, we show the robustness of our measure of overprecision by comparing it to five alternative approaches. These are i) a *standardized* measure, which standardizes each question before aggregating, ii) a *centered* measure, which centers the errors and subjective errors around their mean, allowing us to disentangle the second moment of the distribution (overprecision) from its first moment, iii) a *relative* approach, which takes into account the relative distance between the subjective error and the realized error, iv) an *age-robust* measure, which is constructed using only those questions concerning events that occurred after the respondent was born, and v) a *residual* approach following the regression methodology of Ortoleva and Snowberg (2015b).

[14] Since *gross income* is only available for employed individuals, we code missing variables as 0 and include a dummy that is 1 for missing observations.

[15] Note that we only report coefficients that are statistically significant in Table 1.

Dependent Variable: *sop*	(1)	(2)	(3)	(4)	(5)
age	-0.008***	-0.008***	-0.007**	-0.005	-0.005
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
female	0.074	0.121	0.094	0.134*	0.111
	(0.070)	(0.074)	(0.074)	(0.074)	(0.076)
years education	-0.053***	-0.062***	-0.050***	-0.043***	-0.037***
	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)
answered		0.064***	0.067***	0.086***	0.085***
		(0.023)	(0.022)	(0.023)	(0.023)
gross income			-0.049**	-0.037*	-0.039*
			(0.021)	(0.020)	(0.021)
fin. literacy				-0.481***	-0.398**
				(0.170)	(0.174)
narcissism				0.107**	0.100**
				(0.042)	(0.043)
N	805	805	805	805	805
adj. R^2	0.036	0.046	0.061	0.081	0.098
Constant Term	Yes	Yes	Yes	Yes	Yes
Employment & GDR 1989	No	No	Yes	Yes	Yes
Personal characteristics	No	No	No	Yes	Yes
Fixed Effects	No	No	No	No	Yes

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Determinants of overprecision. In Columns (1)-(5) we run an OLS with *sop* as the dependent variable. In all Columns, we include age, gender, and education. In Column (2) we add the number of answered questions. In Column (3) we include gross income and dummies for labor force status (employed, unemployed, retired, maternity leave, nonworking), and whether the respondent was a citizen of the GDR before 1989. In Column (4) we include further personal characteristics, which are financial literacy, risk aversion, impulsivity, patience, and narcissism. In Column (5) we also include fixed effects for the federal state (*Bundesland*) where the respondents live and the time at which they responded to the questionnaire.

2004; Hamurcu and Hamurcu, 2021). Similarly, the negative correlation between education and overprecision, if we consider education a proxy for cognitive ability (e.g., Duttle, 2016). The effect of gender on overprecision is far from universal in the literature. While Ortoleva and Snowberg (2015b) find that females are significantly less overprecise than males, López-Pérez et al. (2021), Deaves et al. (2009), and Wohleber and Matthews (2016) find no effect. This is supported by Bandiera et al. (2022) who find that both men and women are overconfident with no significant difference between gender by aggregating experimental findings over the last twenty years, which aligns with our results. On the other hand, there is evidence that overprecision increases with age (Prims and Moore, 2017; Ortoleva and Snowberg, 2015b), while we find a slight negative correlation. However, our result is likely to be driven by the question domain as we also find a positive correlation between age and overprecision when controlling for age in the construction of the overprecision measure in Section C. Finally, the literature on overprecision and financial literacy is scarce. Kramer (2016) reports that confidence is negatively correlated with financial advice seeking while objective measures of financial literacy are not. This suggests a negative relationship between overprecision and financial literacy, which aligns with our results.

3.4 Overprecision and the Financial and Political Behavior of Respondents

In this section, we examine how overprecision correlates with respondent behavior in the political and financial domains by testing several predictions from the theoretical literature. In Section 3.4.1, we describe the methodology, and in Section 3.4.2, we present the results and compare them with the theoretical predictions.

3.4.1 Methodology

To test the predictions from the theoretical literature on overprecision, on which we elaborate in more detail below, we use three different procedures. First, we run a regression of each outcome (y_i) on our measure of overprecision and a vector of control variables of the form:

$$y_i = \alpha + \beta sop_i + \boldsymbol{\gamma'} \boldsymbol{X_i} + \epsilon_i, \tag{3}$$

where sop_i denotes the standardized overprecision measure for respondent i, $\boldsymbol{X_i}$ is a vector of control variables, and ϵ_i is the random error term. We include all possible control variables available in the SOEP-IS that might be correlated with the outcome we are studying or with overprecision based on the previous literature. These are age, gender, years of education (which serves as a proxy for cognitive ability), monthly gross labor income, dummy variables for the labor force status (employed, unemployed, maternity-leave, non-working, and retired), measures of impulsivity, patience, narcissism, financial literacy, and risk aversion, a dummy variable for having lived in the German Democratic Republic in 1989, the number of overprecision questions answered by each respondent, state fixed effects, and interview date (month and year) fixed effects. The latter absorbs any variation in the outcome driven by the time the survey was run, e.g., in the development of the asset prices. Additionally, we include a measure of political interest in the political analyses.[16] A test for multicollinearity shows no strong linear dependencies across the explanatory variables. We estimate (3) using OLS and present the point estimate of the standardized overprecision measure sop_i from the full regression and its unadjusted p-value respectively in Columns (1) and (2) of Table 2.[17] Since we test the behavior of respondents across several dimensions, we also report the Sidak-Holm adjusted p-value for multiple hypothesis testing in Column (3).

Second, we derive the "R^2 rank" of our standardized measure of overprecision sop_i from a forward stepwise regression, similar to Cobb-Clark et al. (2019). This is obtained by running a stepwise regression in which we sequentially keep adding variables to the model. To do so, in step 1, we regress the behavior of interest on each of the K control variables in the specification separately. Of these K regressions, we pick the control variable with the highest R^2. Hence, this variable is able to explain most of the observed variation in the data. In step 2, we regress $K-1$ times the behavior of interest on the control variable selected in the first step plus each of the $K-1$ remaining controls. This is continued until all K variables have been added to the model. The resulting R^2 rank is determined by the step at which each control variable was added to the model. Therefore, the better the "R^2 rank" of sop_i, the more the variable can explain the variation in the outcome, i.e., rank 1 delivers the highest R^2. We report the results in Column (4) of Table 2 along with the maximum number of variables to be included in the model as specified above.[18]

Finally, we employ a least absolute shrinkage and selection operator (LASSO) to test whether our overprecision measure has predictive power for the outcome variable in an out-of-sample prediction. LASSO is a machine learning application that is frequently applied to improve the predictive power of statistical models. The objective of the LASSO approach is to choose those variables with the highest predictive power from the set *of all*

[16] In Table B.5 in the appendix, we also include the Big Five personality traits (Rammstedt and John, 2007). These are only available from the 2017 SOEP-IS, and because not all respondents in our sample responded to them, we lose 55 observations. Yet, the results remain robust to the inclusion of the Big Five personality traits.

[17] Adjusting the degrees of freedom by the number of questions used to construct the measure of overprecision does not significantly affect the results.

[18] Note that each binary representation of a possible realization of the fixed effects could be included as an independent variable. This leads to 41 and 42 as the maximum number of variables to be included.

possible control variables. It does so by estimating a penalized regression by minimizing the sum of squared residuals and a penalty term for the sum of the coefficients.[19] This is implemented via cross-validation, i.e., the estimator partitions the data into different folds of training and testing data and selects the penalty term that minimizes the out-of-sample prediction error in the testing data.[20] If sop_i is included in the model, then it has predictive power for the outcome. We report the results in Column (5) of Table 2 along with the number of control variables chosen by LASSO and the resulting R^2 of the model in Column (6). In Column (7), we report the number of observations, which may vary due to missing observations in the outcome variables.[21]

3.4.2 Prediction Results

The results of our three analytical approaches are summarized in Table 2.[22] We first discuss financial behavior outcomes and then outcomes regarding political behavior. For each outcome, we start by presenting the theoretical predictions from the literature.

Financial Behavior Outcomes

The first hypothesis concerns the forecast errors of asset price predictions in the stock market. Benos (1998) and Odean (1998) argue that overprecise investors hold incorrect beliefs about the future valuation of assets because they overweight their private signals when forming expectations. Direct empirical support for the association of overprecision and forecast errors in financial markets is provided by Deaves et al. (2019), who correlate the predictions of German stock market forecasters with a measure of overprecision. Additionally, Hilary and Menzly (2006) provide evidence consistent with this association

[19] Formally $\min_\beta \frac{1}{2N}\sum_{i=1}^{N}(y_i - \alpha - \sum_j \beta_j x_{ij})^2 + \lambda \sum_j |\beta_j|$ for the linear case, where j are the coefficients which are included in the model and λ is a given tuning parameter. See Tibshirani (1996) for more details.

[20] The algorithm proceeds step-wise and estimates the model for each λ starting at the smallest λ that delivers zero non-zero coefficients and ending at a λ of 0.00005 in a grid of 100. In each step, a different number of variables could be added or removed from the model.

[21] A test of the means of personal characteristics for the estimation samples and the entire sample (N=805) shows no significant differences. The only exception is a slightly higher share of male respondents in the stock market regressions. We therefore consider the estimation samples to be representative of the entire sample (N=805).

[22] In Appendix C, we test the robustness of our results using the five alternative approaches mentioned in Section 3.3. The results principally replicate.

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) LASSO R^2	(7) N
Financial Behavior:							
DAX forecast error	0.083*	0.056	0.108	4/41	yes/14	0.10	548
1-year ahead	*0.529*	*0.308*		*7/41*	*yes/21*	*0.16*	*578*
2-year ahead	*3.078***	*0.016*		*4/41*	*yes/8*	*0.04*	*557*
portfolio diversification	-0.131***	0.000	0.002	3/41	yes/19	0.14	774
Political Behavior:							
extremeness	0.087**	0.041	0.117	6/42	yes/14	0.05	716
left-right	-0.008	0.854	0.854	18/42	no/14	0.08	716
non-voter	0.032**	0.011	0.041	3/42	yes/18	0.14	706

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: This table shows the estimation results of Section 3.4. The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 805. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)). Column (4) displays the R^2 rank in the forward stepwise regression as specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)). Variable definitions are in Table B.3 in the appendix.

for North American analysts.[23] Therefore, we expect overprecise respondents to be less accurate in their predictions. To test this prediction we use the absolute distance of the one- and two-year-ahead predictions of the German Stock Index (DAX), Germany's blue-chip stock market index, from the realized value.[24] Since single forecast errors might be prone to random noise, we aggregate both errors using the principal component, which we standardize to be mean zero and standard deviation one (*DAX forecast error*).[25] Additionally, we report the results for both forecast errors separately (*1-year ahead* and *2-year*

[23] However, note that, unlike our method, Hilary and Menzly (2006) and Deaves et al. (2019) rely on indirect proxies to construct their measure of overprecision.

[24] We use the closing price based on the day of the interview to calculate exact forecast errors for each respondent. Note that the one-year-ahead observations from the 2018 waves are almost all from the period before March 2019 and are thus unaffected by the stock market decline caused by the coronavirus crisis in March 2020.

[25] The principal component analysis shows a strong first factor with an eigenvalue of 1.65. All other factors are below the commonly used Kaiser criterion of 1.0.

ahead).[26]

The results in Table 2 show that our measure of overprecision is positively correlated with forecast errors in asset prices. An increase in overprecision of 1 standard deviation is associated with an increase in the principal component in the forecast errors by 0.083 standard deviations. The LASSO estimation results reveal that overprecision is also a good predictor of these forecast errors since it is selected as an explanatory variable for the models of stock market forecasts; it also ranks fourth in the forward stepwise regression approach.

Next, we test the theoretical prediction by Odean (1998) that overprecision is associated with underdiversified portfolios. Intuitively, overprecise investors overweigh their private information, thereby trading too frequently while concentrating on an overly limited number of favorable assets (Odean, 1998). Goetzmann and Kumar (2008) provide empirical evidence supporting this prediction for traders in the US and Merkle (2017) does so for traders in the UK. While the former relies on the asset turnover proxy, the latter elicits overprecision directly through survey questions. We test this hypothesis using a standardized measure with mean 0 and standard deviation of 1 that captures the degree to which a respondent diversifies their hypothetical portfolio among stocks, real estate, government bonds, savings, and gold.[27]

It is important to note that by this approach, we do not make any claim on the optimal degree of portfolio diversification. We rather test whether, conditional on the individual degree of risk aversion, overprecision is associated with a tendency towards a certain asset category. We argue that this approach is in line with the theoretical arguments of Odean (1998) who shows that overconfident traders overreact to their personal information and underdiversify by investing more in a certain asset.

Our results confirm the theoretical prediction that overprecision is associated with underdiversification. The point estimate in Column (1) in Table 2 shows that a 1 standard error increase in overprecision leads to a 0.131 standard deviation decrease in our

[26] We include a dummy variable that indicates asset ownership as possible control variables in the predictions to account for different information sets in a robustness test in Table B.6 in the appendix. The qualitative results remain unaffected by this change, although the sample size decreases.

[27] For a detailed description of the measure, see Table B.3 in the appendix. The measure of diversification displays an inverse-U relation with risk aversion. More risk-averse respondents skew their portfolio toward safer assets such as savings and gold whereas more risk-loving respondents skew their portfolio toward riskier assets such as stocks and real estate. This lends credibility to the diversification measure.

diversification measure. That means that the optimal portfolio of overprecise respondents is skewed towards a certain asset category. Moreover, overprecision is among the variables in the LASSO model and ranked third in the forward stepwise regression approach.

Political Views and Voting Behavior

According to Ortoleva and Snowberg (2015b), overprecision leads people to believe that their own experiences are more informative about politics than they really are. For instance, overprecise people may consult biased media outlets without fully accounting for this bias or exchange information on social media without realizing that much of the information comes from politically like-minded peers. Against this background, the authors show theoretically and empirically that overprecision leads to ideological extremeness and strengthens the identification with political parties, increasing the likelihood to vote. Yet, the literature remains inconclusive on whether these associations hold for liberals and conservatives alike. While Moore and Swift (2011) and Ortoleva and Snowberg (2015b) find that conservatives seem more susceptible to overprecision than liberals, Ortoleva and Snowberg (2015a) show that this association only holds in election years.

To test whether overprecision correlates with the political preferences of respondents, we use their self-reported ideology on a scale from 0 (extreme left) to 10 (extreme right) to construct the variable *left-right*. Using the answer to the same question, we also construct *extremeness*, which measures from 0 to 5 how far away from the political center respondents see themselves. We standardize both variables to be mean zero and standard deviation one. Finally, to study whether overprecise respondents are more likely to vote, we use a dummy that equals 1 if a respondent indicated being a nonvoter in the (ex-post) opinion poll (*Sonntagsfrage*) for the 2017 federal elections to the German *Bundestag* (*nonvoter*).[28]

In line with Ortoleva and Snowberg (2015b), the results of Table 2 suggest that overprecision is positively correlated with ideological extremeness. Overprecision is among the variables chosen by the LASSO estimation and ranks high (sixth) in the forward stepwise regression approach. Confirming Ortoleva and Snowberg (2015a), we do not find evidence that overprecision is associated more strongly with any side of the political spectrum, as

[28] Note that 2018, the year in which the survey took place, is a non-election year. According to Ortoleva and Snowberg (2015a), there should be no correlation between overprecision and conservatism.

it is not correlated with political ideology and is not among the variables chosen by the LASSO estimation. Furthermore, overprecision is ranked quite low (18/42) in the R^2 rank approach. Finally, we find that overprecision is a strong predictor of voting absenteeism, with overprecision being chosen by the LASSO estimation and ranked third in the R^2 rank approach. Hence, it seems that overprecision increases the likelihood of voting absenteeism rather than increasing the likelihood of voting: An increase in the standard deviation for overprecision of 1 results in a 3 percentage point increase in the likelihood of not voting.

The last result seems to be in contradiction with the result of Ortoleva and Snowberg (2015b). However, the voting behavior of overprecise respondents in the United States and Europe is difficult to compare. In Ortoleva and Snowberg (2015b) partisanship is measured *within* the Republican and Democratic parties. Because both of these parties have high chances of winning the elections, those more identified with such parties have stronger incentives to vote for them (e.g., Miller and Conover, 2015). By contrast, in Germany, more extreme respondents gravitate to fringe parties (e.g., Die Linke, AfD, NPD)[29] with smaller chances of winning elections, so the incentives to vote are very different than for those in the dataset used by Ortoleva and Snowberg (2015b).[30] Hence, the theoretical assumptions underlying the predictions made by Ortoleva and Snowberg (2015b) regarding voter turnout and overprecision are a good description of voting behavior in the two-party system of the United States but are not appropriate for the more disperse German system.

The results presented in this section show that the majority of respondents in the sample are overprecise. This cognitive bias is correlated with a range of personal characteristics largely consistent with previous results in the literature. Moreover, in line with the theoretical literature, we find that overprecision is positively correlated with stock market forecast errors, negatively correlated with portfolio diversification, and positively correlated with political extremism. Taken together, these results suggest that our mea-

[29] If we pool all respondents voting for radical parties (AfD, NPD, and Die Linke) and compare them to the voters of the rest of parties, a nonparametric test confirms the tendency of radical party voters to ideological extremeness (Mann-Whitney U p-value<0.001).

[30] Take as an example the explicit (self-imposed) *cordon sanitaire* that all major democratic parties have imposed around the AfD. Angela Merkel's intervention and the series of resignations that followed the 2019 Thuringian election shows how strongly this *cordon* is enforced.

sure of overprecision captures a behavior that is consistent with overprecision. In the following section, we further validate the notion that overprecision is a distinct behavioral trait.

4 Overprecision Across Domains

In Section 3, we showed that overprecision is associated with different aspects of respondents' political and financial behavior. This relationship between overprecision in history questions and behavior in unrelated domains suggests that overprecision is an underlying personality trait that is non-domain-specific and contrasts with previous literature which finds that overconfidence depends on the questions domains (e.g., Klayman et al., 1999). To test whether the Subjective Error Method consistently captures overprecision across different domains and, ultimately, whether overprecision is a persistent personality trait, we run a *pre-registered* online survey in a representative sample of the German population.[31]

4.1 Survey Design

The survey was programmed with Qualtrics and consists of blocks of five questions in five different domains (see the full set of questions in Table B.7 in the appendix). All respondents go through all domains, which are:

1. ***Neutral***: In this domain, respondents are flashed for 8 seconds with five different 20×20 matrices of black triangles and gray squares. After each matrix, they are then asked to estimate the number of black triangles in each of the five shown matrices and to report their subjective error. This task is similar to that of Bosch-Rosa et al. (2020) and has the advantage that it is independent of any socio-economic traits, such as wealth or education, and avoids any heterogeneity in experience and prior knowledge across respondents. Respondents are shown five different matrices with the number of triangles ranging from 70 to 330.[32]

[31] The online survey and its analysis were pre-registered at AsPredicted.org (#118284) and the survey was administered by *Bilendi/Respondi.*

[32] See an example of a matrix with 120 black triangles and 280 gray squares in Figure A.5 in the appendix.

2. ***Contemporary history***: In this domain, we ask the five most answered historical questions in the SOEP survey. The contemporary history domain allows us to compare the results of the online survey and those of the SOEP survey in Appendix D, where we show that the answers of respondents are not substantially different across both surveys. It also allows us to replicate the findings on stock price predictions from Section 3.4.2 in Section F.2 in the appendix.

3. ***General knowledge***: In this domain, respondents answer five general knowledge questions and report their subjective error for each one. Some examples of questions in this domain include the number of teeth of an adult polar bear, the number of keys on a concert piano, or the number of African countries that are part of the UN. This domain captures a wide range of knowledge types encompassing many common topics encountered in daily life.

4. ***Future stock prices***: In this domain, respondents are asked to predict the 28-day forecast for the price of five different assets (Benz, Puma, BMW, Deutsche Post, BASF) and report their subjective error for each one. To do so, we employed a widget from *tradingview.com*, which allowed us to show respondent's an interactive price chart over the past year for each stock. This domain enables us to test overprecision in the financial domain and, importantly, measures overprecision of future events. This is important, because, unlike the other domains, in this domain, respondents are asked to guess about something that *will* happen, not something with a correct answer at the time they are asked.

5. ***Economics***: In this domain, respondents answer five questions related to the German economy and report their subjective error for each one. To maintain consistency in the range of answers across questions, this domain is limited to percentage changes. Some examples include the percentage change in the German CPI from 2011 to 2021, the percentage change in the German nominal GDP from 2006 to 2021, or the percentage change in the German DAX from 2014 to 2021. We include this domain given the importance of overprecision in economic decision-making and its closeness to the financial behavior we analyze in the SOEP survey.

All respondents started with the neutral domain since it began with three practice rounds to familiarize respondents with the matrices.[33] After this, we randomized the order of the remaining domains. Importantly, before each domain, we ask respondents to self-report their knowledge of the topic on a scale of 0 (not knowledgeable at all) to 100 (very knowledgeable).

To ensure data quality, we introduced attention checks throughout the survey. Any respondent who failed these checks was automatically excluded from the survey. To account for the use of *Google* or other search engines, at the end of the survey respondents are asked whether they used such methods to answer the questions. Additionally, in the contemporary history, general knowledge, and economic domains we included an extra question that acts as "*Google* control." These questions are presented in the same format as all other questions but are difficult and unlikely to be known by respondents.[34] Importantly, the answers to these questions are not used to measure the overprecision of respondents.

Besides the key dependent variable, we collect demographic variables such as age, gender, years of education, income, and nationality. We also measure mathematical literacy by asking respondents to solve three mathematical problems and ask respondents to self-report the amount of effort they put into answering the survey. In Table B.9 in the appendix, we list all of the variables collected in the survey.

4.2 Data

We collected 1.000 complete responses. As pre-registered, we exclude all respondents who admit to having used a third party to answer our questions from the analysis, respondents who we identify as 'Googlers' by using our control questions, and the lowest five percentiles on the self-reported effort measure. This leaves us with 839 respondents for the baseline

[33] The difference between practice rounds and the main rounds is that in the practice rounds the correct answer is shown to the respondents after answering the questions, which is not the case in the main rounds.

[34] Following our pre-registration, we consider a respondent to have used a search engine if two conditions are met. In any domain: i) answering the *Google* controls correctly and stating a subjective error of zero, and ii) answering correctly and stating zero subjective error for at least three other questions in this domain. If a respondent is flagged as having used search engines, then she is excluded from the analysis. The control questions we used to check if people were using *Google* included: the year in which Joachim Sauer, husband of Angela Merkel, was born (1949), the upper bound in kilograms of the Bantamweight class in female Olympic boxing (54kg), and the percent change of M1 in the Euro area from 2015 to 2021 (86%).

analysis.[35] In Table B.10 in the appendix, we include the summary statistics of the full and cleaned samples. The data-cleaning process does not substantially alter the sample composition with respect to the collected variables.

4.3 Individual Overprecision

In Figure 5 we plot the distributions of the answers across all five domains, similar to Figure 2.[36] For most of the questions across the five domains, the answers are distributed around the true answer, indicating that respondents were paying attention and could answer the questions presented. Nonetheless, it is also clear from the dispersion of the answers and the distance of the mode to the true answer that some questions were easier than others. In Figure 6, we plot the realized error ($error_{i,j}$) on the vertical axis against the subjective error ($se_{i,j}$) on the horizontal axis for each of the five questions in each domain, similar to Fgure 3.[37] We add a 45-degree line, so that any dot above is an overprecise observation ($error_{i,j} > se_{i,j}$) and any point below is underprecise ($error_{i,j} < se_{i,j}$). The results show that respondents are, on average, overprecise across all questions and domains.

4.4 Overprecision Across Domains

For each domain, we construct an aggregate measure of overprecision following the procedure in Section 3.2.[38] For better comparability, we standardize the aggregate measure for each domain to have a zero mean and a standard deviation of one as we did with the SOEP data.[39,40]

[35] In Appendix E, we use different exclusion restrictions on the sample as specified in the pre-registration and show that our results are robust to the different restrictions.

[36] We plot the distribution of the answers to the three questions which we use to detect respondents who we assume to have used search engines in Figure A.3 in the appendix.

[37] We plot the realized error ($error_{i,j}$) against the subjective error ($se_{i,j}$) for the three questions we use to detect respondents who we assume to have used search engines in Figure A.4 in the appendix.

[38] Before aggregating, we divide the answers in the neutral domain by 4 since the scale differs. In our pre-registration, we specified that we would use both the simple average and the principal component across all domains. However, since the results only marginally change, we do not report the results using the principal component as an aggregation method, which are available upon request.

[39] Standardizing the aggregate measure does not change the qualitative results.

[40] To have meaningful estimates and decrease noise, we follow our pre-registration, and for each domain we only constructed the aggregate overprecision measure if the respondent answered at least four out of the five questions.

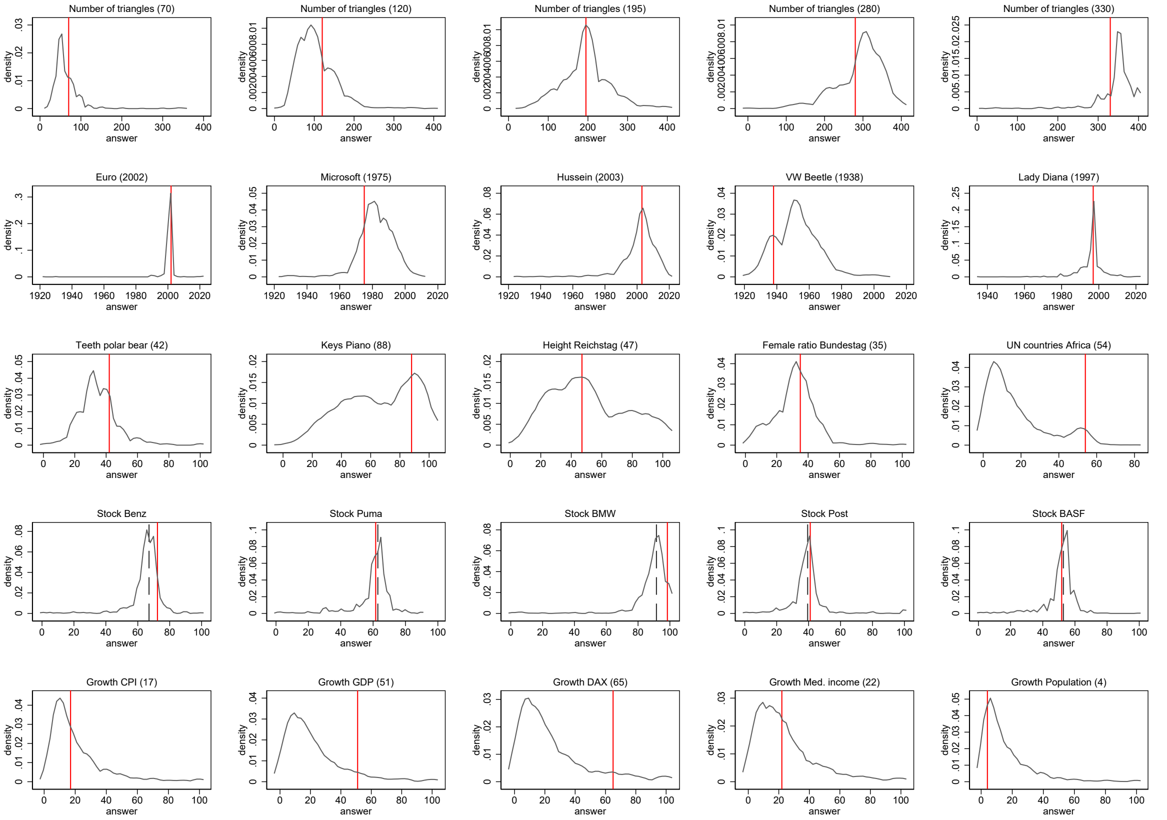


Figure 5: Density of the answers $(a_{i,j})$ for each question. The vertical red line marks the correct answer. For the stocks, the vertical red line marks the weighted average of the 28-day-ahead stock price and the vertical black dashed line the weighted average of the stock price on the day of the interview. Note that the vertical axis is different for each question. The corresponding graphs for the additional questions to detect the use of search engines can be found in Figure A.3 in the appendix.

We analyze the relationship of the domain-specific overprecision measures in three ways: i) via a principal component analysis, ii) via a partial correlation analysis, and iii) via a leave-one-out analysis.[41]

Principal component analysis: The principal component analysis (PCA) is a dimensionality-reducing method that identifies common patterns among the variables and creates new latent variables (the principal components) that capture as much variation (variance and covariance) as possible among the considered variables. All principal components can be ranked by their eigenvalues, with the first principal component capturing the most variation in the data. This first component identifies the most important underlying pat-

[41] We show the robustness of the following results in Appendix E using different exclusion restrictions on the sample as specified in the pre-registration.

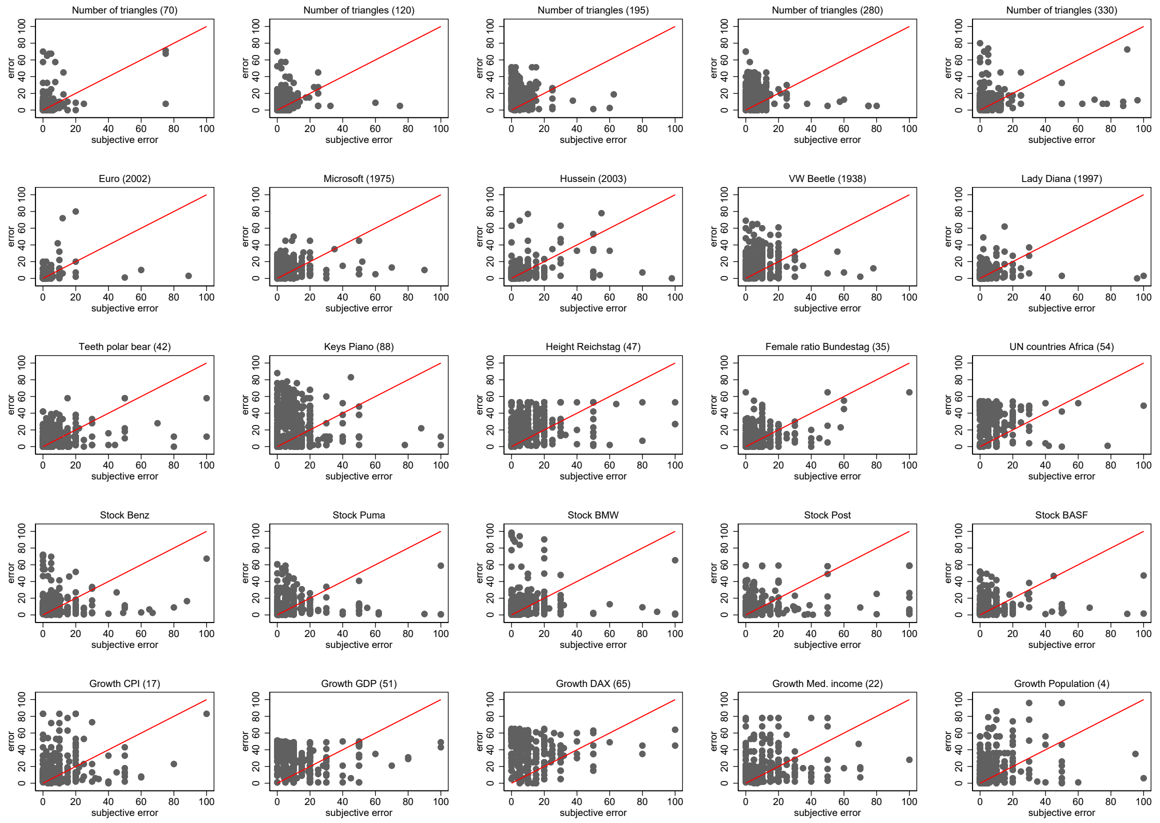


Figure 6: Relation between the realized error ($error_{i,j}$) in the vertical axis and the subjective error ($se_{i,j}$) in the horizontal axis. Any dot above (below) the 45-degree red line is an overprecise (underprecise) answer by the respondent. Note that the results for the neutral domain are divided by four to make the results comparable to the other domains. The corresponding graphs for the additional questions to detect the use of search engines can be found in Figure A.4 in the appendix.

tern in the data and is the most important latent variable in the dataset. Therefore, if it is possible to collapse the different measures of overprecision across domains to a single principal component, it means that there is one "latent trait" across all domains, indicating that overprecision is a domain-free personality trait. The analysis is based on the 552 respondents for which we could calculate an aggregate score for every domain. The results show that only one factor has an eigenvalue above the Kaiser criterion of 1.0. This implies that there is only *one* latent variable that can explain the observed variance and covariances across the five domains. Moreover, the factor loadings across all five domains are positive, which shows that this trait is persistent across the five domains.[42] In other

[42] To be more precise, the eigenvalue of this factor is 2.29, with the respective factor loadings are 0.52, 0.76, 0.74, 0.69, and 0.65 for the neutral, contemporary history, general knowledge, stock prices, and economics domains.

Domain 1	Domain 2	Correlation	Standard error	N
Neutral	History	0.17	0.03	688
Neutral	Knowledge	0.21	0.04	626
Neutral	Stocks	0.30	0.04	676
Neutral	Economics	0.18	0.04	639
History	Knowledge	0.52	0.04	609
History	Stocks	0.35	0.04	638
History	Economics	0.37	0.05	620
Knowledge	Stocks	0.34	0.04	591
Knowledge	Economics	0.32	0.04	580
Stocks	Economics	0.28	0.04	618

Table 3: Partial correlations between domains. This table shows the partial correlations between the aggregate overprecision measures across the five domains. The control variables are age, gender, education, income, nationality, state, mathematical literacy, and a measure of self-reported knowledge on the topic of the domain.

words, the principal component analysis reveals that there is a single underlying factor that can explain the observed variation in the overprecision measures and that all domains contribute positively to it. We interpret this factor as the individual overprecision of respondents.

Partial correlation analysis: Partial correlation analysis allows us to estimate the correlation across two domain-specific overprecision measures, assuming that all other control variables were fixed. Table 3 presents the partial correlation coefficients, adjusted for age, gender, education, income, nationality, state, mathematical literacy, and self-reported domain knowledge. The results show a strong positive correlation across the different domain-specific overprecision measures with the exception of the Neutral domain which exhibits only a moderate positive correlation with the other. We visually confirm the results from Table 3 in Figure 7, where we show a binned scatter plot of the respective aggregate overprecision measure across each pair of domains.

Leave-one-out analysis: The leave-one-out analysis adopts the methodology described in Morrison and Taubinsky (2019). We first partition the overprecision sample for each domain into those respondents in the highest 25% (group 1) and those in the lowest 75% (group 2) according to the overprecision measure in this domain. We then pair this domain with all other domains separately and use a standard t-test to determine if subjects in group 1 also exhibit significantly higher overprecision than respondents in group 2 in the other domains. The intuition of this method is that if the Subjective Error Method

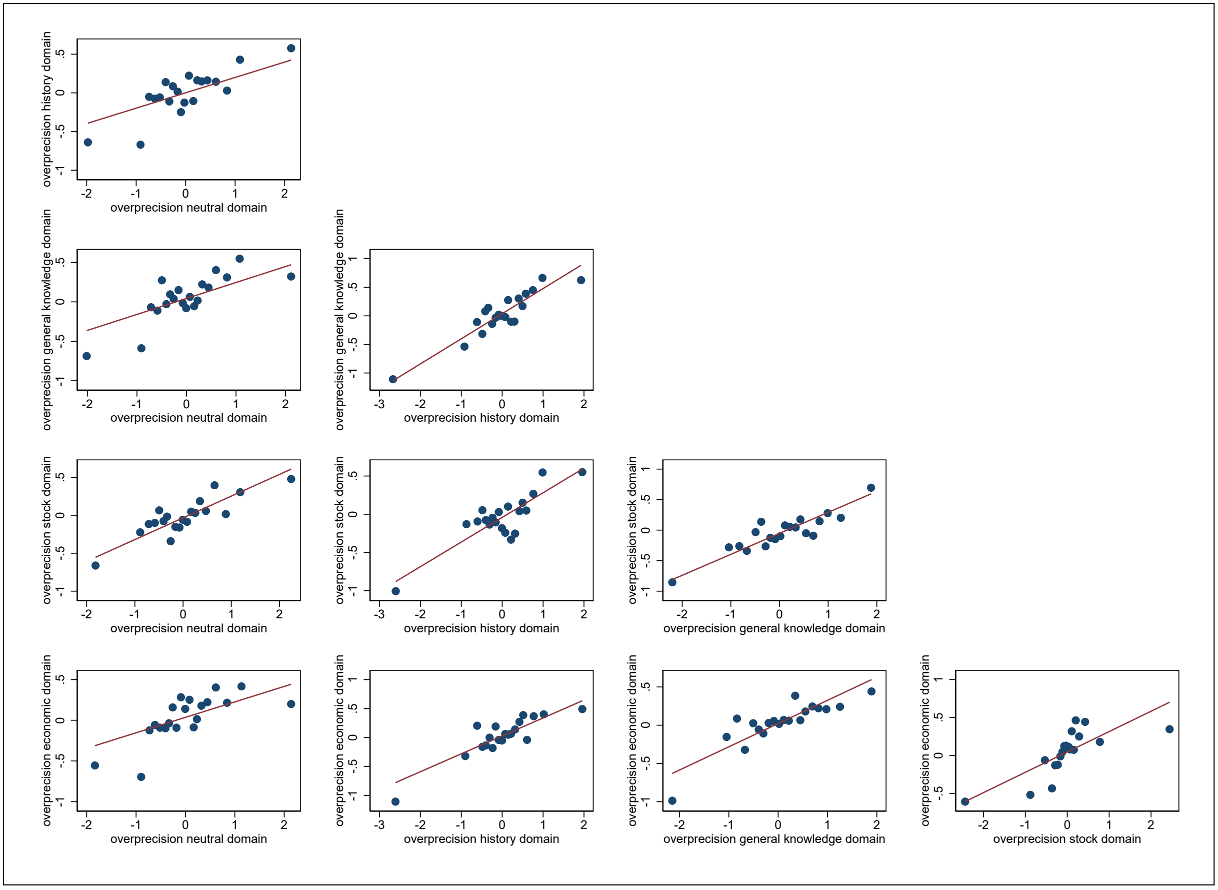


Figure 7: This figure shows binned scatter plots for each pair of overprecision measures. The number of bins is 20 in each plot. The red line is a linear fit between the two dimensions.

is capturing the same trait across domains, then respondents in group 1 in one domain should also be more overprecise than respondents in group 2 in the other domain. In other words, the ranking of overprecision across respondents should be similar across domains. The results are presented in Table 4, where it is clear that the ranking of overprecision is maintained across domains, with more overprecise respondents in one domain also being significantly more overprecise in the other domains.

Taken together, the results from our three pre-registered analyses indicate a strong relationship between our overprecision measures across all domains. This suggests that the Subjective Error Method consistently measures overprecision across domains and that, at a given point in time, it is a stable personal trait. Specifically, the results from the leave-one-out analysis suggest that it is possible to measure overprecision in a different domain than that of the outcome of interest. This supports the analysis in Section 3, where we can correlate the relationship between overprecision measured in the domain of

Sample		Lower 75%			Upper 25%			Difference	
In	Out	mean	sd	N	mean	sd	N	Difference	P-value
Neutral	History	-0.08	1.09	540	0.27	0.71	163	-0.35	< 0.01
Neutral	Knowledge	-0.06	0.88	484	0.35	0.96	149	-0.41	< 0.01
Neutral	Stocks	-0.10	0.92	522	0.24	0.99	168	-0.34	< 0.01
Neutral	Economics	-0.04	0.90	495	0.29	0.69	152	-0.32	< 0.01
History	Neutral	-0.05	0.86	621	0.16	1.23	165	-0.22	0.01
History	Knowledge	-0.09	0.90	489	0.48	0.84	144	-0.57	< 0.01
History	Stocks	-0.11	0.87	536	0.32	1.09	154	-0.43	< 0.01
History	Economics	-0.05	0.90	499	0.34	0.69	148	-0.38	< 0.01
Knowledge	Neutral	-0.05	0.92	629	0.18	1.05	157	-0.23	0.01
Knowledge	History	-0.10	1.06	552	0.35	0.79	151	-0.45	< 0.01
Knowledge	Stocks	-0.09	0.88	541	0.24	1.12	149	-0.32	< 0.01
Knowledge	Economics	-0.03	0.91	499	0.27	0.67	148	-0.30	< 0.01
Stocks	Neutral	-0.10	0.88	621	0.35	1.12	165	-0.45	< 0.01
Stocks	History	-0.09	1.05	551	0.33	0.81	152	-0.42	< 0.01
Stocks	Knowledge	-0.04	0.90	499	0.33	0.90	134	-0.37	< 0.01
Stocks	Economics	-0.05	0.87	495	0.32	0.78	152	-0.37	< 0.01
Economics	Neutral	-0.07	0.93	634	0.27	0.99	152	-0.34	< 0.01
Economics	History	-0.08	1.06	561	0.31	0.77	142	-0.38	< 0.01
Economics	Knowledge	-0.02	0.89	500	0.27	0.97	133	-0.29	< 0.01
Economics	Stocks	-0.10	0.91	542	0.28	1.01	148	-0.38	< 0.01

Table 4: Results of the leave-one-out analysis. *In sample* signifies the domain in which the quartiles based on overprecision were computed. The sample is then partitioned into the lower 75% and upper 25% group. *Out sample* signifies the domain in which overprecision is then measured and tested between both groups.

historical knowledge with respondents' financial and political behavior.

5 Conclusion

We study the correlation between overprecision and the political and financial behavior of a nationally representative sample. To do so, we implement the Subjective Error Method in the 2018 wave of the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). The Subjective Error Method is a new way to measure overprecision that, in contrast to previous methods, is intuitive to respondents, and quick to implement.

Our results show that our overprecision measure aligns with several theoretical predictions from the financial and political science literature. Specifically, overprecision is associated with greater forecasting errors in stock price predictions (Odean, 1998) and lower portfolio diversification (Barber and Odean, 2000). Furthermore, as predicted and shown in Ortoleva and Snowberg (2015a), more overprecise respondents tend to hold more extreme political ideologies. As for the socio-demographic factors influencing overprecision, we find that years of education, age, and gross income reduce respondents' overprecision but do not detect any effect of gender on overprecision. Further, we find a negative relationship between overprecision and financial literacy and, as one would expect, a positive relationship between overprecision and narcissism. Both the relationship with respondents' behavior and with the socio-demographic determinants are robust to a series of robustness tests, further validating the Subjective Error Method as a measure of overprecision.

In fact, to test whether the Subjective Error Method consistently measures respondents' overprecision across domains (and ultimately if overprecision is a personality trait), we conduct a companion online survey across five different domains on a representative sample of the German population. In this survey, we elicit oveprecision in the domains of contemporary history, general knowledge, economics, future stock price predictions, and a "neutral" domain that respondents had not encountered before. The results show a high correlation of our measure across the different domains and confirm the reliability and versatility of the Subjective Error Method. Moreover, the consistency of our measure across domains and its correlation with the behavior of people in different aspects of people's lives suggests that the Subjective Error Method captures a persistent personality trait with real-world implications.

Overall, our work contributes to a literature that tries to understand overconfidence, "the most significant of the cognitive biases" (Kahneman, 2011), and how it affects our lives. Because we show that overprecision is a trait that is robust across domains that can result in reckless behavior and lead to extreme political views, our results and methodology should be of interest not only to economists and political scientists but also to psychologists, financial researchers, policymakers, and educators.

References

Alpert, M. and H. Raiffa (1982): "A progress report on the training of probability assessors," in *Judgment Under Uncertainty: Heuristics and Biases.*, Cambridge University Press, 294–305. Cited on pages 4 and 5.

Bandiera, O., N. Parekh, B. Petrongolo, and M. Rao (2022): "Men Are from Mars, and Women Too: A Bayesian Meta-Analysis of Overconfidence Experiments," *Economica*, 89, 38–70. Cited on page 15.

Barber, B. M. and T. Odean (2000): "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance*, 55, 773–806. Cited on pages 3 and 31.

——— (2001): "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment," *The Quarterly Journal of Economics*, 116, 261–292. Cited on page 2.

Bazerman, M. H. and D. A. Moore (2013): *Judgment in Managerial Decision Making*, New York: Wiley, 8th ed. Cited on pages 2 and 6.

Ben-David, I., J. R. Graham, and C. R. Harvey (2013): "Managerial Miscalibration," *The Quarterly Journal of Economics*, 128, 1547–1584. Cited on page 2.

Benos, A. V. (1998): "Aggressiveness and survival of overconfident traders," *Journal of Financial Markets*, 1, 353–383. Cited on page 18.

Bosch-Rosa, C., D. Gietl, and F. Heinemann (2020): "Risk-Taking under Limited Liability: Quantifying the Role of Motivated Beliefs," Working Paper 210, CRC TRR 190 Rationality and Competition. Cited on page 23.

Camerer, C. and D. Lovallo (1999): "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 89, 306–318. Cited on page 2.

Campbell, W. K., A. S. Goodie, and J. D. Foster (2004): "Narcissism, confidence, and risk attitude," *Journal of Behavioral Decision Making*, 17, 297–311. Cited on page 14.

Cho, E. (2016): “Making Reliability Reliable: A Systematic Approach to Reliability Coefficients,” *Organizational Research Methods*, 19, 651–682. Cited on page 11.

Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamhöfer, and H. Schildberg-Hörisch (2019): “Self-Control: Determinants, Life Outcomes and Intergenerational Implications,” *SOEPpapers on Multidisciplinary Panel Data Research*, 1047. Cited on page 17.

Cronbach, L. J. (1951): “Coefficient alpha and the internal structure of tests,” *Psychometrika*, 16, 297–334. Cited on page 12.

Crosetto, P. and T. De Haan (2022): “Comparing input interfaces to elicit belief distributions,” Working papers, Grenoble Applied Economics Laboratory (GAEL). Cited on page 77.

Daniel, K. and D. Hirshleifer (2015): “Overconfident Investors, Predictable Returns, and Excessive Trading,” *Journal of Economic Perspectives*, 29, 61–88. Cited on page 5.

Deaves, R., J. Lei, and M. Schröder (2019): “Forecaster Overconfidence and Market Survey Performance,” *Journal of Behavioral Finance*, 20, 173–194. Cited on pages 2, 18, and 19.

Deaves, R., E. Lüders, and G. Y. Luo (2009): “An Experimental Test of the Impact of Overconfidence and Gender on Trading Activity,” *Review of Finance*, 13, 555–575. Cited on page 15.

Duttle, K. (2016): “Cognitive Skills and Confidence: Interrelations with Overestimation, Overplacement and Overprecision,” *Bulletin of Economic Research*, 68, 42–55. Cited on page 15.

Enke, B. and T. Graeber (2021): “Cognitive Uncertainty,” Working Paper 26518, National Bureau of Economic Research. Cited on page 9.

Glaser, M. and M. Weber (2007): “Overconfidence and trading volume,” *The Geneva Risk and Insurance Review*, 32, 1–36. Cited on page 4.

Goetzmann, W. N. and A. Kumar (2008): "Equity Portfolio Diversification," *Review of Finance*, 12, 433–463. Cited on pages 2 and 20.

Griffin, D. and L. Brenner (2004): "Perspectives on probability judgment calibration," in *Blackwell Handbook of Judgment and Decision Making*, Blackwell Publishing Ltd., 177–199. Cited on page 6.

Grubb, M. D. (2015): "Overconfident Consumers in the Marketplace," *Journal of Economic Perspectives*, 29, 9–36. Cited on page 2.

Hamurcu, C. and H. D. Hamurcu (2021): "Can financial literacy overconfidence be predicted by narcissistic tendencies?" *Review of Behavioral Finance*, 13, 438–449. Cited on page 15.

Haran, U., D. A. Moore, and C. K. Morewedge (2010): "A simple remedy for overprecision in judgment," *Judgment and Decision Making*, 5, 467–476. Cited on page 6.

Hilary, G. and L. Menzly (2006): "Does Past Success Lead Analysts to Become Overconfident?" *Management Science*, 52, 489–500. Cited on pages 18 and 19.

Johnson, D. D. P. (2004): *Overconfidence and War*, Harvard University Press. Cited on page 2.

Kahneman, D. (2011): *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 1st ed. Cited on pages 2 and 32.

Klayman, J., J. B. Soll, C. Gonzalez-Vallejo, and S. Barlas (1999): "Overconfidence: It depends on how, what, and whom you ask," *Organizational behavior and human decision processes*, 79, 216–247. Cited on page 23.

Kramer, M. M. (2016): "Financial literacy, confidence and financial advice seeking," *Journal of Economic Behavior & Organization*, 131, 198–217. Cited on page 16.

López-Pérez, R., A. Rodriguez-Moral, and M. Vorsatz (2021): "Simplified mental representations as a cause of overprecision," *Journal of Behavioral and Experimental Economics*, 92, 101681. Cited on page 15.

Malmendier, U. and T. Taylor (2015): "On the Verges of Overconfidence," *Journal of Economic Perspectives*, 29, 3–8. Cited on page 2.

Mannes, A. E. and D. A. Moore (2013): "A Behavioral Demonstration of Overconfidence in Judgment," *Psychological Science*, 24. Cited on page 2.

McKenzie, C. R. M., M. J. Liersch, and I. Yaniv (2008): "Overconfidence in interval estimates: What does expertise buy you?" *Organizational Behavior and Human Decision Processes*, 107, 179–191. Cited on pages 4, 9, and 73.

Merkle, C. (2017): "Financial overconfidence over time: Foresight, hindsight, and insight of investors," *Journal of Banking & Finance*, 84, 68–87. Cited on page 20.

Miller, P. R. and P. J. Conover (2015): "Red and Blue States of Mind: Partisan Hostility and Voting in the United States," *Political Research Quarterly*, 68, 225–239. Cited on page 22.

Moore, D. A. and P. J. Healy (2008): "The trouble with overconfidence," *Psychological Review*, 115, 502–517. Cited on page 2.

Moore, D. A. and D. Schatz (2017): "The three faces of overconfidence," *Social and Personality Psychology Compass*, 11, 1–12. Cited on page 2.

Moore, D. A. and S. A. Swift (2011): "The Three Faces of Overconfidence in Organizations," in *Social psychology and organizations*, Routledge/Taylor & Francis Group, 147–184. Cited on page 21.

Moore, D. A., E. R. Tenney, and U. Haran (2015): "Overprecision in Judgment," in *The Wiley Blackwell Handbook of Judgment and Decision Making*, John Wiley & Sons, Ltd, 182–209. Cited on pages 2, 5, and 6.

Morrison, W. and D. Taubinsky (2019): "Rules of Thumb and Attention Elasticities: Evidence from Under- and Overreaction to Taxes," Working Paper 26180, National Bureau of Economic Research. Cited on page 29.

Odean, T. (1998): "Volume, Volatility, Price, and Profit When All Traders Are Above Average," *The Journal of Finance*, 53, 1887–1934. Cited on pages 3, 5, 18, 20, and 31.

Önkal, D., J. F. Yates, C. Simga-Mugan, and S. Öztin (2003): "Professional vs. amateur judgment accuracy: The case of foreign exchange rates," *Organizational Behavior and Human Decision Processes*, 91, 169–185. Cited on pages 9 and 73.

Ortoleva, P. and E. Snowberg (2015a): "Are conservatives overconfident?" *European Journal of Political Economy*, 40, 333–344. Cited on pages 2, 21, 31, 58, and 65.

——— (2015b): "Overconfidence in Political Behavior," *American Economic Review*, 105, 504–535. Cited on pages 2, 3, 5, 6, 12, 14, 15, 16, 21, 22, 53, 55, 58, and 60.

Pennycook, G., Z. Epstein, M. Mosleh, A. Arechar, D. Eckles, and D. Rand (2021): "Shifting attention to accuracy can reduce misinformation online," *Nature*, 592, 590–595. Cited on page 5.

Prims, J. P. and D. A. Moore (2017): "Overconfidence over the lifespan," *Judgment and Decision Making*, 12, 29–41. Cited on pages 16 and 58.

Rammstedt, B. and O. P. John (2007): "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, 41, 203–212. Cited on page 17.

Richter, D. and J. Schupp (2015): "The SOEP Innovation Sample (SOEP IS)," *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 135, 389–400. Cited on page 9.

Russo, J. E. and P. J. H. Schoemaker (1992): "Managing Overconfidence," *Sloan Management Review*, 33, 7–17. Cited on page 6.

Scheinkman, J. A. and W. Xiong (2003): "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111, 1183–1220. Cited on page 5.

Stone, D. F. (2019): ""Unmotivated bias" and partisan hostility: Empirical evidence," *Journal of Behavioral and Experimental Economics*, 79, 12–26. Cited on pages 2 and 5.

Svenson, O. (1981): "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica*, 47, 143–148. Cited on page 2.

TEIGEN, K. H. AND M. JØRGENSEN (2005): "When 90% Confidence Intervals are 50% Certain: On the Credibility of Credible Intervals," *Applied Cognitive Psychology*, 19, 455–475. Cited on pages 2 and 5.

THALER, M. (2023): "The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News," Tech. rep., mimeo (forthcoming AEJ: Microeconomics). Cited on pages 2 and 5.

TIBSHIRANI, R. (1996): "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288. Cited on page 18.

WOHLEBER, R. W. AND G. MATTHEWS (2016): "Multiple facets of overconfidence: Implications for driving safety," *Transportation Research Part F: Traffic Psychology and Behaviour*, 43, 265–278. Cited on page 15.

A Extra Figures

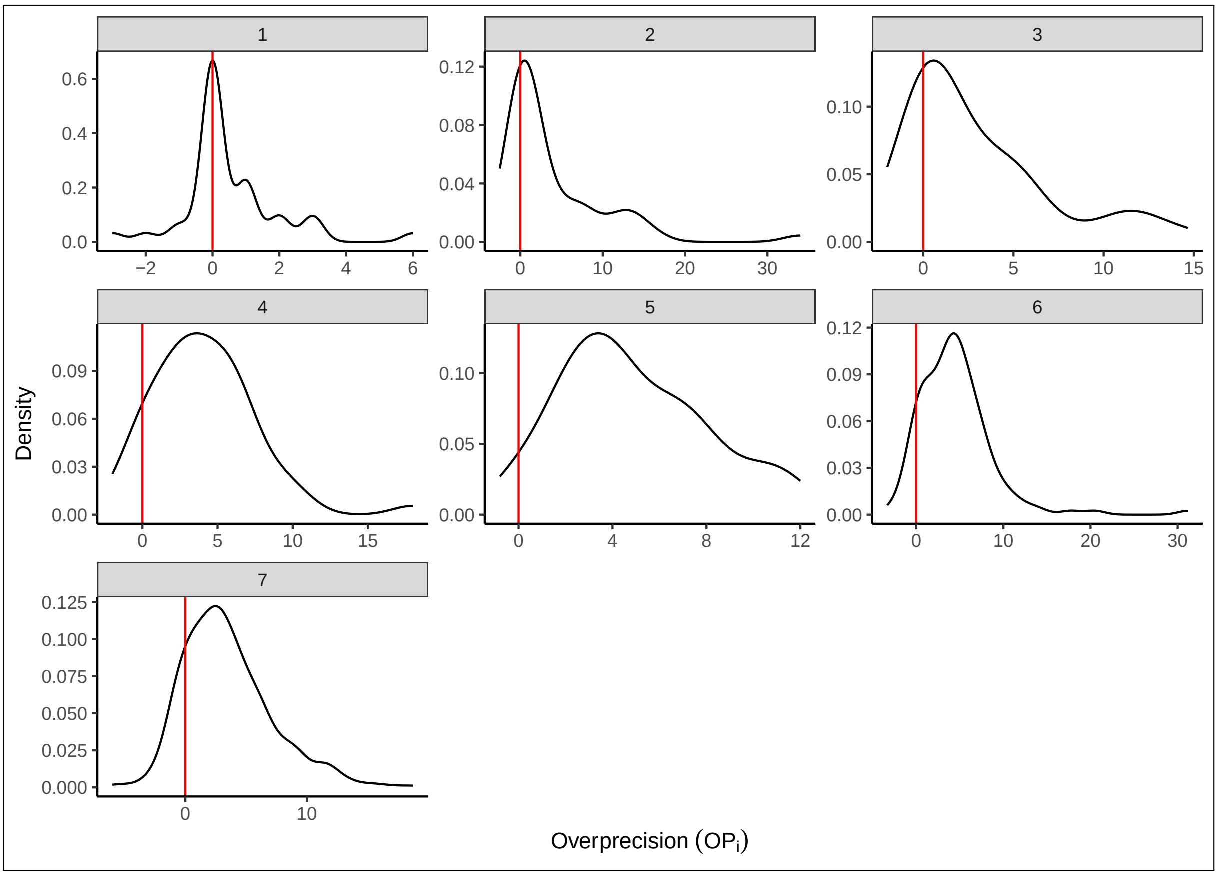


Figure A.1: Density of Overprecision (op_i) for each of the subsets of questions answered. We plot from left to right the densities of op_i for those respondents who answered from the minimum number of answers (1) to the maximum number of answers (7). In the title, we report the number of respondents for each density. Notice that the scale of the Y-axis changes across panels.

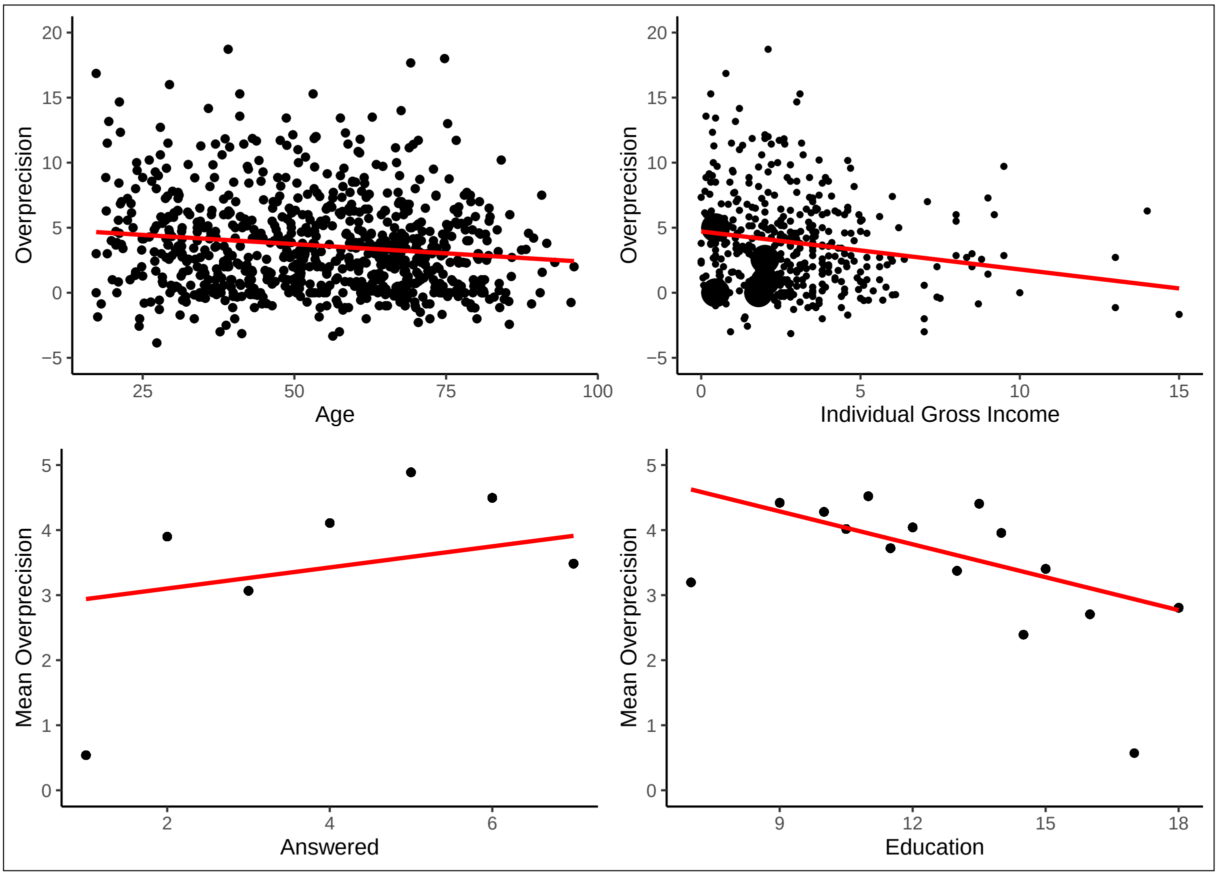


Figure A.2: Correlation of Overprecision. In the vertical axis of each panel, we plot the overprecision (upper row) and mean overprecision across all groups which we plot in the horizontal axis (lower row). In all four cases, the red line is the fitted linear regression. We dropped one individual outlier in all cases to make the graphs more readable.

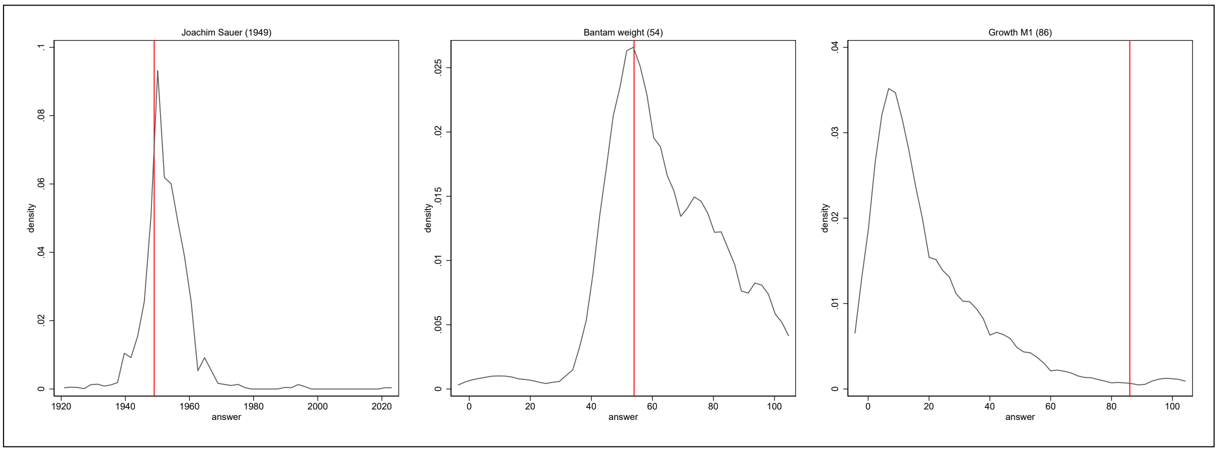


Figure A.3: Density of the answers ($a_{i,j}$) for each of the questions which we use to detect respondents who we assume to have used search engines. The vertical line marks the correct answer. Note that the vertical axis is different for each question.

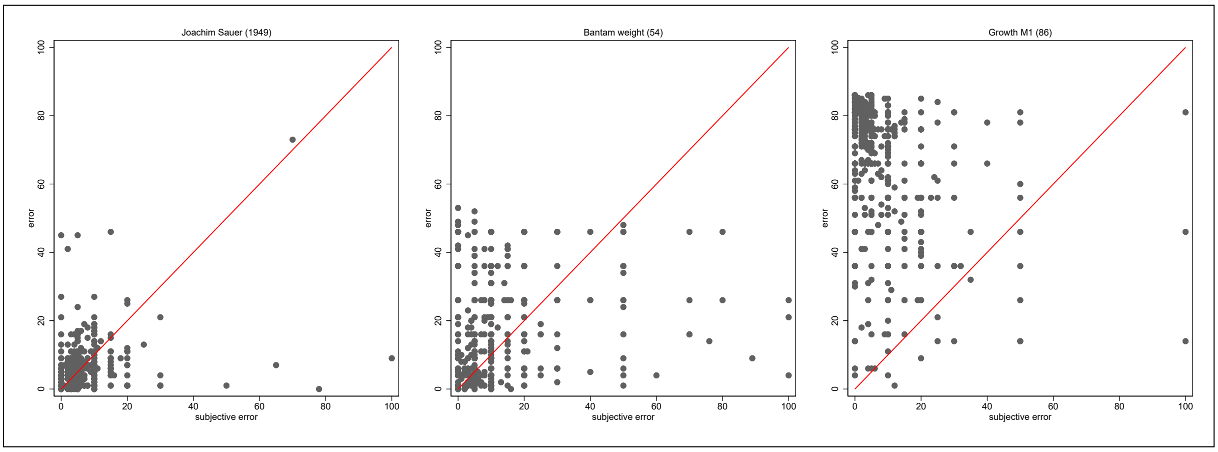


Figure A.4: Relation between the realized error ($error_{i,j}$) in the vertical axis and the subjective error ($se_{i,j}$) in the horizontal axis for each of the questions which we use to detect respondents who we assume to have used search engines. Any dot above (below) the 45-degree red line is an overprecise (underprecise) answer by the respondent.

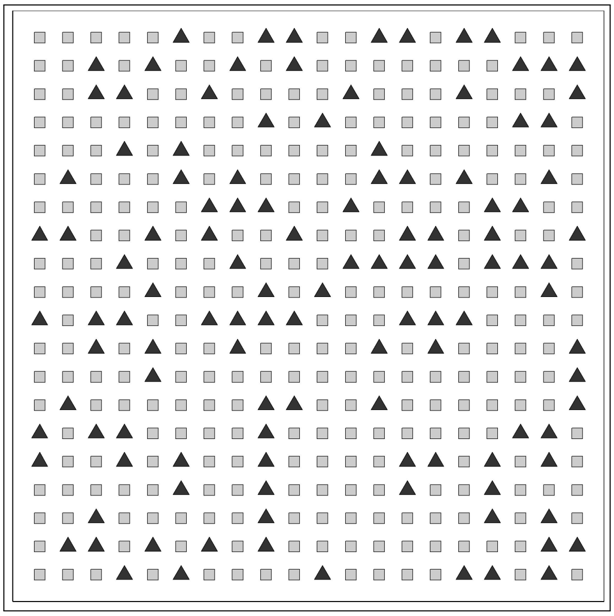


Figure A.5: Matrix Example. Example of one of the matrices use in the neutral domain of the online survey. The matrix contains 120 black triangles and 280 gray squares.

B Extra Tables

SOEP-IS Code	Question (a)	Answer
Q467 - IGEN02a	In which year were euro notes and coins introduced?	2002
Q470 - IGEN03a	In which year was Microsoft (Publisher of the software package Windows) founded?	1975
Q473 - IGEN04a	In which year was the movie “Das Boot” (directed by Wolfgang Peterson) first shown in German cinemas?	1981
Q476 - IGEN05a	In which year was Saddam Hussein captured by the US army?	2003
Q479 - IGEN06a	In which year was the first Volkswagen Type 1 (also known as “Volkswagen Beetle”) produced?	1938
Q482 - IGEN07a	In which year did the Korean War end with a truce?	1953
Q485 - IGEN08a	In which year did Lady Diana, Prince Charles’ first wife, die?	1997
	Question (b)	
	What do you think: How far is your answer off the correct answer?	

Table B.1: Original history questions in English language from the 2018 SOEP-IS

SOEP-IS Code	Questions (a)	Answer
Q467 - IGEN02a	In welchem Jahr wurden Euro-Geldscheine und -Münzen eingeführt?	2002
Q470 - IGEN03a	In welchem Jahr wurde das Unternehmen Microsoft (Herausgeber des Betriebssystems Windows) gegründet?	1975
Q473 - IGEN04a	In welchem Jahr kam der Film “Das Boot” (Regie: Wolfgang Petersen) in die deutschen Kinos?	1981
Q476 - IGEN05a	In welchem Jahr wurde Saddam Hussein von der US-Armee gefangen genommen?	2003
Q479 - IGEN06a	In welchem Jahr wurde der erste Volkswagen Typ 1(auch bekannt als “Käfer”) produziert?	1938
Q482 - IGEN07a	In welchem Jahr endete der Korea-Krieg mit einem Waffenstillstand?	1953

Q485 - IGEN08a	In welchem Jahr starb Lady Diana, die erste Frau von Prinz Charles?	1997
	Question (b)	
	Was schätzen Sie: wie viele Jahre liegt Ihre Antwort von der richtigen Antwort entfernt?	

Table B.2: Original history questions in German language from the 2018 SOEP-IS

Variable	Definition
Financial Behavior:	
DAX forecast error	Principal component of the absolute distance between one- and two-year-ahead prediction of the DAX realization and the actual realization over the horizon. The closing price of the date of the respective interview was used. Standardized to have mean 0 and standard deviation 1.
portfolio diversification	Aggregate diversification measure over five asset classes. For each asset class, a penalty score is calculated expressing the distance to an equally diversified portfolio. Diversification equals the maximum attainable penalty score less the actual penalty. The diversification measure is standardized to have mean 0 and standard deviation 1.
Political Behavior:	
extremeness	Absolute distance to the center of an ideology scale from 0 (left) to 10 (right). Standardized to have mean 0 and standard deviation 1.
left-right	Location on an ideology scale from 0 (left) to 10 (right). Standardized to have mean 0 and standard deviation 1.
non-voter	=1 if respondent indicated not to vote in the Sonntagsfrage (ex-post) for the Bundestagswahl 2017.
Controls:	
age	Difference between interview month/year and birth month/year in years.
female	=1 if female.
GDR 1989	=1 if living in East Germany in 1989.

Variable	Definition
years education	Years of education (including any further education after primary and secondary education)=.
gross income	Monthly gross labor income in thousands. Missings are coded with a zero.
missing income	=1 if missing gross income.
fin. literacy	Share of correct answers to 6 questions related to financial knowledge.
risk aversion	Location on a risk scale from 0 (risk avers) to 10 (risk loving). Standardized to have mean 0 and standard deviation 1.
narcissism	Average narcissism measure over 6 items on a scale from 1 to 6. Standardized to have mean 0 and standard deviation 1.
impulsivity	Location on impulsivity scale from 0 (not impulsive) to 10 (fully impulsive). Standardized to have mean 0 and standard deviation 1.
patience	Location on the patience scale from 0 (not patient) to 10 (fully patient). Standardized to have mean 0 and standard deviation 1.
employed	=1 if employed.
unemployed	=1 if unemployed.
nonwork	=1 if non-working.
matedu	=1 if on maternity, educational, or military leave.
retired	=1 if retired.
answered	Number of questions answered for overprecision.
Additional Controls:	
assets	=1 if owning financial assets.
pol. interest	Political interest on a scale from 1 (high) to 4 (low). Reversed and standardized to have mean 0 and standard deviation 1.

Table B.3: Overview and definition of the variables from the SOEP used in the analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	SOEP-IS		SOEP Core		Difference		
	mean	sd	mean	sd	difference	p-value	N[Core]
age	53.914	(0.627)	50.535	(0.180)	-3.379	0.000	30,997
female	0.508	(0.018)	0.508	(0.005)	0.000	0.989	30,997
german	0.933	(0.009)	0.877	(0.003)	-0.056	0.000	30,997
east (current)	0.174	(0.013)	0.172	(0.003)	-0.001	0.916	30,997
GDR 1989	0.186	(0.014)	0.198	(0.004)	0.012	0.404	24,591
years education	12.704	(0.098)	12.276	(0.027)	-0.428	0.000	28,482
employed	0.534	(0.018)	0.593	(0.005)	0.058	0.001	30,967
retired	0.229	(0.015)	0.221	(0.004)	-0.007	0.627	30,967
gross income	2.943	(0.112)	2.837	(0.029)	-0.106	0.359	17,829
married	0.568	(0.017)	0.521	(0.005)	-0.047	0.009	30,896
N[SOEP-IS]	805						

Table B.4: Representativeness of the SOEP-IS subsample. This table shows the descriptives of selected personal characteristics of the respondents for the subsample of the SOEP-IS and the SOEP-Core. The results for the SOEP-IS in Columns (1) and (2) are unweighted whereas the results for the SOEP-Core in Columns (3) and (4) are weighted using the sampling weights provided. Columns (5) and (6) show a simple t-test on the difference between the means. Column (7) shows the sample size of the SOEP-Core. The sample size varies due to missing observations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Point	Unadj.	SH	R^2	LASSO		
	estimate	p-value	p-value	rank	included	R^2	N
Financial Behavior:							
DAX forecast error	0.104**	0.024	0.071	4/46	yes/13	0.09	510
1-year ahead	*0.631*	*0.251*		10/46	yes/18	0.14	*537*
2-year ahead	*3.944******	*0.004*		4/46	no/0	0.00	*519*
portfolio diversification	-0.12***	0.002	0.009	4/46	yes/15	0.13	719
Political Behavior:							
extremeness	0.077*	0.059	0.115	8/47	yes/18	0.07	716
left-right	-0.019	0.643	0.643	24/47	no/17	0.10	716
non-voter	0.029**	0.015	0.060	3/47	yes/10	0.12	706

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: This table shows the estimation results of Section 3.4 including the Big Five personality traits. The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 750. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)). Column (4) displays the result from the R^2 procedure as specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Point	Unadj.	SH	R^2	LASSO	LASSO	
	estimate	p-value	p-value	rank	included	R^2	N
Financial Behavior:							
DAX forecast error	0.081*	0.065	0.125	6/42	yes/16	0.10	545
1-year ahead	*0.477*	*0.360*		8/42	yes/24	0.17	*574*
2-year ahead	*3.044***	*0.018*		5/42	no/0	0.00	*553*
portfolio diversification	-0.131***	0.000	0.002	4/42	yes/18	0.13	763
Political Behavior:							
extremeness	0.082*	0.057	0.161	6/43	yes/14	0.06	706
left-right	-0.002	0.966	0.966	18/43	no/14	0.08	706
non-voter	0.031**	0.014	0.054	3/43	yes/10	0.10	694

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.6: This table shows the estimation results of Section 3.4 including asset ownership as control. The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 791. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)). Column (4) displays the result from the R^2 procedure as specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the R^2 of the estimated model (Column (6)).

Qualtrics Code	Question (a)	
PX110 a	How many black triangles were in the matrix?	70
PX110 b	How many black triangles were in the matrix?	120
PX110 c	How many black triangles were in the matrix?	195
PX110 d	How many black triangles were in the matrix?	280
PX110 e	How many black triangles were in the matrix?	330
PX120 a	In which year were euro notes and coins introduced?	2002
PX120 b	In which year was Microsoft (Publisher of the software package Windows) founded?	1975
PX120 c	In which year was Saddam Hussein captured by the US army?	2003
PX120 d	In which year was the first Volkswagen Type 1 (also known as "Volkswagen Beetle") produced?	1938
PX120 e	In which year did Lady Diana, King Charles' first wife, die?	1997
PX120 f	In which year was Joachim Sauer (husband of Angela Merkel) born?	1949
PX120 g	Which year is between 1993 and 1995?	1994
PX130 a	How many teeth does an adult polar bear have?	42
PX130 b	How many keys (black AND white) does a grand piano have?	88
PX130 c	How high (in meters) is the Reichstag building in Berlin?	47
PX130 d	What percentage of seats in the 20th German Bundestag (elected on September 26, 2021) are occupied by female members of the Bundestag	35
PX130 e	How many countries on the African continent are members of the United Nations?	54
PX130 f	What is the upper bantamweight limit in women's Olympic boxing weight classes?	54
PX130 g	What whole number is between 83 and 85?	84
PX140 a-e	What do you estimate: Where (in euros) will the share price be in exactly four weeks?	+ 28 days
PX140 f	In this question, the graph is empty. Please just enter the number 42 as your answer. What do you estimate: where (in euros) will the share price be in exactly four weeks?	42

PX150 a	The Consumer Price Index (CPI) is a measure of the average percentage change in the price level of certain goods and services purchased by households for consumption. The change in the consumer price index compared to the same month of the previous year or the previous year is also referred to as the rate of inflation. By how much (in percent) did the German CPI increase from the beginning of 2011 to the end of 2021?	17
PX150 b	The gross domestic product (GDP) indicates the total value of all goods and services that were produced as end products within the national borders of an economy during a year, after deduction of all intermediate consumption. By how much (in percent) is the German GDP at market prices (nominal) increased from 2006 to 2021?	51
PX150 c	The DAX (Deutscher Aktienindex) is a stock index that measures the performance of the 40 largest companies in the German stock market. By how much (in percent) did the DAX rise from the beginning of 2014 to the end of 2021?	65
PX150 d	The middle income or median income in a society or group describes the income level at which the number of households (or persons) with lower incomes is equal to the number of households with higher incomes. By how much (in percent) did median income in Germany increase from 1991 to 2018?	22
PX150 e	A census is a legally ordered survey of statistical population data. The last census in Germany took place in 2022. By how much (in percent) did the German population grow from 1991 to 2021?	4
PX150 f	M1 money supply describes the amount of cash in circulation and the amount of sight deposits (e.g. savings accounts). By how much (in percent) did the M1 money supply in the euro zone increase from the beginning of 2015 to the end of 2021?	86
	Question (b)	
	What do you think: How many [unit] is your answer away from the correct answer?	

Table B.7: Original questions in English language from the online survey

Qualtrics Code	Questions (a)	Answer
PX110 a	Wie viele schwarze Dreiecke waren in der Matrix?	70
PX110 b	Wie viele schwarze Dreiecke waren in der Matrix?	120
PX110 c	Wie viele schwarze Dreiecke waren in der Matrix?	195
PX110 d	Wie viele schwarze Dreiecke waren in der Matrix?	280
PX110 e	Wie viele schwarze Dreiecke waren in der Matrix?	330
PX120 a	In welchem Jahr wurden Euro-Geldscheine und -Münzen eingeführt?	2002
PX120 b	In welchem Jahr wurde das Unternehmen Microsoft (Herausgeber des Betriebssystems Windows) gegründet?	1975
PX120 c	In welchem Jahr wurde Saddam Hussein von der US-Armee gefangen genommen?	2003
PX120 d	In welchem Jahr wurde der erste Volkswagen Typ 1 (auch bekannt als “Käfer”) produziert?	1938
PX120 e	In welchem Jahr starb Lady Diana, die erste Frau von König Charles III.?	1997
PX120 f	In welchem Jahr wurde Joachim Sauer (Ehemann von Angela Merkel) geboren?	1949
PX120 g	Welches Jahr liegt zwischen 1993 und 1995?	1994
PX130 a	Wie viele Zähne hat ein ausgewachsener Eisbär?	42
PX130 b	Wie viele Tasten (schwarz UND weiß) hat ein Konzertflüegel?	88
PX130 c	Wie hoch (in Metern) ist das Berliner Reichstagsgebäude?	47
PX130 d	Wie viel Prozent der Sitze im 20. Deutschen Bundestag (gewählt am 26. September 2021) sind durch weibliche Bundestagsabgeordnete besetzt?	35
PX130 e	Wie viele Staaten auf dem Afrikanischen Kontinent sind Mitglied der Vereinten Nationen?	54

PX130 f	Bei wie viel Kilogramm liegt die Obergrenze des Bantamgewichts in den Gewichtsklassen der Frauen beim Olympischen Boxen?	54
PX130 g	Welche ganze Zahl liegt zwischen 83 und 85?	84
PX140 a-e	Was schätzen Sie: wo (in Euro) liegt der Aktienkurs in genau vier Wochen?	+28 Tage
PX140 f	In dieser Frage ist die Grafik leer. Bitte geben Sie einfach die Zahl 42 als Antwort ein. Was schätzen Sie: wo (in Euro) liegt der Aktienkurs in genau vier Wochen?	42
PX150 a	Der Verbraucherpreisindex (VPI) ist ein ist ein Maß der durchschnittlichen prozentualen Veränderung des Preisniveaus bestimmter Waren und Dienstleistungen, die von privaten Haushalten für Konsumzwecke gekauft werden. Die Veränderung des Verbraucherpreisindex zum Vorjahresmonat bzw. zum Vorjahr wird auch als Teuerungsrate oder als Inflationsrate bezeichnet. Um wie viel (in Prozent) ist der deutsche VPI von Anfang 2011 bis Ende 2021 gestiegen?	17
PX150 b	Das Bruttoinlandsprodukt (BIP) gibt den Gesamtwert aller Waren und Dienstleistungen an, die während eines Jahres innerhalb der Landesgrenzen einer Volkswirtschaft als Endprodukte hergestellt wurden, nach Abzug aller Vorleistungen. Um wie viel (in Prozent) ist das deutsche BIP zu Marktpreisen (nominal) von 2006 bis 2021 gestiegen?	51
PX150 c	Der DAX (Deutscher Aktienindex) ist ein Aktienindex, der die Wertentwicklung der 40 größten Unternehmen des deutschen Aktienmarkts misst. Um wie viel (in Prozent) ist der DAX von Anfang 2014 bis Ende 2021 gestiegen?	65
PX150 d	Das mittlere Einkommen oder Medianeinkommen in einer Gesellschaft oder Gruppe bezeichnet die Einkommenshöhe, von der aus die Anzahl der Haushalte (bzw. Personen) mit niedrigeren Einkommen gleich groß ist wie die der Haushalte mit höheren Einkommen. Um wie viel (in Prozent) ist das Medianeinkommen in Deutschland von 1991 bis 2018 gestiegen?	22

PX150 e	Eine Volkszählung oder auch Zensus ist eine gesetzlich angeordnete Erhebung statistischer Bevölkerungsdaten. Der letzte Zensus in Deutschland fand 2022 statt. Um wie viel (in Prozent) ist die deutsche Bevölkerung von 1991 bis 2021 gewachsen?	4
PX150 f	Die Geldmenge M1 bezeichnet die Menge an Bargeld im Umlauf sowie die Höhe an Sichteinlagen (bspw. Sparkonten). Um wie viel (in Prozent) ist die Geldmenge M1 in der Eurozone von Anfang 2015 bis Ende 2021 gestiegen?	86
	Question (b)	
	Was schätzen Sie: wie viele [Einheit] liegt Ihre Antwort von der richtigen Antwort entfernt?	

Table B.8: Original questions in German language from the online survey

Variable	Definition
Controls:	
age	Reported age.
gender	=1 if female.
education	Years of primary and secondary education.
income	Monthly gross labor income in thousands, reported in bins.
nationality	Reported primary nationality.
mathematical literacy	Share of correct answers to three statistical problems.
expertise	Self-reported expertise on specific domain on a scale from 0 to 100.

Table B.9: Overview and definition of the variables from the survey used in the analysis.

	count	mean	sd	p25	p50	p75
Full sample						
age	1000	44.950	14.315	33.000	46.000	58.000
gender	998	0.489	0.500	0.000	0.000	1.000
german	1000	0.962	0.191	1.000	1.000	1.000
east (current)	1000	0.200	0.400	0.000	0.000	0.000
education	988	10.685	1.871	10.000	10.000	12.000
gross income	924	2.456	1.856	1.500	2.500	4.000
N	1000					
Subsample						
age	839	44.897	14.213	33.000	46.000	57.000
gender	837	0.483	0.500	0.000	0.000	1.000
german	839	0.969	0.173	1.000	1.000	1.000
east (current)	839	0.203	0.402	0.000	0.000	0.000
education	831	10.693	1.863	10.000	10.000	12.000
gross income	778	2.485	1.898	1.500	2.500	4.000
N	839					

Table B.10: Summary statistics of the online companion survey. This table shows summary statistics for the main variables in the survey.

C Alternative Measures of Overprecision

To test the robustness of our overprecision measure, in Section C.1 we discuss five alternative measures of overprecision, which are variations of our measure. In Section C.2 we use these alternative measures to test the robustness of our results from Section 3.3 regarding the socio-demographic characteristics and Section C.3 the robustness of the predictions in Section 3.4.2.

C.1 Alternative Measures

Standardized measure (op_i'): Since the overprecision measure of Ortoleva and Snowberg (2015b) standardizes the measure with respect to the entire population, we further construct a *standardized* measure op_i' of overprecision where we standardize the absolute measure op_i of the respective question to be mean zero and standard deviation one before aggregation to avoid the aggregated measure to be biased by a specific question and to relate the level to the entire population. The mean is used again to aggregate across the seven questions.

Centered measure (op_i''): Respondents might not only differ with respect to the perceived variance of the distribution of the error to their answer, but also with respect to the mean of the distribution. Hence, the baseline overprecision measure might capture both overprecision and a miscalibration of the mean. To separate both of them, we construct a *centered* measure of overprecision. To correct for the difference in the means of the distributions and center the distributions around zero, for each question, we subtract the sample mean from the true and subjective error. Any remaining systematic deviation of the subjective error from the realized error should be exclusively due to over- or underprecision.

Relative measure (op_i'''): To circumvent the classification problem of the residual approach (op_i''''') we compute a *relative* measure op_i''' by dividing the absolute measure op_i in a specific question with the respective subjective error. Taking the relative distance into account makes the measure more comparable across respondents while still keeping the relative distance between the subjective error and the realized error (see Figure C.1).

Assume that, similar to the example in Figure 1, the true error is normally distributed with mean 0 and variance σ^2 (solid curve). Moreover, the perceived distribution by the

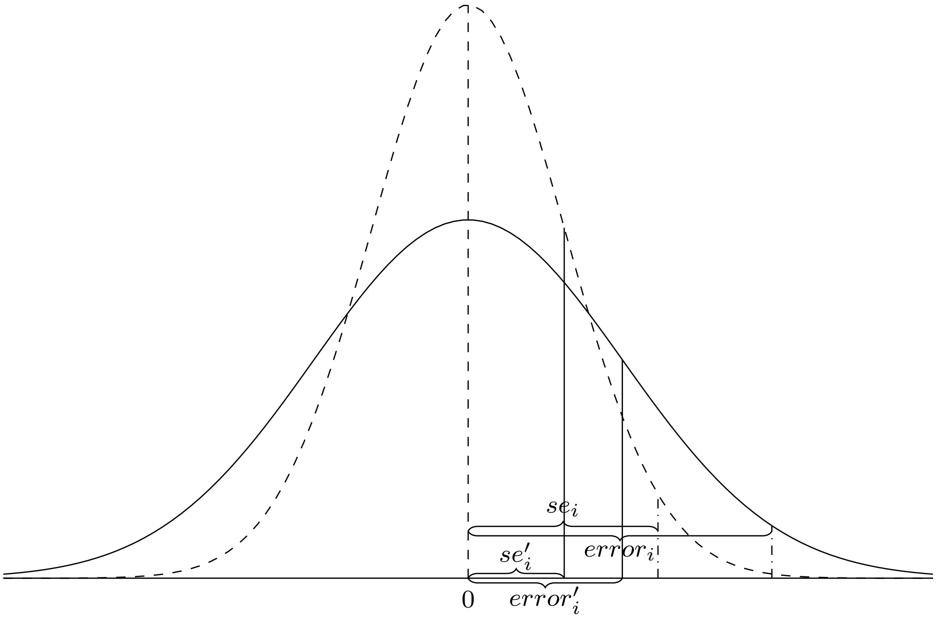


Figure C.1: Two distributions of the (subjective) error. The solid curve shows the true distribution of the error with a standard deviation of 2 (precision of .25). The dashed red curve shows the perceived distribution by an overprecise respondent with a standard deviation of 1.25 (precision of .64). The solid and dashed vertical lines indicate the subjective errors (*se*) and the true errors (*error*) resulting from respondents with two different ideas about the nature of the subjective error asked in the second question.

respondents might not necessarily coincide with the true distribution. If the perceived variance $\hat{\sigma}^2$ is smaller, i.e., the precision $\rho = 1/\hat{\sigma}^2$ is larger, then we call this respondent overprecise (dashed curve). As long as respondents have the same idea in mind when asking for the error they expect to make, the absolute overprecision measure is comparable across subjects. However, when respondents substantially differ, e.g., by having different confidence intervals in mind, the ranking as computed with the absolute measure might not be consistent anymore whilst the sign of the deviation still being correct. Taking the example in Figure C.1, where the respondents have the same degree of overprecision since the perceived precision of .64 deviates from the true precision of .25, for a respondent with having 95% confidence in mind (*se* and *error*) the absolute overprecision measure would yield 1.47 whereas for the respondent with having 68% confidence in mind (se' and $error'$) it would yield .75. Thus, the second respondent would incorrectly be classified as less overprecise.

The relative measure corrects this inconsistency by scaling the absolute overprecision measure with the subjective absolute error, making the measure comparable across subjects. In the above example, the relative measure yields .6 in both cases, which is precisely the relative difference between the standard deviations of the respective distributions and, thus, directly proportional to the relative difference between the degree of precision.

Turning to the SOEP data, the correlation between the absolute and relative measure across the seven questions ranges from $\rho^{Spearman} = .91$ to $\rho^{Spearman} = .96$ which is consistent with the respondents interpreting the subjective error question in the same way.[43] Given the high correlation between both approaches, using the absolute measure is preferable as it avoids having to drop the observations of respondents whose subjective error is zero.

Age-robust measure (op_i''''): The negative correlation between age and overprecision in our sample is likely to be driven by the type of questions that were asked in the survey. Since we asked about specific historical events within the last 100 years, respondents who lived during these events might be better calibrated. This becomes obvious in Figure C.2 where, for every question, we split the density of our overprecision measure $op_{i,j}$ between those respondents born before and after the event. As expected, those subjects born before the event are better calibrated than those born after. As a robustness test, we construct, for every respondent, an *age-robust* measure of overconfidence (op_i''''). We construct this measure following the formulation described in Section 2.1, but using only those questions about events that happened *after* the respondent was born. The drawback of this approach is that we lose a substantial amount of information and give more weight to events that occurred later in time. Taking fewer questions into account also comes at the risk that the aggregate measure is biased by one specific question.

Residual measure (op_i'''''): The *residual* measure is a measure of overprecision obtained by the estimation method of Ortoleva and Snowberg (2015b). Ortoleva and Snowberg (2015b) construct their measure of overconfidence by asking respondents about their assessment of the current and one year-ahead inflation rate and the unemployment rate as well as their confidence about the respective answers using a six-point scale. They then regress participants' confidence on a fourth-order polynomial of accuracy to isolate the effect of knowledge. The principal component of the four residuals is then used as their measure of overconfidence. To replicate their approach as closely as possible, we construct a measure of respondent confidence by inverting the reported subjective error and computing quintiles. We then regress the respondents' "confidence" about the answer on

[43] Note that the relationship between the absolute and the relative measure is non-linear. Therefore, we report the Spearman correlation coefficient only.

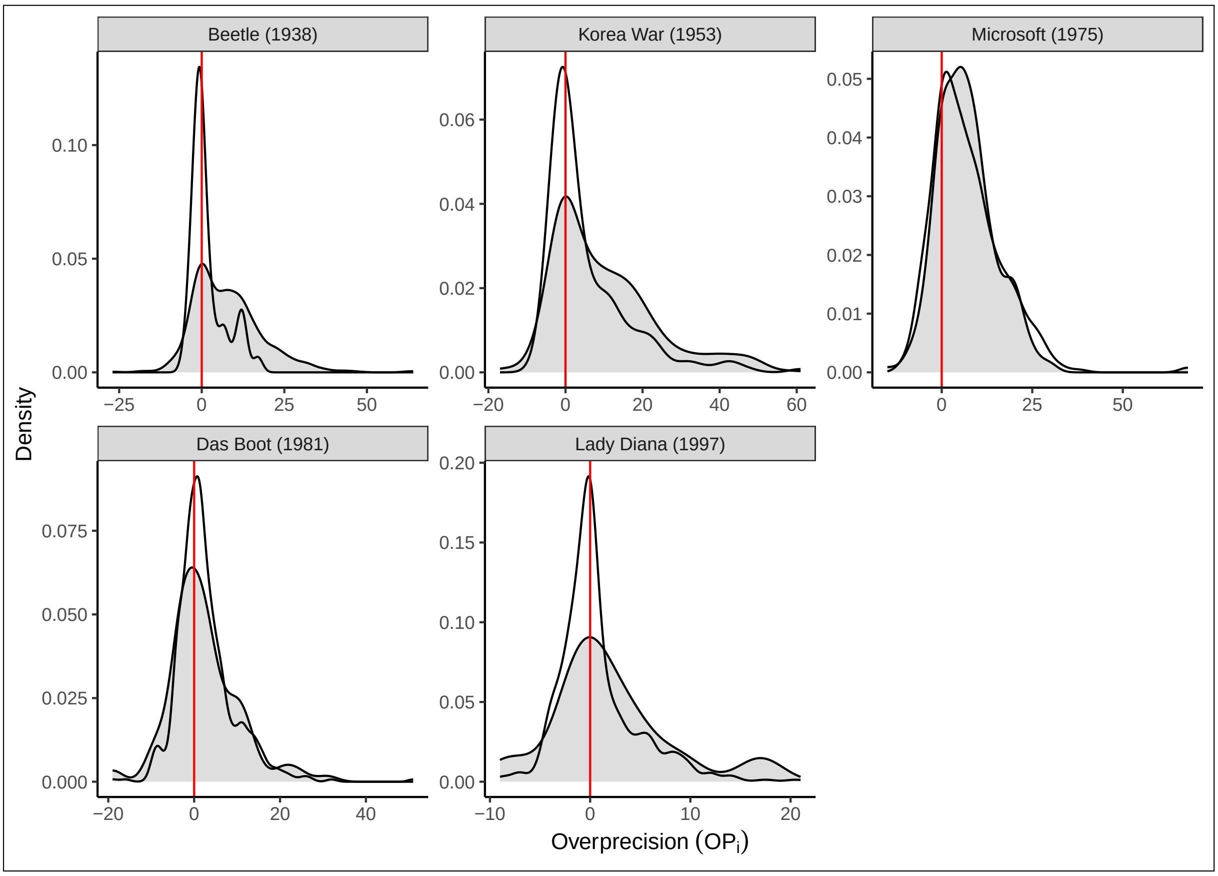
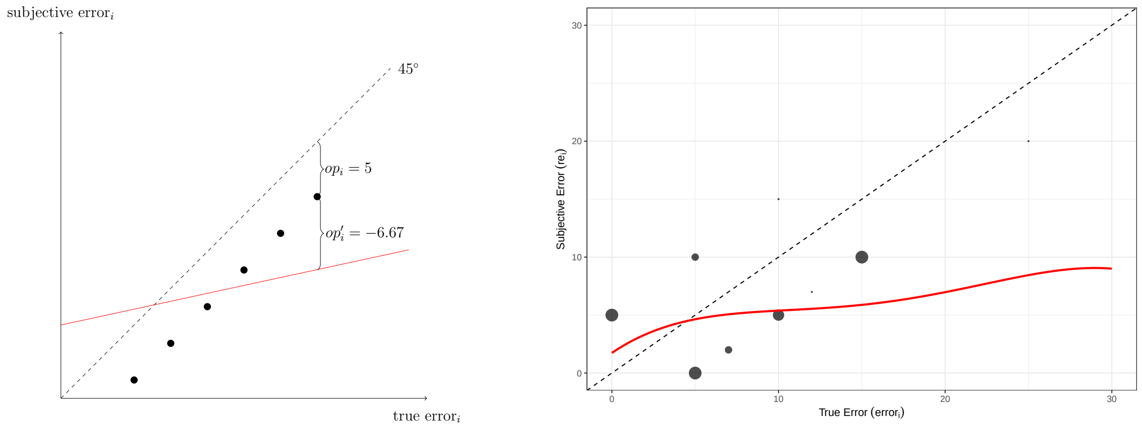


Figure C.2: Density of Overprecision (op_{ij}) and Age. From left (less recent) to right (more recent) We plot the density of the measured overprecision ($op_{i,j}$) for each question j. In gray, we plot the density of the measured overprecision for the question ($op_{i,j}$) of those subjects that were born after the event took place. With no color, we plot the density of all respondents born at the year of the event or before. Note that the scale of the vertical axis is different across the five plots. Questions with (correct) answers after 2000 are omitted as there were no underage respondents.

a fourth-order polynomial of the realized error and take the principal component of the residuals across all seven questions to create our new individual measure of overprecision $op_i^{'''''}$.

The residual measure of overprecision ($op_i^{'''''}$) mechanically differs from our baseline measure (op_i) because it effectively calculates the distance between the subjective error and the fitted fourth-order polynomial instead of the distance between the subjective error and the realized error. This approach comes with the caveat that, if a respondents deviation is small relative to that of the population, then, when computing the residuals for the seven questions, the measure classifies the respondent as underconfident even if their realized error is larger than their subjective error (for an illustrative example see Figure C.3). Thus, for every measure of $op_i^{'''''}$, the residual measure takes into account the relationship between the subjective error and the realized error for the entire *population*



(a) Illustration

(b) Data

Figure C.3: Misspecification of participants. This figure shows the difference between the Subjective Error Method and the *residual* approach for a theoretical illustration in (a) and for the answers to one of the overprecision questions in (b). Any observation in both panels above the 45° line represents underprecise individuals and any observation below represents overprecise individuals. Note that the axes are changed as compared to Figure 3. In panel (a), the dots represent observations for respondents for whom, in the example, the Subjective Error Method yields $op_i = error_i - se_i = 5$ in a specific question in the set of questions. The red line illustrates the fitted line of a simplified version of the *residual* approach using only a first order polynomial ($se_i = \alpha + \beta error_i + \epsilon_i$). In panel (b), the dots represent respondents for whom the Subjective Error Method yields an overprecision of 5 and -5 respectively. The red line indicates the fitted line of the *residual* approach using a fourth-order polynomial.

of respondents. In contrast, our approach focuses on the respondent's signal processing only by comparing the realized error with the subjective error.

C.2 Robustness of Descriptive Results

In Table C.1 we replicate Table 1 using each of the measures described in Section C.1 (Columns (2) to (6)) and our baseline measure sop_i in Column (1).

Column (2) of Table C.1 shows the results for the *standardized* measure (op'_i). The results show no qualitative changes with respect to the baseline except for the coefficient of the number of questions that were answered. However, the results are less significant. Column (3) shows the results using the *centered* measure (op''_i). The results remain largely robust with the coefficient for *gender* becoming larger and the coefficient for *answered* turning negative and insignificant.

Column (4) replicates the baseline estimations using the *relative* approach (op_i'''). The qualitative results remain similar except for less precisely estimated coefficients which can be explained by the decreased sample size. Column (5) uses the *age-robust* measure (op_i''''). The results show that, if we exclude the mechanical effect of age, then overprecision and age are positively correlated which is consistent with the earlier results from the literature (e.g., Ortoleva and Snowberg, 2015a,b; Prims and Moore, 2017). Otherwise, all of our results remain robust.

Column (6) of Table C.1 shows the results for the *residual* approach (op_i'''''). For the most part, the outcome replicates the results of Ortoleva and Snowberg (2015b), with females being less overprecise and income and education not showing up as statically relevant. Moreover, age is positively correlated with the estimated overprecision. Surprisingly, the number of answered questions has a negative effect on overprecision. In other words, contrary to the observed measure of overprecision, if we estimate overprecision using the methodology of Ortoleva and Snowberg (2015b), then the more questions a respondent answers, the less overprecise she is.

Given the results in Table C.1, we believe that our baseline measure is the best alternative. It is a simple and straightforward approach that can easily be implemented and which does not require the specification of an econometric model such as the approach of Ortoleva and Snowberg (2015b). It does not miss-classify respondents and uses all of the available information into account. Moreover, it is highly correlated to both the standardized measure ($\rho^{Pearson} = .85$; $\rho^{Spearman} = .86$; $N = 805$), the relative measure ($\rho^{Pearson} = .68$; $\rho^{Spearman} = .82$; $N = 801$), as well as the centered measure ($\rho^{Pearson} = .96$; $\rho^{Spearman} = .93$; $N = 801$) and therefore robust to transformations. All of these results are confirmed in Appendix C.3 where we test the predictive power of all robustness measures.

	Baseline	Standardized	Centered	Relative	Age robust	Residual
	(1)	(2)	(3)	(4)	(5)	(6)
age	-0.005	-0.000	-0.005*	0.003	0.018***	0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
female	0.111	0.087	0.177**	0.079	-0.013	-0.199***
	(0.075)	(0.076)	(0.075)	(0.083)	(0.074)	(0.074)
years education	-0.037***	-0.014	-0.037***	-0.003	-0.008	0.002
	(0.014)	(0.014)	(0.014)	(0.016)	(0.014)	(0.014)
answered	0.085***	-0.031	-0.010	0.033	0.075***	-0.104***
	(0.021)	(0.022)	(0.021)	(0.027)	(0.021)	(0.021)
gross income	-0.039*	-0.040**	-0.033	-0.019	-0.032	0.010
	(0.023)	(0.024)	(0.023)	(0.025)	(0.023)	(0.023)
fin. literacy	-0.398**	-0.306*	-0.364**	-0.469*	-0.347**	-0.149
	(0.155)	(0.158)	(0.155)	(0.175)	(0.154)	(0.153)
risk aversion	0.048	0.037	0.036	-0.016	0.017	0.007
	(0.037)	(0.038)	(0.037)	(0.043)	(0.037)	(0.037)
impulsivity	-0.014	-0.005	-0.018	-0.006	0.020	0.034
	(0.037)	(0.038)	(0.037)	(0.041)	(0.037)	(0.037)
patience	0.038	0.042	0.038	0.058*	0.033	0.044
	(0.035)	(0.036)	(0.035)	(0.039)	(0.035)	(0.035)
narcissism	0.100**	0.079*	0.098**	0.118**	-0.003	0.042
	(0.037)	(0.038)	(0.037)	(0.041)	(0.037)	(0.037)
N	805	805	805	702	800	805
adj. R^2	0.098	0.066	0.100	0.045	0.121	0.117
Constant Term	Yes	Yes	Yes	Yes	Yes	Yes
Employment & GDR 1989	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.1: Determinants of overprecision using alternative measures of overprecision. For comparison, in Column (1) we run an OLS with the baseline measure. In Column (2) - (6), we run an OLS using the *standardized* measure, the *centered* measure, the *relative* measure, the *age-robust* measure, and the *residual* measure respectively. All include dummies for the labor force status (employed, unemployed, retired, maternity leave, non-working), whether the respondent was a GDR citizen before 1989, and further personal characteristics. We also control for the federal state (*Bundesland*) where the respondent lives and the time at which he/she responded to the questionnaire.

C.3 Predictions Using Alternative Overprecision Measures

In the following, we will show the results for the *residual* approach following Ortoleva and Snowberg (2015b), the *relative* measure, the *standardized*, the *age-robust* measure, and the *centered* measure of overprecision. Table C.2 shows the results from the predictions using the *standardized* measure of overprecision instead. The results only slightly change with respect to the baseline, with the coefficients for the prediction errors becoming insignificant. However, the sign of the coefficient remains unchanged. The predictive power with respect to the LASSO estimations remains strong despite a slight decrease in the ranking as calculated by the R^2 method.

Table C.3 shows the results from the predictions using the *centered* measure of overprecision. Since the correlation between the centered and the baseline measure is .96, the results remain mostly unchanged.

Table C.4 shows the results from the predictions using the *relative* measure of overprecision instead. The advantage is that it makes the measure more comparable across subjects. However, we lose those observations with a reported zero subjective error due to mathematical reasons. The results, as compared to those in the baseline in Table 2, remain qualitatively similar.

Table C.5 shows the results from the predictions using the *age-robust* measure of overprecision instead. The results are at large in line with the results of the baseline estimations. The *age-robust* overprecision measures still predicts the outcomes according to the LASSO estimations. The point estimates slightly decrease in size and significance. However, as pointed out above, this measure considers fewer answers of the respondents and puts more weight on the more recent events since it only considers the questions on events after the respondent was born. Thus, the aggregate measure is calculated across fewer answers which might bias the measure. Therefore, these results have to be taken with a grain of salt.

Table C.6 shows the results from the predictions using the residual approach. Compared to the baseline measure, the alternative measure does not significantly predict any of the predictions derived from the theory. This is most likely because, applied to our data, this approach misclassifies certain respondents in the data as discussed in Appendix C.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Point	Unadj.	SH	R^2	LASSO		
	estimate	p-value	p-value	rank	included	R^2	N
Financial Behavior:							
DAX forecast error	0.11***	0.010	0.039	3/41	yes/15	0.10	548
1-year ahead	*0.382*	*0.450*		*19/41*	*no/15*	*0.14*	*578*
2-year ahead	*4.776****	*0.000*		*1/41*	*no/0*	*0.00*	*557*
portfolio diversification	-0.104***	0.004	0.019	3/41	yes/18	0.13	774
Political Behavior:							
extremeness	0.09**	0.025	0.072	3/42	yes/13	0.06	716
left-right	-0.017	0.662	0.662	18/42	no/14	0.08	716
non-voter	0.026**	0.025	0.050	4/42	yes/19	0.13	706

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.2: This table shows the estimation results of Section 3.4 using the *standardized* overprecision measure. The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 805. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)). Column (4) displays the result from the R^2 procedure specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Point	Unadj.	SH	R^2	LASSO		
	estimate	p-value	p-value	rank	included	R^2	N
Financial Behavior:							
DAX forecast error	0.08*	0.066	0.127	4/41	yes/14	0.10	548
1-year ahead	*0.387*	*0.456*		*12/41*	*yes/21*	*0.16*	*578*
2-year ahead	*3.202***	*0.012*		*4/41*	*no/1*	*0.00*	*557*
portfolio diversification	-0.125***	0.001	0.003	3/41	yes/19	0.13	774
Political Behavior:							
extremeness	0.081*	0.059	0.166	6/42	yes/14	0.05	716
left-right	0.003	0.944	0.944	18/42	no/14	0.08	716
non-voter	0.026**	0.034	0.128	4/42	yes/18	0.13	706

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.3: This table shows the estimation results of Section 3.4 using the *centered* overprecision measure. The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 805. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)). Column (4) displays the result from the R^2 procedure specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Point	Unadj.	SH	R^2	LASSO		
	estimate	p-value	p-value	rank	included	R^2	N
Financial Behavior:							
DAX forecast error	0.062	0.166	0.421	3/41	yes/16	0.11	501
1-year ahead	*0.19*	*0.721*		*9/41*	*no/20*	*0.16*	*530*
2-year ahead	*2.822***	*0.033*		*2/41*	*no/1*	*0.01*	*510*
portfolio diversification	-0.068*	0.080	0.284	18/41	yes/14	0.11	681
Political Behavior:							
extremeness	0.113***	0.007	0.033	2/42	yes/15	0.06	624
left-right	-0.03	0.465	0.713	18/42	no/15	0.08	624
non-voter	0.003	0.800	0.800	3/42	no/4	0.09	616

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: This table shows the estimation results of Section 3.4 using the *relative* overprecision measure. The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 805. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)) which is slightly less conservative than the Bonferroni adjustment. Column (4) displays the result from the R^2 procedure specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Point	Unadj.	SH	R^2	LASSO		
	estimate	p-value	p-value	rank	included	R^2	N
Financial Behavior:							
DAX forecast error	0.036	0.399	0.399	8/41	no/8	0.07	546
1-year ahead	*-0.442*	*0.379*		*4/41*	*no/16*	*0.15*	*576*
2-year ahead	*2.868***	*0.021*		*5/41*	*no/0*	*0.00*	*555*
portfolio diversification	-0.048	0.200	0.359	6/41	yes/16	0.12	769
Political Behavior:							
extremeness	0.065	0.102	0.350	5/42	yes/13	0.05	712
left-right	-0.061	0.119	0.315	6/42	no/11	0.07	712
non-voter	0.026**	0.022	0.106	3/42	yes/19	0.14	701

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: This table shows the estimation results of Section 3.4 using the *age-robust* overprecision measure. The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 805. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)). Column (4) displays the result from the R^2 procedure specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Point	Unadj.	SH	R^2	LASSO		
	estimate	p-value	p-value	rank	included	R^2	N
Financial Behavior:							
DAX forecast error	-0.021	0.626	0.948	14/41	no/12	0.09	548
1-year ahead	*-0.269*	*0.606*		*9/41*	*no/20*	*0.16*	*578*
2-year ahead	*-0.267*	*0.835*		*12/41*	*no/7*	*0.04*	*557*
portfolio diversification	-0.029	0.440	0.902	13/41	no/19	0.12	774
Political Behavior:							
extremeness	-0.046	0.256	0.773	8/42	yes/13	0.05	716
left-right	-0.008	0.831	0.831	17/42	no/14	0.08	716
non-voter	-0.003	0.798	0.959	18/42	no/17	0.13	706

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: This table shows the estimation results of Section 3.4 using the *residual* aggregation method of Ortoleva and Snowberg (2015a). The number of observations (Column (7)) varies due to missing observations in the outcome variable. The maximum number of observations is 805. Column (1) lists the point estimate of the standardized overprecision measure *sop* from the full regression as specified in Section 3.4.1 along with the unadjusted p-value (Column (2)) and the Sidak-Holm adjusted p-value for multiple hypothesis testing (Column (3)). Column (4) displays the result from the R^2 procedure specified in Section 3.4.1 along with the maximum possible variables to be included in the model. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest. Column (5) specifies the result of the LASSO procedure as specified in Section 3.4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

D Comparing the Survey Data to the SOEP Data

In the following, we compare the Subjective Error Method answers to the contemporary history questions for the online survey with those in the SOEP survey. To avoid any distortion from outliers, all variables are trimmed at the 1 and 99 percentile in both datasets.

Figure D.1 plots the distributions of answers to the history questions in the SOEP and the online survey. With the exception of the Beetle question, the distributions are relatively similar across samples. Table D.1 shows the results of a standard t-test on the means and of a test on the equality of the standard deviations of both samples. Except for the mean of the question regarding the VW Beetle (1938), the means of the answers to the questions differ by little, even if this difference is statistically significant in most cases. Interestingly, in most cases, the SOEP sample is closer to the correct answer. This could be explained by more effort of the respondents in the face-to-face interviews than in the online survey and speaks in favor of the 'Google' filters we introduced to filter our online sample. The dispersion of the answers around the mean is relatively similar for both samples except for the question regarding Hussein (2003) where SOEP respondents are more closely around the mean. However, the analysis suggests that knowledge about contemporary history is relatively similarly distributed among both samples.

Figure D.2 plots the distributions of the overprecision measure for the respective history questions contained in both the SOEP survey and the companion online survey. The distributions are graphically relatively similar in both samples. The statistical tests in Table D.2, however, show that the respondents in the SOEP sample, on average, tend to be slightly more overprecise for most of the questions. One explanation for this result could be that respondents try to show off in front of the interviewers in the face-to-face interviews by stating lower subjective errors. Taken together, the distributions of the overprecision measures only marginally differ among both samples, pointing to the robustness of the *Subjective Error Method.*

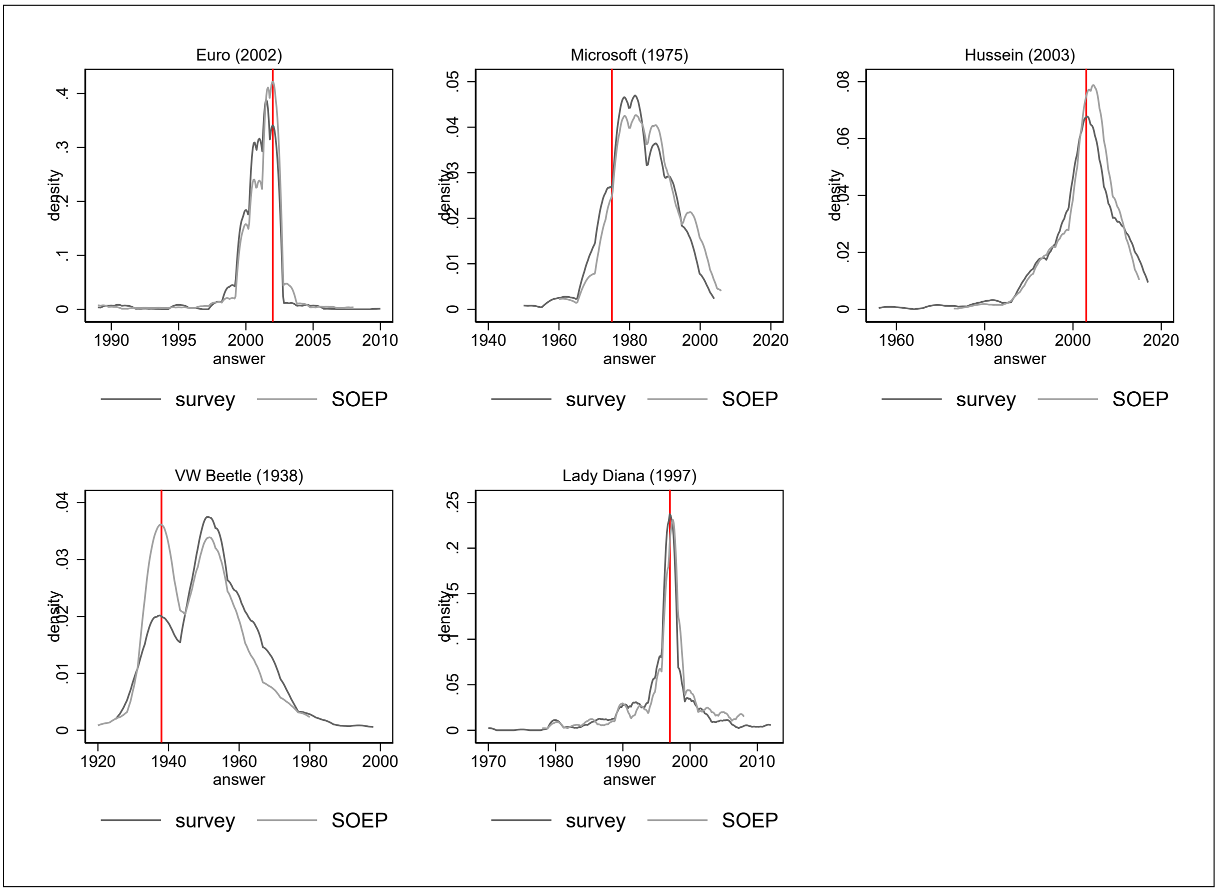


Figure D.1: Density of the answers ($a_{i,j}$) for each of the history questions contained in both the SOEP survey and the companion online survey. The vertical line marks the correct answer. Note that the vertical axis is different for each question.

question	mean (survey)	mean (SOEP)	Δ	p-val	sd (survey)	sd (SOEP)	Δ	p-val
Euro (2002)	2000,835	2001,124	-0,289	0,004	1,980	2,055	-0,075	0,294
Microsoft (1975)	1982,691	1984,929	-2,237	0,000	8,974	9,108	-0,134	0,697
Hussein (2003)	2002,212	2002,955	-0,742	0,077	8,718	6,568	2,150	0,000
VW Beetle (1938)	1952,268	1948,535	3,734	0,000	12,561	11,502	1,059	0,016
Lady Diana (1997)	1995,975	1996,680	-0,704	0,010	5,367	5,141	0,226	0,240

Table D.1: Statistical comparison between the online and SOEP survey. This table shows the statistical comparison of the answers to the history questions between the survey and the SOEP sample.

question	mean (survey)	mean (SOEP)	Δ	p-val	sd (survey)	sd (SOEP)	Δ	p-val
Euro (2002)	0,143	0,106	0,037	0,728	2,474	1,562	0,913	0,000
Microsoft (1975)	2,953	6,662	-3,708	0,000	8,687	7,979	0,708	0,031
Hussein (2003)	0,293	1,562	-1,269	0,000	6,707	4,554	2,153	0,000
VW Beetle (1938)	8,222	7,238	0,984	0,076	11,352	9,290	2,062	0,000
Lady Diana (1997)	0,096	0,663	-0,567	0,008	4,412	3,622	0,790	0,000

Table D.2: Statistical comparison of the overprecision measures for the online and SOEP survey. This table shows the statistical comparison of the overprecision measures by question between the survey and the SOEP sample.

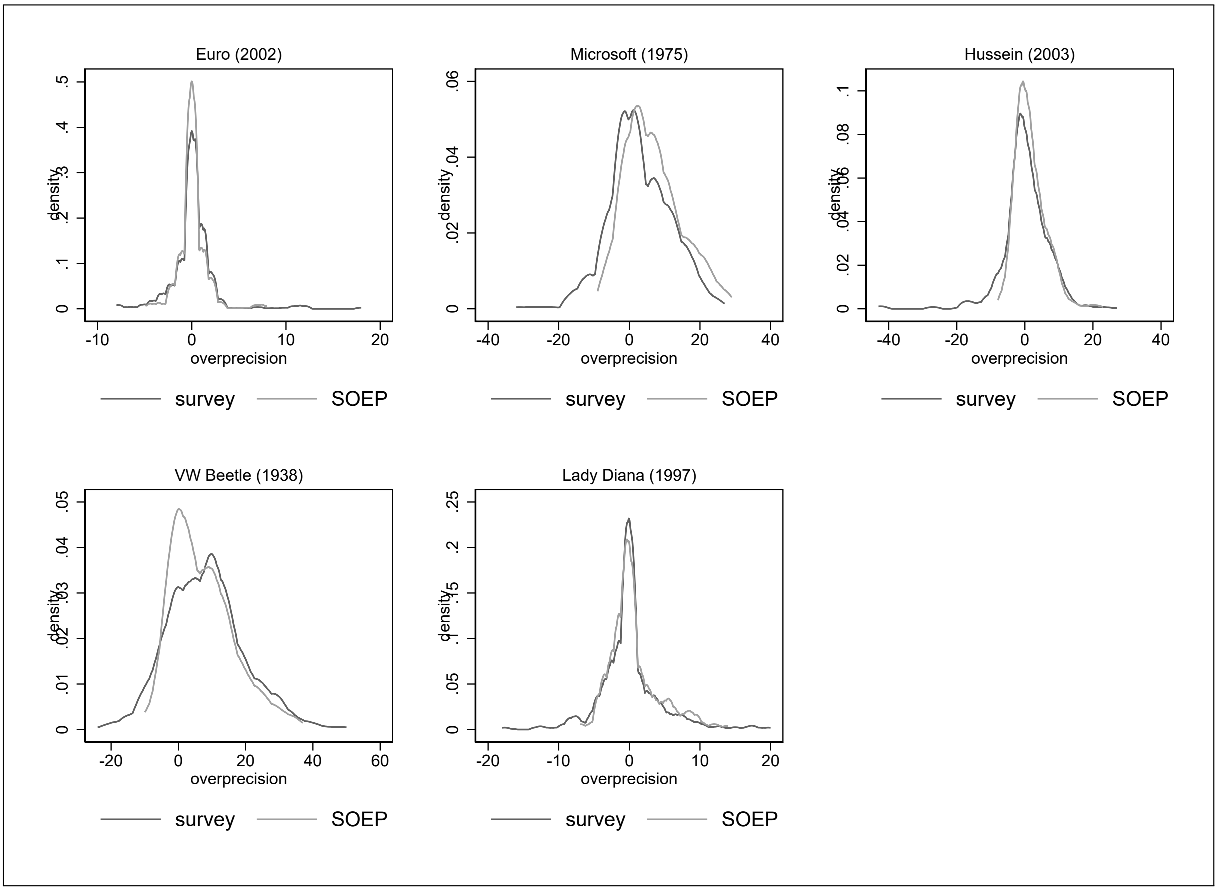


Figure D.2: Density of the overprecision measure ($op_{i,j}$) for each of the history questions contained in both the SOEP survey and the companion online survey. Note that the vertical axis is different for each question.

E Robustness Tests Companion Online Survey

In the following, we test the robustness of the analyses of the online experiment as outlined in the pre-registration. All robustness tests are concerned with the exclusion restrictions specified in the pre-analysis plan. In the baseline analysis, we exclude all respondents who admit to using a third party to answer our questions and respondents who we identify as 'Googlers' by using our control questions. We do so by excluding from the entire analysis those respondents who answer the Google controls correctly and state a subjective error of zero while displaying the same behavior for at least three other questions within the same domain. Additionally, we exclude the lowest fifth percentile on the self-reported effort measure to ensure data quality.

In total we have seven different robustness tests. Robustness tests (1) and (2) change the exclusion restriction to detect 'Googlers'. Robustness test (1) drops all respondents that have one of the Google control questions correct while stating a subjective error of zero, as opposed to the original restriction where we required respondents to answer all Google control questions correctly. Robustness test (2) drops all respondents who get at least four of the five answers in any domain correct. Robustness test (3) also excludes all respondents who are in the lowest decile of time spent on each survey question at least 20% of the time across the entire survey, while robustness test (4) repeats the same for the highest decile. Robustness test (5) increases the threshold of self-reported quality to 10% and robustness test (6) to 20%. Robustness test (7) additionally excludes those respondents who gave the same answer to more than two of the respective first questions within a domain to control for response patterns.

Table E.1 shows the robustness test results for the factor analysis, Table E.2 for the partial correlation analysis, and Table E.3 for the leave-one-out analysis. The results show that by adding more stringent exclusion restrictions concerning 'Googlers' does not substantially change the results. Adding further restrictions, which improve the quality of answers but reduced the number of observations, only marginally changes the results, predominantly improving the results.

		Robustness						
	Baseline	1	2	3	4	5	6	7
Eigenvalue	2.29	2.30	2.28	2.34	2.23	2.27	2.32	2.25
Loading neutral	0.52	0.52	0.52	0.53	0.60	0.51	0.49	0.48
Loading history	0.76	0.76	0.76	0.76	0.72	0.76	0.78	0.77
Loading knowledge	0.74	0.75	0.74	0.75	0.72	0.73	0.75	0.74
Loading stocks	0.69	0.69	0.68	0.71	0.68	0.68	0.67	0.68
Loading economics	0.65	0.65	0.65	0.65	0.61	0.66	0.67	0.65
N	552	544	547	487	440	535	476	500

Table E.1: Robustness of factor analysis. Column (Baseline) shows the baseline results with the exclusion restrictions specified in Section 4.2. Each of the robustness tests changes one of the exclusion restrictions. Robustness tests (1) and (2) change the exclusion restriction to detect 'Googlers'. Robustness test (1) drops all respondents that have one of the hard questions, which we use to detect respondents who we assume to have used search engines, correct while stating a subjective error of zero. Robustness test (2) drops all respondents who get at least four of the five answers in any domain correct. Robustness test (3) additionally excludes all respondents who are in the lowest decile of time spent on each survey question at least 20% of the time across the entire survey, while robustness test (4) repeats the same for the highest decile. Robustness test (5) increases the threshold of self-reported quality to 10% and robustness test (6) to 20%. Robustness test (7) additionally excludes those respondents who gave the same answer to more than two of the respective first questions within a domain to control for response patterns.

				Robustness											
		Baseline		1		2		3		4		5		6		7	
Domain 1	Domain 1	ρ	N	ρ	N	ρ	N	ρ	N	ρ	N	ρ	N	ρ	N	ρ	N
Neutral	History	0.17	688	0.17	680	0.17	679	0.18	588	0.19	560	0.16	658	0.18	578	0.16	632
Neutral	Knowledge	0.21	626	0.22	618	0.20	617	0.23	539	0.22	503	0.19	603	0.18	538	0.18	570
Neutral	Stocks	0.30	676	0.30	668	0.29	670	0.32	577	0.33	546	0.28	648	0.24	575	0.25	619
Neutral	Economics	0.18	639	0.18	632	0.18	633	0.17	552	0.20	518	0.17	613	0.19	542	0.17	580
History	Knowledge	0.52	609	0.52	600	0.52	600	0.52	531	0.49	488	0.52	586	0.55	522	0.53	554
History	Stocks	0.35	638	0.35	629	0.35	632	0.37	557	0.29	514	0.35	614	0.37	544	0.35	584
History	Economics	0.37	620	0.37	612	0.37	614	0.37	542	0.31	499	0.37	596	0.38	526	0.37	565
Knowledge	Stocks	0.34	591	0.34	582	0.33	585	0.37	516	0.32	471	0.33	573	0.34	509	0.32	537
Knowledge	Economics	0.32	580	0.33	572	0.32	574	0.33	507	0.28	461	0.32	562	0.35	502	0.30	525
Stocks	Economics	0.28	618	0.28	610	0.27	613	0.30	540	0.24	500	0.29	594	0.29	529	0.28	561

Table E.2: Robustness of partial correlation analysis. Column (Baseline) shows the baseline results with the exclusion restrictions specified in Section 4.2. Each of the robustness tests changes one of the exclusion restrictions. Robustness tests (1) and (2) change the exclusion restriction to detect 'Googlers'. Robustness test (1) drops all respondents that have one of the hard questions, which we use to detect respondents who we assume to have used search engines, correct while stating a subjective error of zero. Robustness test (2) drops all respondents who get at least four of the five answers in any domain correct. Robustness test (3) additionally excludes all respondents who are in the lowest decile of time spent on each survey question at least 20% of the time across the entire survey, while robustness test (4) repeats the same for the highest decile. Robustness test (5) increases the threshold of self-reported quality to 10% and robustness test (6) to 20%. Robustness test (7) additionally excludes those respondents who gave the same answer to more than two of the respective first questions within a domain to control for response patterns.

				Robustness											
		Baseline		1		2		3		4		5		6		7	
In	Out	Δ	p	Δ	p	Δ	p	Δ	p	Δ	p	Δ	p	Δ	p	Δ	p
Neutral	History	-0.348	< 0.01	-0.335	< 0.01	-0.363	< 0.01	-0.394	< 0.01	-0.335	< 0.01	-0.346	< 0.01	-0.352	< 0.01	-0.335	< 0.01
Neutral	Knowledge	-0.406	< 0.01	-0.422	< 0.01	-0.373	< 0.01	-0.519	< 0.01	-0.368	< 0.01	-0.397	< 0.01	-0.409	< 0.01	-0.396	< 0.01
Neutral	Stocks	-0.339	< 0.01	-0.343	< 0.01	-0.341	< 0.01	-0.354	< 0.01	-0.338	< 0.01	-0.301	< 0.01	-0.273	< 0.01	-0.328	< 0.01
Neutral	Economics	-0.323	< 0.01	-0.322	< 0.01	-0.340	< 0.01	-0.326	< 0.01	-0.338	< 0.01	-0.311	< 0.01	-0.364	< 0.01	-0.311	< 0.01
History	Neutral	-0.218	0.01	-0.184	0.03	-0.221	0.01	-0.243	0.01	-0.288	< 0.01	-0.207	0.02	-0.273	< 0.01	-0.163	0.07
History	Knowledge	-0.566	< 0.01	-0.572	< 0.01	-0.547	< 0.01	-0.605	< 0.01	-0.547	< 0.01	-0.549	< 0.01	-0.534	< 0.01	-0.556	< 0.01
History	Stocks	-0.432	< 0.01	-0.436	< 0.01	-0.453	< 0.01	-0.439	< 0.01	-0.477	< 0.01	-0.421	< 0.01	-0.368	< 0.01	-0.387	< 0.01
History	Economics	-0.385	< 0.01	-0.393	< 0.01	-0.383	< 0.01	-0.366	< 0.01	-0.439	< 0.01	-0.370	< 0.01	-0.405	< 0.01	-0.390	< 0.01
Knowledge	Neutral	-0.230	0.01	-0.244	< 0.01	-0.215	0.01	-0.247	0.01	-0.186	0.05	-0.206	0.02	-0.150	0.10	-0.213	0.02
Knowledge	History	-0.451	< 0.01	-0.480	< 0.01	-0.463	< 0.01	-0.381	< 0.01	-0.454	< 0.01	-0.468	< 0.01	-0.419	< 0.01	-0.488	< 0.01
Knowledge	Stocks	-0.323	< 0.01	-0.336	< 0.01	-0.314	< 0.01	-0.349	< 0.01	-0.341	< 0.01	-0.315	< 0.01	-0.271	< 0.01	-0.325	< 0.01
Knowledge	Economics	-0.301	< 0.01	-0.307	< 0.01	-0.291	< 0.01	-0.301	< 0.01	-0.308	< 0.01	-0.281	< 0.01	-0.320	< 0.01	-0.279	< 0.01
Stocks	Neutral	-0.449	< 0.01	-0.418	< 0.01	-0.436	< 0.01	-0.451	< 0.01	-0.467	< 0.01	-0.421	< 0.01	-0.406	< 0.01	-0.407	< 0.01
Stocks	History	-0.421	< 0.01	-0.408	< 0.01	-0.421	< 0.01	-0.416	< 0.01	-0.467	< 0.01	-0.416	< 0.01	-0.411	< 0.01	-0.361	< 0.01
Stocks	Knowledge	-0.368	< 0.01	-0.408	< 0.01	-0.360	< 0.01	-0.432	< 0.01	-0.381	< 0.01	-0.364	< 0.01	-0.352	< 0.01	-0.325	< 0.01
Stocks	Economics	-0.371	< 0.01	-0.395	< 0.01	-0.369	< 0.01	-0.437	< 0.01	-0.394	< 0.01	-0.397	< 0.01	-0.459	< 0.01	-0.348	< 0.01
Economics	Neutral	-0.338	< 0.01	-0.313	< 0.01	-0.358	< 0.01	-0.344	< 0.01	-0.346	< 0.01	-0.323	< 0.01	-0.302	< 0.01	-0.319	< 0.01
Economics	History	-0.384	< 0.01	-0.367	< 0.01	-0.386	< 0.01	-0.336	< 0.01	-0.445	< 0.01	-0.376	< 0.01	-0.359	< 0.01	-0.341	< 0.01
Economics	Knowledge	-0.292	< 0.01	-0.306	< 0.01	-0.277	< 0.01	-0.273	0.01	-0.320	< 0.01	-0.281	< 0.01	-0.301	< 0.01	-0.262	< 0.01
Economics	Stocks	-0.381	< 0.01	-0.386	< 0.01	-0.330	< 0.01	-0.364	< 0.01	-0.360	< 0.01	-0.363	< 0.01	-0.359	< 0.01	-0.340	< 0.01

Table E.3: Robustness of leave-one-out-analysis. Column (Baseline) shows the baseline results with the exclusion restrictions specified in Section 4.2. Each of the robustness tests changes one of the exclusion restrictions. Robustness tests (1) and (2) change the exclusion restriction to detect 'Googlers'. Robustness test (1) drops all respondents that have one of the hard questions, which we use to detect respondents who we assume to have used search engines, correct while stating a subjective error of zero. Robustness test (2) drops all respondents who get at least four of the five answers in any domain correct. Robustness test (3) additionally excludes all respondents who are in the lowest decile of time spent on each survey question at least 20% of the time across the entire survey, while robustness test (4) repeats the same for the highest decile. Robustness test (5) increases the threshold of self-reported quality to 10% and robustness test (6) to 20%. Robustness test (7) additionally excludes those respondents who gave the same answer to more than two of the respective first questions within a domain to control for response patterns.

F Further Exploratory Analyses in the Companion Survey

Given the rich data we collect in the companion online survey, we conduct further exploratory analyses which have not been pre-registered in the pre-analysis plan.

F.1 Overprecision and Expertise

The literature on overconfidence and overprecision has shown that knowledge and expertise increase the accuracy of answers, but also the confidence of respondents. These two effects cancel out making experts and non-experts equally overconfident. An example of this effect is Önkal et al. (2003), who show that newcomers and experts have similar levels of overconfidence when predicting foreign exchange rates. While experts' predictions are more accurate, they also show more confidence in their answers, resulting in similar degrees of overconfidence. Similarly, McKenzie et al. (2008) show that the confidence intervals of experts are closer to the truth but also narrower than those of newcomers.

In the following, we test this relationship in our data using a self-reported measure of expertise. To do so, we regress the mean error, the mean subjective error, and the aggregate measure of overprecision in each domain on self-reported expertise in the same domain. Because self-reported expertise is likely to be influenced by overprecision, we control for the aggregate measure of overprecision. That is, for each domain, we predict the first component of a principal component analysis of the standardized aggregate overprecision measures across all domains excluding the one that is analyzed.

The results in Table F.1 show that self-reported expertise in a domain results in both a decrease in the mean realized error but also in a decrease of the mean subjective error. Importantly, the results in Columns (3) and (4) show that overprecision also affects self-reported expertise. Controlling for overprecision, the effects on the mean error and the mean subjective error are relatively similar. Therefore, consistent with the literature, we do not find an effect of expertise on the overprecision measure within the same domain.

	mean error		mean subjective error		mean overprecision	
	(1)	(2)	(3)	(4)	(5)	(6)
History:						
expertise	-0.036***	-0.039***	-0.044***	-0.041***	0.001	0.000
	(0.008)	(0.008)	(0.013)	(0.014)	(0.002)	(0.002)
overprecision		0.389*		-3.437***		0.581***
		(0.208)		(0.856)		(0.129)
N	703	552	703	552	703	552
adj. R^2	0.032	0.042	0.025	0.312	-0.001	0.291
Knowledge:						
expertise	-0.030*	-0.030*	-0.054***	-0.056***	0.002	0.002
	(0.015)	(0.017)	(0.017)	(0.017)	(0.002)	(0.002)
overprecision		1.287***		-4.277***		0.484***
		(0.414)		(0.596)		(0.055)
N	633	552	633	552	633	552
adj. R^2	0.005	0.028	0.023	0.370	0.001	0.272
Stocks:						
expertise	-0.026**	-0.028**	-0.058***	-0.046***	0.003**	0.002
	(0.011)	(0.012)	(0.012)	(0.011)	(0.001)	(0.001)
overprecision		0.983*		-3.655***		0.443***
		(0.555)		(0.735)		(0.075)
N	690	552	690	552	690	552
adj. R^2	0.005	0.020	0.024	0.214	0.004	0.219
Economics:						
expertise	-0.039***	-0.043***	-0.066***	-0.050***	0.002	0.001
	(0.011)	(0.011)	(0.015)	(0.014)	(0.002)	(0.001)
overprecision		0.429*		-3.839***		0.376***
		(0.250)		(0.747)		(0.067)
N	647	552	647	552	647	552
adj. R^2	0.019	0.025	0.030	0.230	0.002	0.188

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F.1: Self-reported expertise and the overprecision measure. This table reports the results of regressing the mean error, the mean subjective error, and the overprecision measure of one domain on the self-reported expertise in the same domain. In Columns (2), (4), and (6) we additionally control for overprecision, which is measured as the first component of a principal component analysis of the overprecision measures across all domains but the one that is analyzed.

F.2 Overprecision and Forecast Errors

This section replicates the results for the stock price forecast errors in Section 3.4.2. We calculate the absolute forecast error for each stock and regress the resulting forecast error on a measure of overprecision with and without control variables (Columns (1) to (10) in Table F.2). We also compute the principal component across all five stocks and regress this aggregate forecast error on a measure of overprecision and control variables in Column (11).

The first panel replicates the analysis in Section 3.4.2 using the standardized aggregate overprecision measure in the contemporary history domain. The results show that overprecision measured in the history domain is weakly positively correlated with the forecast error in two of the stocks but not significantly correlated with the forecast error in the other three stocks. The correlation with the principal component across all forecast errors is positive but insignificant. The results, hence, support the findings for the one-year and two-year ahead forecasts in Section 3.4.2. In the second panel, we use the standardized aggregate overprecision measure from the stock forecast domain instead of the history domain. The results for all stocks and the principal component show a positive and significant correlation between overprecision measured in the stock forecast domain and the forecast errors. However, it is likely that the measured relationship is mechanical.

Because of the potential mechanical relationship, in the third and fourth panel, we use the principal component of the standardized aggregate overprecision measures across all five domains.[44] This reduces the number of observations since it requires an overprecision measure for each domain per respondent, however, since the measure is based on more questions it becomes more reliable.[45] The results once more show a positive and significant correlation between overprecision and the absolute forecast error for most of the stocks and the principal component. In the fourth panel, we winsorize the forecast errors at 5th and 95th percentile to avoid outliers influencing the results. As expected, the coefficients are smaller in terms of size, but the qualitative result is unaffected.

[44] The principal component analysis only yields one factor with an eigenvalue of 2.29 and relatively similar factor loadings across domains (0.52,0.76,0.74,0.69,0.65).

[45] The principal component calculated across all questions, instead of across all domain-specific aggregate overprecision measures, exhibits a Spearman correlation coefficient of 0.96 with the principal component calculated across all domain-specific aggregate overprecision measures.

Forecast error in:	Benz		Puma		BMW		Post		BASF		PC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Overprecision in history domain:											
sop	0.924**	0.779*	0.303	0.167	0.066	-0.074	-0.431	-0.577	0.581*	0.441	0.018
	(0.466)	(0.446)	(0.432)	(0.428)	(0.498)	(0.501)	(0.557)	(0.568)	(0.297)	(0.297)	(0.041)
N	665	663	669	667	665	663	657	655	664	662	641
adj. R^2	0.008	0.036	-0.000	0.005	-0.001	0.031	0.000	0.028	0.005	0.001	0.028
controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Overprecision in stock forecast domain:											
sop	5.400***	5.141***	4.981***	4.831***	6.460***	6.025***	2.718***	2.467***	3.631***	3.463***	0.566***
	(1.000)	(0.963)	(0.831)	(0.845)	(1.503)	(1.462)	(0.794)	(0.819)	(0.660)	(0.666)	(0.106)
N	687	683	687	683	686	682	683	679	687	683	666
adj. R^2	0.272	0.276	0.231	0.225	0.224	0.225	0.063	0.071	0.198	0.175	0.314
controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Principal component of overprecision:											
sop	2.769***	2.638***	1.970**	1.862**	2.990**	2.740**	0.845	0.555	1.864***	1.736***	0.243**
	(0.903)	(0.885)	(0.806)	(0.806)	(1.241)	(1.217)	(0.781)	(0.788)	(0.577)	(0.570)	(0.098)
N	550	548	551	549	549	547	548	546	551	549	539
adj. R^2	0.083	0.094	0.043	0.043	0.052	0.070	0.005	0.021	0.057	0.041	0.075
controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Principal component of overprecision & winsorized outcomes:											
sop	1.011***	0.880***	1.328***	1.215**	0.521	0.421	0.969*	0.767	1.134***	1.016***	0.094**
	(0.323)	(0.313)	(0.510)	(0.505)	(0.332)	(0.331)	(0.516)	(0.513)	(0.321)	(0.313)	(0.043)
N	550	548	551	549	549	547	548	546	551	549	539
adj. R^2	0.029	0.045	0.032	0.042	0.005	0.015	0.016	0.040	0.040	0.025	0.036
controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Principal component of overprecision without finance domain:											
sop	1.515**	1.339**	0.705	0.508	1.405*	1.029	0.143	-0.188	0.971**	0.788*	0.086
	(0.636)	(0.618)	(0.648)	(0.649)	(0.842)	(0.835)	(0.693)	(0.709)	(0.441)	(0.438)	(0.071)
N	556	554	557	555	556	554	552	550	557	555	542
adj. R^2	0.024	0.039	0.004	0.008	0.010	0.033	-0.002	0.018	0.014	0.001	0.023
controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F.2: Forecast errors and the overprecision measure. This table reports the results of regressing the forecast errors for each of the stocks (Columns (2) to (10)) and the principal component across all stocks on overprecision.

In the fifth panel, we exclude the stock forecast domain from the aggregate overprecision measure. The relationship between the forecast errors and overprecision measured in all other domains but the stock forecast domain is weaker, however, the qualitative results remain similar.

G Calibration Analysis

Individuals might have different concepts of the second question of the Subjective Error Method in mind when answering the question. Consider the example in Figure G.1, where we plot a hypothetical distribution of the subjective error for two individuals. Both individuals have the same knowledge about the question, equally and correctly assess the distribution of the subjective error (dotted normal distribution), and make the same error (vertical red line). Since the SEM asks for the absolute error, we transform the dotted subjective error distribution to a half-normal distribution and its cumulative distribution (both continuous lines).[46] As can be seen from the graph, individual i, which has a 90% confidence interval in mind when answering our question, would give a larger number as subjective error than individual j, which has a 68% confidence interval in mind. Therefore, two persons with the same expected error distribution and the same realized error, would be classified as over or underprecise, depending on how they interpret the second answer to the SEM.

To test what percentile respondents have in mind when answering the second question of the SEM, we run an online companion survey using a representative sample of the German population.

G.1 Survey Design

As the first companion survey, this second survey asks respondents to estimate the year in which the five most answered questions in the SOEP survey happened as well as estimating their subjective error. The difference is that in this survey subjects not only reported a point prediction for the year they guessed the event happened, but we also elicited a full probability distribution of the year in which the event happened using the click-and-drag interface of Crosetto and De Haan (2022).[47] By comparing the point answer of the year in which the event happened to the probabilistic distribution reported by the subject, we can get a full distribution of the subjective errors, with errors larger or equal to +/-20

[46] Note that this assumes that the distribution is symmetric around zero.

[47] Each respondent is shown a graph with 20 years above and below their answer to the first question on the x-axis. By clicking and dragging points for each bin in the graph, the respondents can build a complete probability distribution which is normalized to 100% after submitting the response.

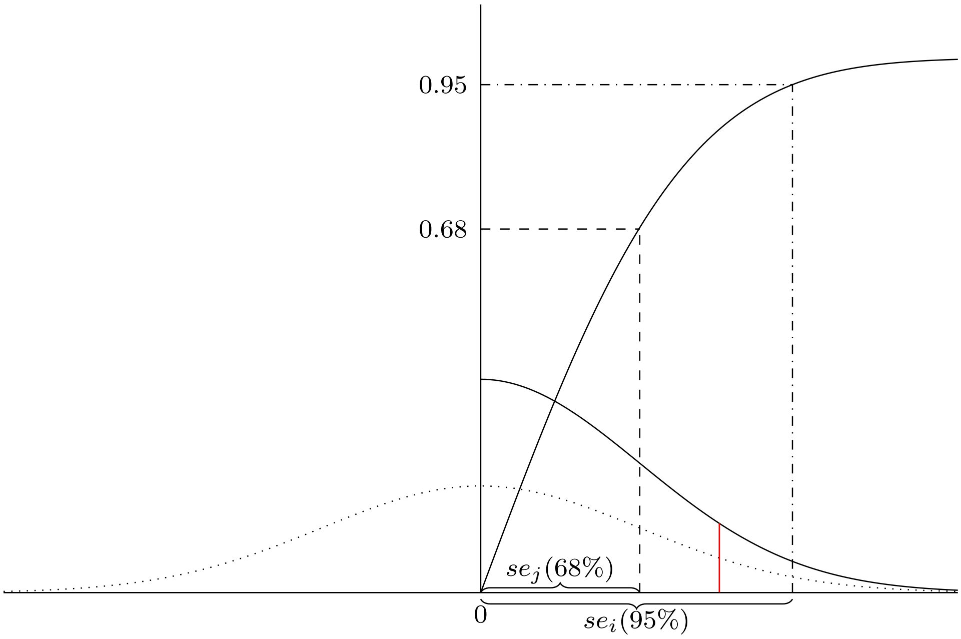


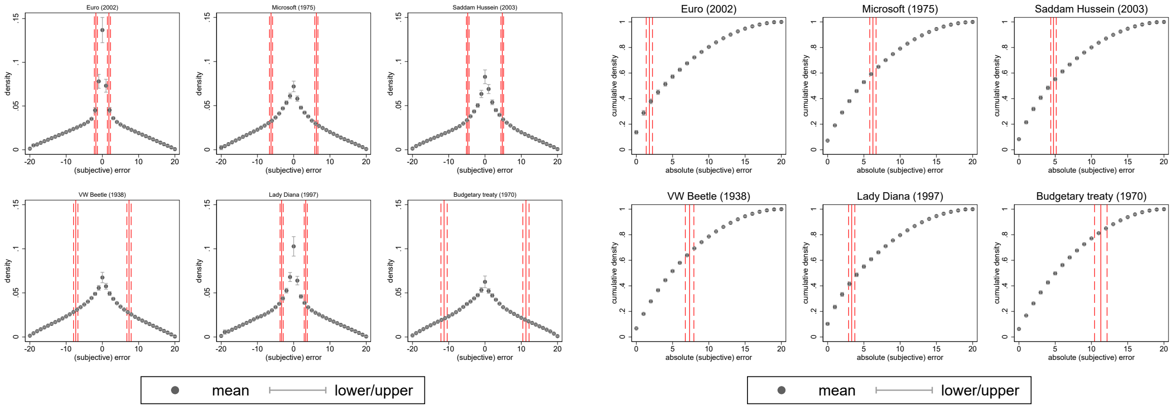
Figure G.1: Hypothetical distribution of the subjective error. The dotted normal distribution with a mean of zero and a standard deviation of two represents the subjective error for subjects i and j. The vertical

summarized in the outer bins.[48] Half of the respondents received the distribution question before reporting their estimated subjective error and the other half *vice versa*.

To ensure the quality of the data, we introduced attention checks during the survey. Any respondent that failed them was automatically excluded from the survey. To account for the use of Google or other search engines, at the end of the survey respondents are asked whether they used such methods to answer the questions. Additionally, we included an extra question that acts as "Google control." This question is presented in the same format as all other questions but is more difficult and unlikely to be known by respondents. In this case, we asked the year in which the first budgetary treaty of the EU was signed (1970).[49] Importantly, in this survey we pre-registered that we would use the answers to the Google control in the analysis of the data, as we feared being underpowered.

[48] For example, say a subject answers 1995 to the question about the year of the death of Lady Diana with a subjective error of 1 year. If she then builds a distribution that has 60% mass at 1995, 10% at 1994 and 1996 each, 5% at 1993 and 1997, and 2.5% at 1992 and 1998 then we have a full distribution of the subjective errors. By comparing the reported subjective error to the distribution, we can infer that when answering the subjective error question, she was reporting a 80% confidence interval. Notice that subjects were not required to build symmetric distributions.

[49] Following our pre-registration, we consider a respondent to have used a search engine if two conditions are met: i) answering the Google controls correctly and stating a subjective error of zero, and ii) answering correctly and stating zero subjective error for at least three other questions. If a respondent is flagged as having used search engines, then she is excluded from the analysis



(a) PDF

(b) CDF

Figure G.2: Subjective error and the probability distributions of the subjective error in the raw data.

At the end of the survey, all respondents answer a demographic survey consisting of age, gender, years of education, income, nationality, and mathematical literacy determined by solving three mathematical problems. We also ask respondents to self-report their knowledge of contemporary history on a scale of 0 (not knowledgeable at all) to 100 (very knowledgeable). Finally, we ask respondents to self-report the amount of effort they put into answering the survey.

G.2 Data

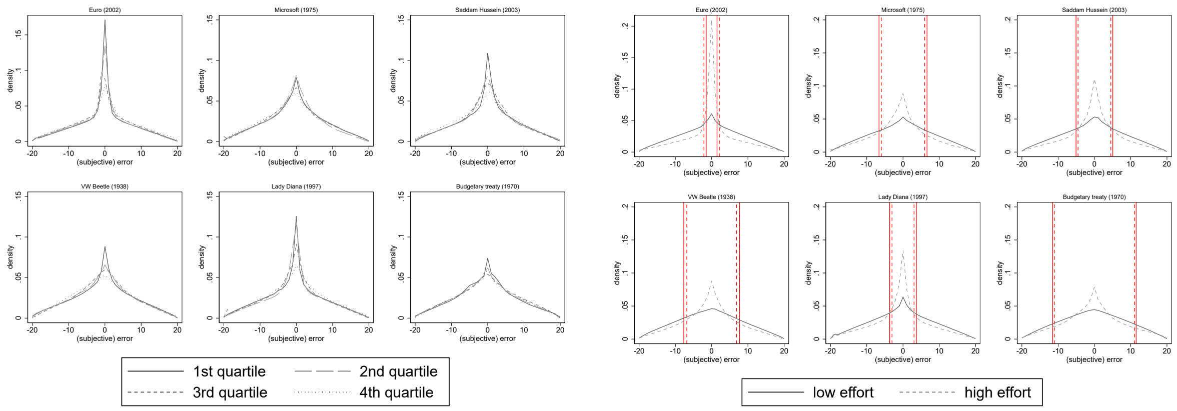
We collected 1.000 complete responses. As pre-registered, and to ensure that we only keep informative responses, we exclude all respondents who admit to using a third- party to answer our questions from the analysis. Additionally, we exclude respondents who we identify as 'Googlers' by using the control questions we describe in section 4.1, and drop the lowest five percentiles on the self-reported effort measure. This leaves us with 818 respondents for the baseline analysis. Of these, we keep the individuals that gave us at least four probability distributions, which reduces the number to 650. We further exclude 6 respondents that entered inconsistent probability distributions.

G.3 Results

Figure G.2a shows the probability density functions of the subjective error averaged across the sample for each of the questions. The gray dots depict the mean probability of each potential subjective error along with the 95% confidence interval of the mean (gray bars). The red solid vertical lines denote the average subjective error along with the 95% confidence interval of the mean (dashed red vertical lines). It is clear that knowledge affects the distributions provided by the respondents as well as the subjective errors. For the presumably easier questions (e.g. the introduction of the Euro in 2002), the distributions are narrower and the subjective errors are smaller. Figure G.2b plots the corresponding cumulative distribution functions of the *absolute* subjective error averaged across the sample along with the average subjective error. From the raw data, it seems that the location of the answer to the second question of the Subjective Error Method, i.e. the subjective error, within the subjective error distribution varies with the difficulty of the question. For example, while the average subjective error corresponds to approximately the 40th percentile of the subjective error distribution for the Euro question, it corresponds to the 80th percentile for the 'hard' question that is used to detect the use of search engines. Hence, for more difficult questions, where respondents on average are more uncertain about the correct answer and provide wider distributions, they also give relatively larger subjective errors.

However, the raw data above does neither control for the knowledge nor for the effort of respondents in answering the distribution questions. To address the first issue, Figure G.3a plots the average provided distributions for four quartiles of the realized error for each question, which is used as a proxy for knowledge. The results show that the lower the realized error, i.e. the better the knowledge of the respondent, the narrower the provided distributions.

The click-and-drag interface allows eliciting distributions in a user-friendly manner by clicking and dragging points along a distribution. The drawback of this interface is, that the first click produces a triangular distribution. Many individuals only clicked on the interface a few times suggesting a lower level of effort. To address this issue, Figure G.3b proxies this finding by splitting the sample according to the median time spent on building the respective distribution. For the low-effort group, i.e., the ones that spent less time answering the question, the distributions are close to triangular distributions which arise



(a) Knowledge

(b) Effort

Figure G.3: Subjective error distributions and knowledge/effort.

when only clicking once in the interface. For the high-effort group, the distributions are narrower. This is mechanical since respondents that are more certain about their answers require more clicks to produce narrower distributions. Interestingly, the subjective error does not substantially differ between the two groups.

To further support this finding, we split the respondents into two groups according to the shape of the provided distributions in Figure G.4. 'Low effort' includes all respondents that provide a perfectly triangular distribution, which requires only one click, while 'high effort' includes all other respondents. While for the low-effort group (dashed lines), the percentile to which the subjective error corresponds to varies between 15% and 85%, and it is relatively stable around 60% for the respondents that provide more effort (solid lines). Hence, on average, after controlling for effort, the location of the subjective error corresponds to the 60th percentile of the subjective error distribution. Given the two previous findings, it is important to control for both knowledge and effort.

To get a sense of how consistently the respondents answer the second question of the Subjective Error Method, we calculate the within-individual standard deviation of the percentiles of the subjective error distributions to which the answer to the second question corresponds to, i.e., the location of the subjective error, across all questions. The larger the standard deviation, the more inconsistently individuals answer. We only

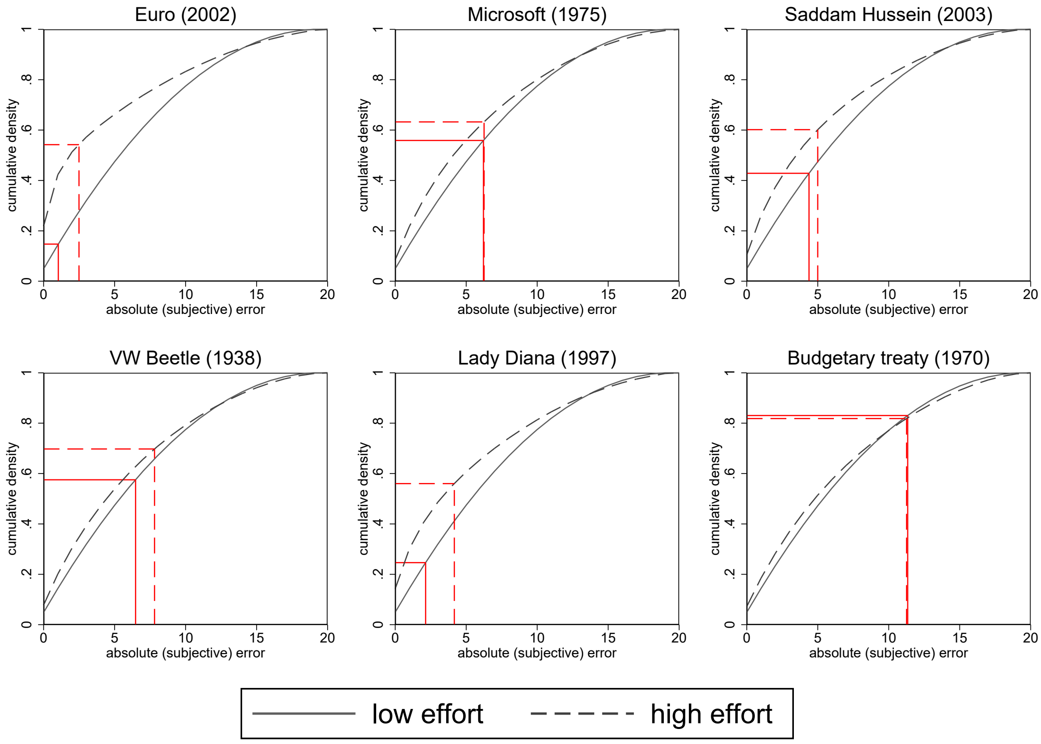
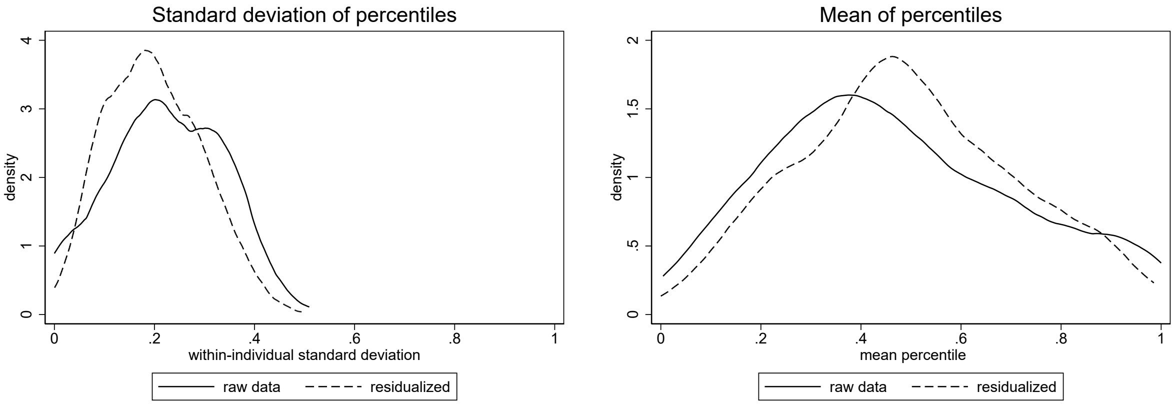


Figure G.4: Subjective error distributions and effort.

compute the within-individual standard deviation if individuals answered at least four of the distribution questions. In addition to the raw measure, we also construct residualized percentiles corrected for the difficulty of the answer and the amount of effort provided by the respondent.[50] Figure G.5a plots the density of both resulting variables. The figure shows that controlling for the individual effort per question as well as the proxy for knowledge decreases the within-individual dispersion of the location. The median standard deviation of the residualized within-individual standard deviation is 0.19 with relatively little dispersion around the median. Hence, individuals are relatively consistent in answering the second question after controlling for the question characteristics.

We then calculate the individual-specific mean of the (up to) six percentiles to obtain an aggregate measure for each respondent and to control for potential within-individual noise. The resulting aggregate measure allows us to examine what respondents on average

[50] For each question, we regress the observed percentile on the realized error, the time spent on the distribution, a dummy that equals one if the respondent provided a perfectly triangular distribution, and individual specific control variables. We then subtract the estimated marginal effects of knowledge and effort from the observed percentiles to control for their effects. This part of the analysis was not pre-registered.



(a) Standard deviation

(b) Mean location

Figure G.5: Densities of the within-individual standard deviation of the location of the subjective error and the individual-specific mean percentile.

have in mind when answering the second question in the Subjective Error Method. To control for potential question characteristics, we again also compute a residualized variable similar to above. Figure G.5b shows the resulting densities of both variables. Similar to before, controlling for question characteristics decreases the dispersion. The aggregated percentiles are centered around 0.48. About 50% of the respondents give a location of the subjective error that corresponds to a percentile between 33% and 64%.