The Effects of Overconfidence on the Political and Financial Behavior of a Representative Sample *

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

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

February 17, 2021

Abstract

We study how overconfidence affects the political and financial behavior of a nationally representative sample. To do so, we introduce a new method of eliciting overconfidence that is simple to understand, quick to implement, and captures the excess confidence of respondents on 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, more extreme political views, and a change in voting behavior. These results validate our method and show how overconfidence is a bias that permeates several aspects of peoples' life.

Keywords Overconfidence, SOEP, Survey

JEL Classification C83 · D91 · G41

*Corresponding author: cirilbosch@gmail.com. We would like to thank Thomas Gräber, Frank Heinemann, Guillem Riambau, Agustin Casas, Antoni-Italo de Moragas, and the participants of the CESifo Area Conference in Behavioral Economics for their insightful comments. We acknowledge the financial support from the Deutsche Forschungsgemeinschaft (DFG) through the CRC TRR 190 "Rationality and Competition." The order of the authors is not alphabetical as the first two authors contributed relatively 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 considered to be one of the most pervasive and potent biases in human judgment (Mannes and Moore, 2013; Kahneman, 2013). It leads to fighting wars (Johnson, 2009), to excessive entry into markets (Camerer and Lovallo, 1999), or (less critically) for 80% of the people to think that they are above median drivers (Svenson, 1981). However, overconfidence is a general term that encompasses three different phenomena which are: overestimation, overplacement, and overprecision (Moore and Healy, 2008; Moore and Schatz, 2017). Overestimation has to do with absolute values, thinking that you are better than you really are. Overplacement has to do with relative values, thinking that your performance is better than that of others. Finally, overprecision has to do with the degree of certainty with which one judges her own information. 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).

Of the three types of overconfidence, overprecision is the most robust and least studied (Moore et al., 2015). 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 the under-diversification of portfolios as well as asset price volatility (Goetzmann and Kumar, 2008; Scheinkman and Xiong, 2003). In a political context, overprecision leads to ideological extremeness, strong partisan identification (Ortoleva and Snowberg, 2015a,b; Stone, 2019), and increased susceptibility to "fake news" (Thaler, 2020). However, even though overprecision has such negative consequences, we understand very little of it (Mannes and Moore, 2013).

One of the reasons for our lack of understanding of overprecision is that, because it deals with the second moment of the belief distribution, it is hard to measure. The most common way was introduced by Alpert and Raiffa (1982) and consists of asking respondents for the confidence intervals (CI) of a series of numerical questions (e.g., how long is the river Nile).[1] However, the literature has shown that this method creates implausibly high measures of overprecision as respondents do not fully understand how to use CI's (Moore et al., 2015). Other alternatives to measure overprecision are the two-alternative forced-choice (2AFC) (Griffin and Brenner, 2004) or the Subjective Probability Interval Estimate (SPIES) (Haran et al., 2010) which are either not fully suited to elicit individual-level measures of overprecision or too time-consuming (Moore et al., 2015).

In this paper, we contribute to the literature by studying how overprecision affects the political and financial behavior of a nationally representative sample – the 2018 Innovation Sample of the German Socio-Economic Panel (SOEP-IS). To do so, we introduce a new way of eliciting overprecision which we call the "Reported Error Method". This method consists of a two-step procedure where we first ask participants a numerical question (e.g., in which year was Saddam Hussein captured by the US army?) and then ask them to estimate how "far away (in years)" is their response to the first question from the correct answer. In other words, in the second step, we ask respondents to report the absolute error they expect to make in the first question. The two questions are easy to understand and, by comparing the true "realized error" to their reported expected absolute error, we can measure the degree of precision of respondents in a simple and direct way.

The richness of our data-set allows us to study not only the socio-demographic determinants of overprecision, but also to test whether overprecision correlates with several respondent's life outcomes (e.g., portfolio diversification, political attitudes, and voting behavior) and how this behavior aligns with the theoretical predictions. The results show that overprecision (as measured using the Reported Error Method) is negatively correlated with age, years of education, and gross income, but does not differ across genders. More importantly, we find that overprecision has strong predictive power for several theoretical conjectures. For example, our measure predicts larger forecast errors in respondents' stock price predictions and 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),

[1] The idea behind this method is that a perfectly calibrated respondent should get nine out of ten correct answers within the confidence intervals. An overprecise subject would get less than ten out of ten within her CI while an underprecise respondent would get all ten questions within the CI's. In their seminal paper, the 98% CI's of their participants (MBA students) contain the correct answer only 60% of the time.

our measure of overprecision is associated with voting absenteeism rather than with an increased likelihood to vote. This difference might be attributed to the different electoral systems in Germany and the US.

The literature on overprecision is rich, yet only a small number of papers study the effects of overprecision in a representative sample.[2] Ortoleva and Snowberg (2015a,b) estimate a regression-based measure of individual overprecision using a representative sample of the US adult population and study the influence of overprecision on ideological extremeness, partisan identification, and voter turnout. Using the same data set, Stone (2019) studies partisan hostility. However, these papers focus on political preferences and voting behavior, while we go beyond this by including data on financial behavior. Moreover, while Ortoleva and Snowberg (2015a,b) and Stone (2019) have to estimate the individual measure of overprecision of respondents, our new method allows us to elicit, directly and more precisely, the overprecision of respondents.

To summarize, our paper contributes to the existing literature on overprecision in three dimensions: first, we introduce a novel technique, the Reported Error Method, which gives a direct measure of overprecision, is easy to understand, and can be quickly implemented in surveys. Second, we show that in line with the theoretical predictions, our measure shows that a higher degree of overprecision results in lower portfolio diversification and larger stock price forecasting errors as well as ideological extremism. Third, 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) use IT professionals), we test the theoretical predictions using a representative sample of the German population.

Our paper proceeds as follows: Section 2 discusses the notion of overprecision, introduces our measure of overprecision, the Reported Error Method, and presents the SOEP IS dataset. In Section 3 we correlate overprecision with various personal characteristics. In Section 4 we use our measure of overprecision to predict various outcomes in the domains of prediction errors, portfolio diversification, or voting behavior as predicted by theory. The last section concludes.

[2]See Moore et al. (2015) for a detailed survey.

2 Overprecision, the Reported Error Method, and Data Details

2.1 Measuring Overprecision

Overprecision (also known as miscalibration) is a type of overconfidence that results from an excess of confidence in one's own judgment (Moore et al., 2015). Because overprecision deals with the way agents process information, it is widely used in finance and political science to model overconfident agents. For example, in Odean (1998) overconfident traders trade excessively and hold underdiversified portfolios because they believe their private signals to be 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.[3] In the political science literature, Ortoleva and Snowberg (2015b) show that more overprecise people tend to vote more, to hold more extreme political views, and show stronger partisan identification. Stone (2019) suggests that overprecision increases partisanship through excessively strong inferences from (biased) information sources. In fact, a recent strand in the literature has begun to study the role that overprecision plays in the dissemination of fake news (Pennycook et al., 2020; Thaler, 2020).

In all of the mentioned cases, overprecision explains how different types of agents reach different conclusions about the "state of the world." The reason is that overprecision deals with the second moment of the belief distribution and, therefore, directly affects how information is processed. However, it is 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 is by asking for 90% confidence intervals (CI) for a series of numerical questions (e.g., how long is the river Nile). Using this paradigm, a perfectly calibrated respondent would get wrong one question out of every ten. However, the literature has shown that such a method creates implausible high measures of overprecision, with the stated 90% CI's of respondents containing the correct answer only between 30% to 60% of the time (e.g., Russo and Schoemaker, 1992; Bazerman and Moore, 2013; Moore et al., 2015). The best explanation for such results is that respondents are not used to using CI's and do not fully grasp what they are being asked (Moore et al., 2015). This was made clear by Teigen and Jørgensen (2005) who

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

show that the elicited intervals resulting from asking 90% CI's are practically identical to those resulting from asking for 50% CI's.

While there are some alternatives to CI's to measure overprecision, these tend to be either time-consuming or limited in the information they provide. For example, the two-alternative forced-choice (2AFC) by Griffin and Brenner (2004) has respondents 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 if, on average, respondents are overconfident. However, this method has several drawbacks. First, since it cannot distinguish between different types of overconfidence, is not suitable for individual measures of overconfidence. Second, the resulting data has several statistical limitations such as non-continuous probability distributions (see Moore et al. (2015); Griffin and Brenner (2004) for a further discussion of the 2AFC method). Another approach to measure overprecision is the Subjective Probability Interval Estimates (SPIES) method by Haran et al. (2010). The SPIES method elicits complete probability distributions from respondents. While it seems to measure overprecision more accurately than CI's (Moore et al., 2015), it is time-consuming as it requires respondents to first understand the concept of probability distributions and then to build such distributions for each question. 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.

2.2 The Reported Error Method

In contrast to the methods listed in the previous section, we introduce the Reported Error Method, a method that allows us to directly measure the overprecision of respondents in a way that is both easy to understand and simple to implement. The Reported 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

[4]Some examples are, the result multiplying 385 times 67, the length of the Nile, or the year of the death of Lady Diana. 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.

the true answer. In other words, the second question asks respondents to report their expected absolute error. An example would be:

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

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

By comparing the *reported error* of respondents (b) to the realized *true error* in question (a), we get a measure of how over-/underprecise the respondent is about her knowledge.

To fix ideas, assume that the true error of a respondent is normally distributed with mean 0 and variance σ^2 as shown with the blue curve in Figure 1. A perfectly calibrated individual would correctly assess the distribution of the true error when answering questions in the Reported Error Method.[5] However, the perceived distribution of most respondents might not necessarily coincide with the true distribution. If the respondent is overprecise, then her perceived variance $\hat{\sigma}^2$ is smaller than the true variance of the error, i.e., the precision $\rho = 1/\hat{\sigma}^2$ is larger (red curve in Figure 1). In this case, the reported error would consistently deviate from the true error, resulting in a systematic deviation across all questions.[6]

Formally, call the answer of respondent i to question j $a_{i,j}$, her reported error for question j $re_{i,j}$, and the true answer to the question ta_j. Hence, our measure of overprecision for respondent i for question j is:

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

$$overprecision_{i,j} = error_{i,j} - re_{i,j}, \tag{2}$$

[5] An important decision of our design is that we are ambiguous on our definition of "expected absolute error." While most respondents will report their absolute expected average error, there might be some that report a maximum or even their minimum absolute expected error. However, we are willing to accept a small number of outliers (i.e., those subjects reporting maximums or minimums) in exchange for an intuitive design that is easy to understand, fast to implement, and which, overall, produces a robust measure of overprecision (see Section 4). Additionally, we also assume that respondents correctly assess the first moment of the distribution and. With this, we can abstract from the case in which respondents are miscalibrated for *both* the mean and the precision of their errors. We address this case in Appendix C.

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

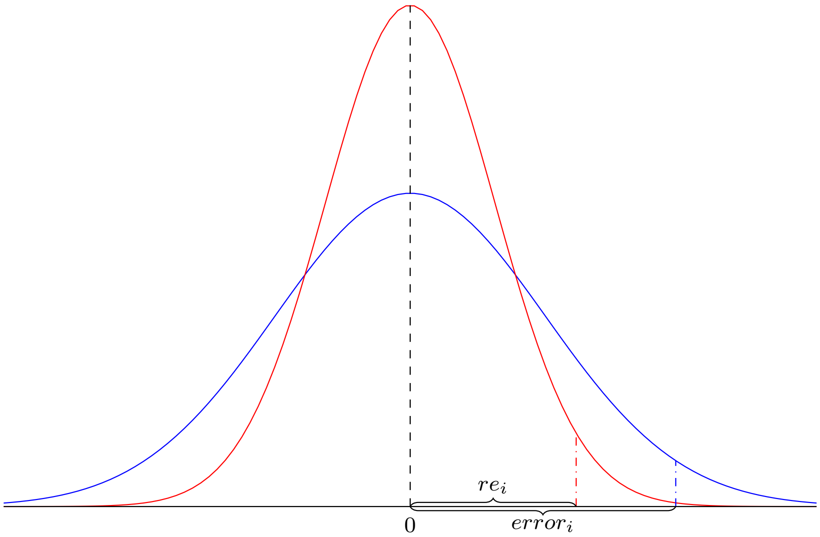


Figure 1: The figure shows two hypothetical normal distributions of the error. The blue curve shows the true distribution of the error with a standard deviation of 2 (precision of .25). The red curve shows the perceived distribution by an overprecise respondent with a standard deviation of 1.25 (precision of .64). The dashed vertical lines indicate the reported error re_i and the true errors $error_i$ resulting from an overprecise respondent.

where equation (1) measures the true error ($error_{i,j}$) of respondent i to question j. Notice, that this equation calculates the *absolute error*, that is, we do not care about the direction of the error, but rather the size of the error. In equation (2), we calculate the difference between the reported error ($re_{i,j}$) and the true error ($error_{i,j}$) of respondents i to question j. This time we do care about the direction of the error, as a respondent who underestimates her expected absolute error (i.e., $error_{i,j} > re_{i,j}$) is considered to be *overprecise*, while a respondent who overestimated the expected absolute error (i.e., $error_{i,j} < re_{i,j}$) is *underprecise*. Finally, those respondents who correctly guessed their expected absolute error (i.e., $error_{i,j} = re_{i,j}$) are considered to be perfectly calibrated for that question.

Eliciting overprecision using the Reported Error Method rather than confidence intervals has several advantages. First and foremost, respondents do not need to have any statistical knowledge to answer the questions and the setup is easy to explain. Additionally, questions can be answered quickly, and it can be implemented easily in either computerized or pen and paper surveys. Another important advantage of the Reported Error Method is that it is easy to make it incentive-compatible as one can put a payment mechanism such as a quadratic scoring rule (Brier, 1950) or the binarized scoring rule (Hossain and Okui, 2013) on top of each question, and then pay randomly only one of the two outcomes to avoid hedging across questions. This stands in contrast with the more complicated scoring rules necessary to make CIs incentive-compatible (e.g., Jose and Winkler, 2009).

In a recent paper Enke and Graeber (2019) study the "subjective uncertainty about the optimal action" that online subjects have when confronted with choices across different economic domains. To measure such uncertainty, they take an approach very similar to the Reported Error Method, allowing 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 of Enke and Graeber (2019) is not designed to measure overprecision, it lends support to the Reported Error Method as a robust tool to elicit the degree of uncertainty of respondents for a given answer.

2.3 Data

The data that we use comes from the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). The Innovation Sample is a subset of the larger SOEP-Core panel (SOEP-Core, which has approximately 30,000 individual respondents) and it is designed to host and test novel survey items (see, Richter et al., 2015). We use the 2018 wave of the SOEP-IS which in total had 4,860 individual respondents distributed across 3,232 different households. As in the SOEP-Core, all interviews of the SOEP-IS are conducted face-to-face by a professional interviewer.

The data we use for the construction of our measure is composed of seven different questions. The questions are presented in the same order to all respondents and in each question we ask respondents to answer two things, (a) the year of a specific historical event that occurred not further away than 100 years (b) the distance (in years) between their answer to (a) and the correct answer to (a).[7] In other words, we ask respondents to answer a general knowledge question and then we ask them to report the absolute error they expect to make, i.e., their reported error (see Section 2.2).

[7]The precise formulation was 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?*.

We asked seven different questions about events taking place between 1938 and 2003. The questions were thought 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 Microsoft was founded (1975) or when Saddam Hussein was captured by the US Army (2003) (see Table B.1 in the appendix for all of the questions and their correct answers).[8] These questions were asked to a subset (902) of the respondents in the SOEP-IS 2018 who joined the panel in 2016. We supplement the data with additional personal characteristics from the survey years 2016-2018. We drop 55 respondents who did not answer any of the overprecision questions, since this is the 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.[9]

3 Socio-demographic Determinants of Overprecision

In Figure 2, we plot the density of the answer $a_{i,j}$ for each question j. The red vertical line marks the correct answer. It is clear from the dispersion of the densities that some questions were easier for respondents than others. In Figure 3, we plot the true error ($error_{i,j}$) in the vertical axis and reported error ($re_{i,j}$) 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 ($error_{i,j} > re_{i,j}$) in her answer to the question, and any point below corresponds to a respondent who is underprecise ($error_{i,j} < re_{i,j}$). It is clear from the figure that respondents are overprecise in their answers across all questions, independent of the difficulty.

Since overprecision is measured across seven different questions, internal consistency across them is important. To measure such consistency we use congeneric reliability, commonly referred to as coefficient omega (see e.g., Cho, 2016). The congeneric reliability measure is defined as $T_i = \mu_i + \lambda_i F + e_i$, where T_i is the outcome of item i with mean μ_i, e_i

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

[9]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 sub-sample is at large still representative of the larger SOEP-Core with only some significant but small differences. When applying the sampling weights to our estimation sample as well, the differences disappear.

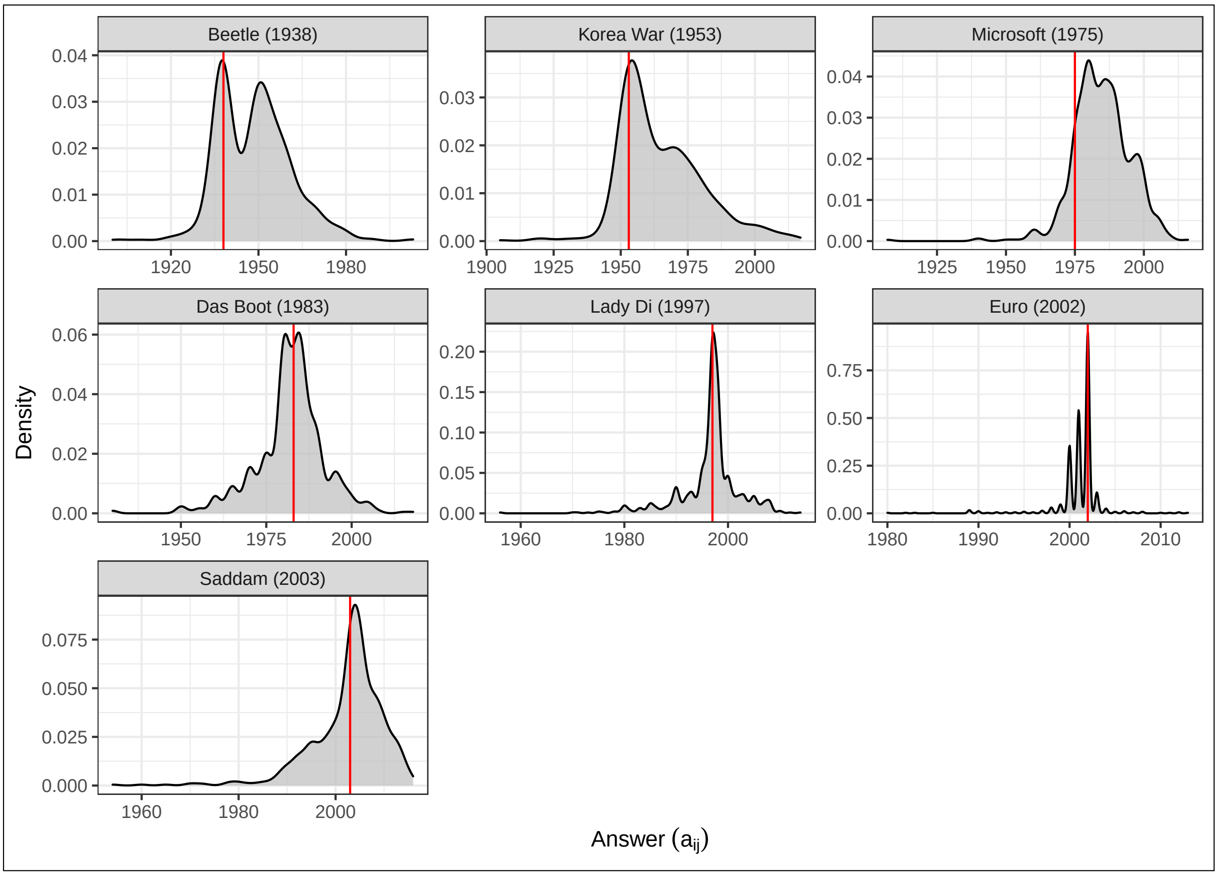


Figure 2: Density of the answers ($a_{i,j}$) for each question. The red vertical line marks the correct answer. Note that the vertical axis is different for each question.

is the score error and λ_i is the factor loading on the latent common factor F (Morera and Sotkes, 2016). It is a generalized version of the Cronbach α (Cronbach, 1951), but allowing for different factor loadings of the latent common factor.[10] To construct the congeneric reliability measure, we estimate the factor loadings, $\hat{\lambda}_i$, for the overprecision measure of each question with respect to a common factor (i.e., overprecision) and estimate the congeneric reliability according to the formula $\frac{(\sum \hat{\lambda}_i)^2}{(\sum \hat{\lambda}_i)^2 + \sum \hat{\sigma}^2_{e_i}}$, where $\hat{\sigma}^2_{e_i}$ is the estimated variance of the error. This results in a congeneric reliability of .76.

To combine the overprecision measures across all seven questions into a unique value for each respondent (OP_i), we average the measure of overprecision ($overprecision_{i,j}$) for each respondent (i) across all questions (j).[11] We plot the density of OP_i in Figure 4a.

[10]For the case of τ-equivalence, i.e., $\lambda_i = \lambda_j \; \forall i, j$, all factor loadings are equal and both measures coincide.

[11]An alternative would be to construct the composite measure OP_i using a principal component approach as in Ortoleva and Snowberg (2015b). The result of using such an approach is very similar to

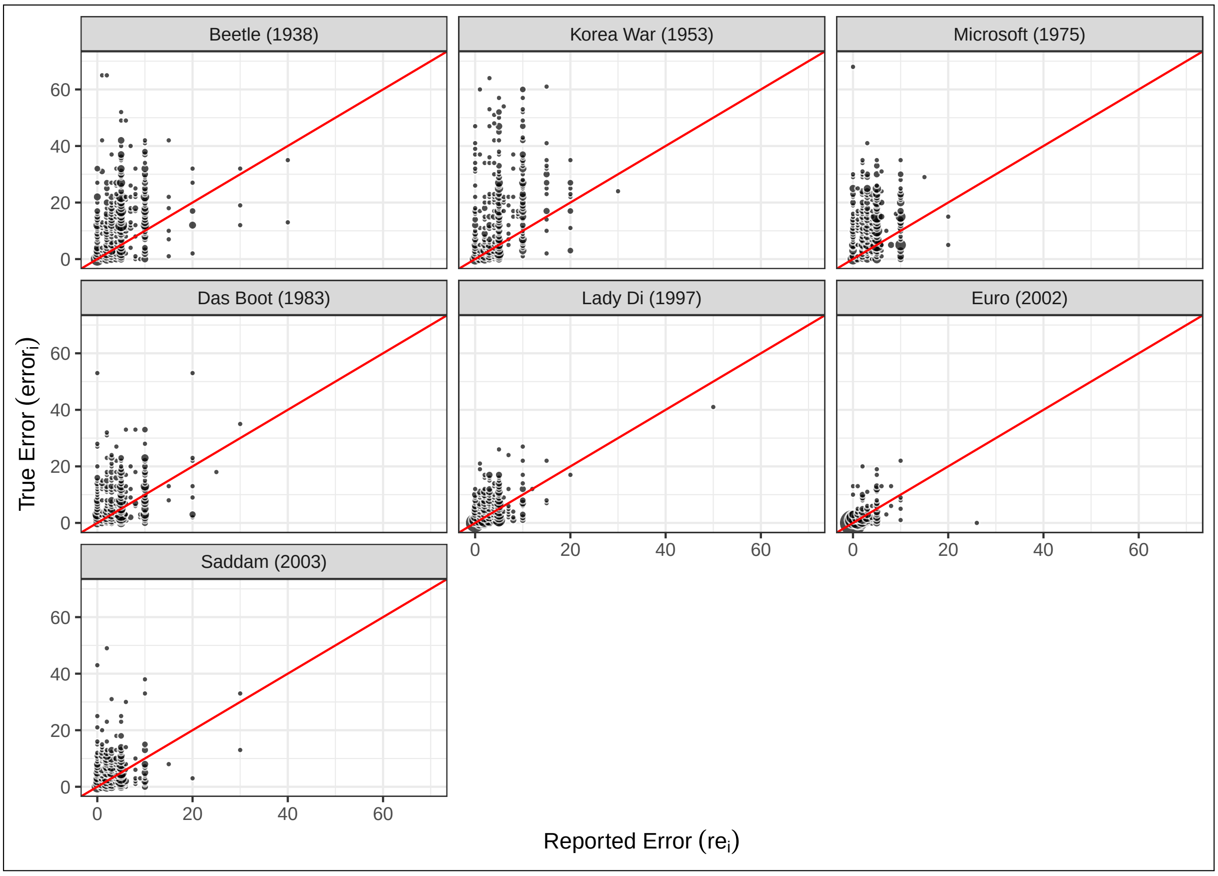
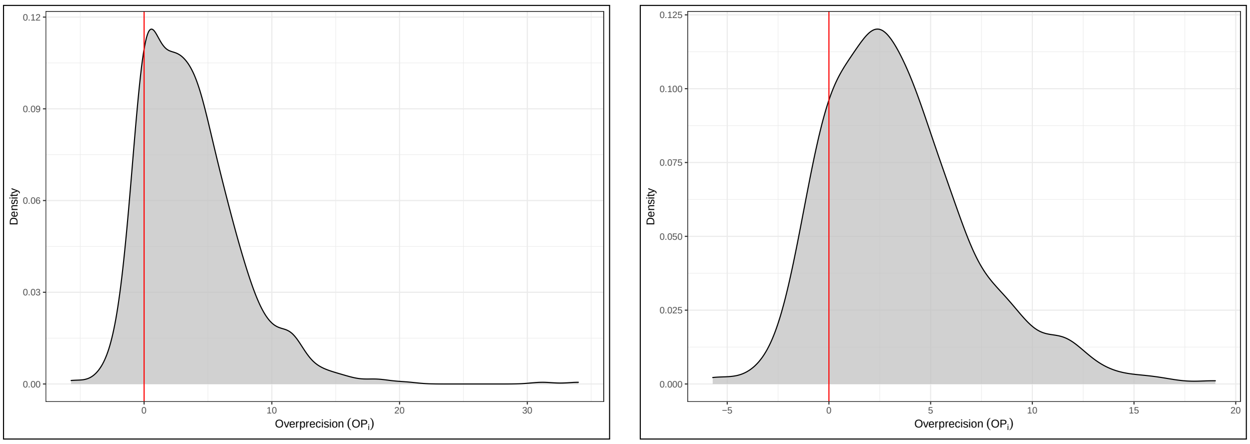


Figure 3: Relation between the true error ($error_{i,j}$) in the vertical axis and the reported error ($re_{i,j}$) in the horizontal axis. Any dot above (below) the 45-degree red line is an overprecise (underprecise) answer by the respondent.

In line with Figure 3, Figure 4a shows that the large 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 respondents seem to be perfectly calibrated (vertical red line in Figure 4a) in the aggregate measure. Of these 56 respondents, 88% are perfectly calibrated across all the questions they answer. However, one should take into account that respondents could decide not to answer a question; 51% of the respondents answered all questions with 5% answering only one (see Figure A.2 in the appendix for a detailed breakdown). Of those respondents that are perfectly calibrated 39% answered only one question and only 7% answered all seven. This means that what we see in Figure 4a is an

using the average ($\rho^{Pearson} = .88; \rho^{Spearman} = .84; N = 805$).



(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.

"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*, then respondents are substantially less calibrated, with the mode of OP_i shifting to the right and leaving only 1% of the respondents being perfectly calibrated, while, at the same time, having an increase in the proportion of underprecise respondents (15%).

For ease of interpretation, we standardize the aggregate score (OP_i) to be mean zero and standard deviation one (Sop_i, henceforth). In Table 1 we regress Sop_i on a series of socio-demographic variables using four different OLS models. In all the models we control for age, gender, and years of education. Of these, both age and education are significantly and negatively correlated with overprecision. Given the units, both seem to have a relatively large impact. For example, for every two years of education overprecision is reduced by about one-tenth of a standard deviation. It is also important to note that the number of questions answered by respondents (*Answered*), which we include in Column (2), is not random, with overprecision increasing as subjects answer more questions (see Figures A.1 and A.2 in the appendix for a graphical overview of these results). In all of the subsequent analysis, we control for the number of answered questions.

In Columns (3) and (4) we add the monthly gross individual income (*Gross Income*) measured in thousands of euros.[12] Additionally, we add dummies for the labor force

[12]Since this item is only available for employed individuals, we code missing variables as 0 and include a dummy that is one for missing observations.

Dependent Variable: *Sop*	(1)	(2)	(3)	(4)
Age	-0.008***	-0.007***	-0.007**	-0.007**
	(0.002)	(0.002)	(0.003)	(0.003)
Female	0.085	0.132*	0.103	0.082
	(0.069)	(0.070)	(0.073)	(0.072)
Years Education	-0.053***	-0.063***	-0.050***	-0.044***
	(0.013)	(0.013)	(0.014)	(0.014)
Answered		0.063***	0.066***	0.070***
		(0.020)	(0.020)	(0.021)
Gross Income			-0.051**	-0.051**
			(0.023)	(0.023)
Constant	1.056***	0.772***	0.652**	0.502
	(0.209)	(0.227)	(0.311)	(0.377)
N	805	805	805	805
adj. R^2	0.035	0.046	0.060	0.083
Fixed Effects	No	No	No	Yes
Employment Status Dummy	No	No	Yes	Yes

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Determinants of overprecision. In Columns (1) - (4) we run an OLS with *Sop* as the dependent variable. In Column (3) we include dummies for the labor force status (employed, unemployed, retired, maternity leave, non-working), and whether the respondent was a citizen of the GDR before 1989. In Column (4) we also include fixed effects for the federal state (Bundesland) where the respondent lives and the time at which he/she responded to the questionnaire.

status (e.g., employed, unemployed, maternity leave, etc.) as well as a dummy for those respondents that were living in East Germany in 1989. The results show a relatively strong effect of income on overprecision, with every two thousand euros reducing overprecision by almost one-tenth of a standard deviation. Finally, in Column (4) we add federal state (Bundesland) and time of interview fixed effects, which have no qualitative effects on the results of the model of Column (3). We include all of the above-mentioned variables as control variables in the subsequent analyses.

The results from Table 1 are in contrast with those of Ortoleva and Snowberg (2015b) who find that neither income nor education is correlated with their measure of overprecision and that females are significantly less overprecise than males. In Appendix C we test the robustness of our approach to measuring overprecision by contrasting it to five alternative approaches in all of the analyses. Those are i) a *residual* approach following the regression methodology of Ortoleva and Snowberg (2015b), ii) a *relative* approach which takes into account the relative distance between the reported error and the true error, iii) a *standardized* measure which standardizes each question before aggregating them, iv) an *age-robust* measure which is constructed using only those questions of events which occurred after the respective respondent was born, and v) a *centered* measure which centers the errors and reported errors around their mean allowing us to disentangle the second moment of the distribution (overprecision) from its first moment.

For all five alternative measures, we run the same OLS models as in Table 1. The results can be found in Table C.1 in the appendix, and show that the impact of the socio-demographic characteristics is overall robust across the different measures of overprecision we build.[13] Furthermore, there is a high correlation between the baseline measure and 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 ($\rho^{Pearson} = .96$; $\rho^{Spearman} = .93; N = 805$). We, therefore, proceed to apply our original measure to analyze the impact on real-life outcomes. Any concerns about confounding effects should be mitigated by the inclusion of the determinants as controls.

4 Overprecision and Real Life Outcomes

In this section, we examine how overprecision correlates with real-life outcomes in the domains of financial markets and politics. In Section 4.1 we describe the empirical methodology to then present the predictions derived from the literature and our results in Section 4.2.

4.1 Methodology

To test the predictions from the theoretical literature on overprecision 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:

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

[13] There are a few exceptions such as being female, which has a negative effect on overprecision only in one of the five measures.

where Sop_i denotes the standardized overprecision measure, $\boldsymbol{X_i}$ a vector of control variables, and ϵ_i is the random error term. We include all possible control variables we assume to be correlated either with the dependent variable or with overprecision. These are: age, gender, years of education (which serves as a proxy for cognitive ability), the 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 GDR (German Democratic Republic) in 1989 as well as interview date (month and year) and state fixed effects. Additionally, we include a measure of political interest in the political analyses, and as mentioned in section 3, we control for the number of overprecision questions answered by each respondent. [14] A test for multicollinearity shows no strong linear dependencies across 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.[15] Since we test several different outcomes, we also report the Sidak-Holm adjusted p-value for multiple hypothesis testing in Column (3).

Second, we follow Cobb-Clark et al. (2019) and estimate the "R^2 rank" of our variable of interest (Sop_i). This is obtained by running a step-wise regression in which we sequentially keep adding variables to the model. To do so, we regress the outcome on each variable in our vector of possible controls separately and pick the variable that delivers the highest R^2. We then proceed and regress the outcome on the variable from the previous round and each of the remaining variables separately. This is continued until all variables have been added to the model. The higher the rank of our variable of interest, 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. The regressions with political outcomes as dependent variable additionally include a self-reported measure of political interest.

[14]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.

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

Finally, we employ a least absolute shrinkage and selection operator (LASSO) to test whether our overprecision measure has predictive power for the outcome variable. 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 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.[16] 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.[17] If our variable of interest, in our case Sop_i, is included in the model, then it has predictive power for the outcome. We report the results of this exercise in Column (5) of Table 2 along with the number of control variables in the model chosen by LASSO and the resulting R^2 of the model in Column (6) and the number of observations in Column (7). The number of observations varies due to missing observations in the outcome variables.[18]

4.2 Prediction Results

The results of our three analytical approaches are summarized in Table 2. We first discuss financial market outcomes and then outcomes regarding political behavior.

Financial Market Outcomes

We first test predictions regarding financial market outcomes. The first set of predictions concerns forecast errors of asset price predictions in the stock market and the real estate market. The theory of overprecision in financial markets argues that overprecise investors overweigh their private signals when forming expectations, leading investors to hold incorrect beliefs about the future valuation of an asset (e.g., Benos, 1998; Odean, 1998).

[16] 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.

[17] Thereby, the algorithm proceeds stepwise 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.

[18] 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).

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) LASSO R^2	(7) N
A Prediction error:							
err_dax	1.153**	0.022	0.105	2/38	yes/15	0.15	578
opt_dax	0.091***	0.009	0.061	3/38	yes/11	0.39	578
err_rent	0.348*	0.051	0.145	2/38	yes/13	0.07	670
err_buy	0.160	0.264	0.458	9/38	no/0	0.00	644
B Diversification:							
std_divers	-0.129***	0.000	0.000	3/38	yes/19	0.13	774
C Ideological Positioning:							
std_extreme	0.091**	0.032	0.122	6/39	yes/13	0.05	716
std_lr	-0.011	0.801	0.801	18/39	no/11	0.07	716
D Voting behavior:							
non_voter	0.032***	0.010	0.059	3/39	yes/18	0.14	706

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

Table 2: This table shows the estimation results of Section 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 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 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 4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

Moreover, the disagreement regarding the fundamentals of assets driven by overprecision leads traders to contribute to asset price bubbles as they over-optimistically believe that in the future they will find a buyer paying an even higher price (e.g., Scheinkman and Xiong, 2003; Hong et al., 2006). Direct empirical support for this association of overprecision and forecast errors in financial markets is provided by Deaves et al. (2019) who analyze the predictions of German stock market forecasters and correlate them to an accompanying survey-based measure of overprecision (using confidence intervals). Additionally, Hilary and Menzly (2006) provide evidence consistent with this association for Northern American analysts and Hayunga and Lung (2011) the US real estate market. The former derive a proxy of overprecision from forecasts, while the latter proxy overprecision with excessive asset turnover.

Following the logic outlined above, we expect that overprecise respondents exhibit a lower forecast performance and that overprecise respondents systematically make prediction errors in the positive direction, i.e., that they overestimate returns to the stock market and the real estate market. We test the first prediction using the absolute distance of one year ahead predictions of the German Stock Index (DAX), Germany's blue-chip stock market index, from the realized value (*err_dax*).[19] We test the second prediction using the standardized difference between the one year ahead prediction and the realized value so that a positive value denotes an overestimation of the stock market realization (*opt_dax*). Analogously, following Hayunga and Lung (2011), we expect overprecise respondents to systematically make prediction errors regarding the development of real estate markets. Therefore, we test the predictive power of overprecision on the absolute distance of two years ahead predictions of German housing and rental prices from the realized values (*err_rent* and *err_buy* respectively).[20]

The results in Table 2 show that our measure of overprecision is an important predictor of forecast errors in asset prices. A one standard deviation increase in overprecision is associated with an increase in the absolute forecast error of 1.15 percentage points and a .09 standard deviation increase in the overestimation of the one-year-ahead stock market forecast. Moreover, the results show that a one standard deviation increase in overprecision leads to an increase in the absolute forecast error of rental and housing prices whereby the latter is less strong. The results from the LASSO estimation reveal that overprecision is also a good predictor of these forecast errors since it is selected as an explanatory variable for the models of the stock market forecast and the rental prices, and ranking high (between 2 and 9) in the R^2 rank approach.

Next, we test the theoretical prediction of Odean (1998) that overprecision is associated with underdiversified portfolios. Intuitively, overprecise investors overweight their private information, thereby trading too frequently while being concentrated on too few favorable assets. Goetzmann and Kumar (2008) provide empirical evidence in support of this prediction traders in the US and Merkle (2017) for traders in the UK. While the former relies on the asset turnover proxy, the latter elicits overprecision directly through

[19]Note that the observations from the 2018 waves are almost all within the period before March 2019 and are thus unaffected by the stock market decline caused by the Corona-crisis in March 2020.

[20]We include a dummy variable that indicates house ownership as well as 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 with the sample size decreasing.

survey questions. We test this hypothesis for the representative German sample using a standardized measure with mean zero and standard deviation of one that captures the degree to which a respondent diversifies her hypothetical portfolio among stocks, real estate, government bonds, savings, and gold (*std_divers*).[21]

Our results for the German sample confirm the theoretical prediction that overprecision is associated with underdiversification. The point estimate in Column (1) in Table 2 shows that a one standard error increase in overprecision leads to a .13 standard deviation decrease in our diversification measure. That means that their optimal portfolio is skewed towards a certain asset category. Moreover, overprecision is among the variables chosen by the LASSO estimation and ranked third in the R^2 rank approach.

Political Views and Voting Behavior

For the last set of predictions, we study the political views and voting behavior of the respondents. Ortoleva and Snowberg (2015b) define overprecision in the political context as the belief that one's own experiences are more informative about actual politics than they really are. For instance, overprecise people may visit biased media outlets, without fully accounting for this bias, exchange information on social media without realizing that much of the information comes from politically like-minded peers, or inferring from adverse local economic conditions on the general state of the economy opting for drastic policy responses. Against this background, the authors show theoretically and empirically that overprecision in one's own beliefs leads to ideological extremeness and strengthens the identification with political parties, increasing the likelihood to vote. Yet, the literature remains inconclusive 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.

We use several measures to test the predictive power of overprecision for political views and voting behavior. First, to test whether overprecision predicts ideological extremeness, we use a measure on a scale from 0 to 5 that indicates how far away from the center respondents see themselves in the political spectrum (*std_extreme*). Second,

[21]For a detailed description of the measure refer to Table B.3 in the appendix.

to test whether overprecision tends to push respondents towards one end of the ideological spectrum, we use respondents' answers to the question where they see themselves in the political spectrum with 0 being extreme left and 10 being extreme right (*std_lr*). Third, to test whether overprecise respondents a more likely to vote, we use a dummy that equals one if a respondent indicated to be a non-voter in the (ex-post) opinion poll (Sonntagsfrage) for the 2017 federal elections to the German Bundestag (*non_voter*).

In line with Ortoleva and Snowberg (2015b), our results suggest that overprecision is a good predictor of ideological extremeness. Overprecision is among the variables chosen by the LASSO estimation and ranking high (sixth) in the R^2 rank approach. Confirming Ortoleva and Snowberg (2015a), we do not find evidence that overprecision is associated more strongly with conservatism or liberalism, as overprecision is neither correlated to political ideology nor among the variables chosen by the LASSO estimation. Furthermore, overprecision is ranked quite low (18/39) 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, as it seems, overprecision increases the likelihood of voting absenteeism rather than increasing the likelihood of voting: with a one standard deviation increase in overprecision resulting 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, one should be cautious when comparing the voting behavior of overprecise respondents in the US and Europe. 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). As opposed, in Germany, more extreme respondents gravitate to fringe parties (e.g., Die Linke, AfD, NPD)[22] with smaller chances of winning elections, so the incentives to vote are very different than for those in the dataset of Ortoleva and Snowberg (2015b).[23] Hence, the theoretical assumptions underlying Ortoleva and Snowberg (2015b)'s predictions of voter turnout and

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

[23]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 resulted after the 2019 Thuringian election shows how strongly such *cordon* is enforced.

overprecision are a good description of voting behavior in the two-party system of the US, but are not appropriate for the more disperse German system.

5 Conclusion

We study how overconfidence correlates with the behavior of a nationally representative sample. To do so, we implement the Reported Error Method in the 2018 wave of the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). The Reported Error Method is a new way to measure overprecision that, in contrast to previous methods, is intuitive to respondents and quick to implement. We exploit the rich data of the SOEP-IS to test how well our new measure of overprecision can predict the financial and political behavior of respondents using both linear regressions and LASSO techniques.

The results show that our novel measure of overprecision lends empirical support to several theoretical predictions from the financial and political science literature. For example, overprecision predicts larger forecast errors in the prediction of blue-chip stock prices (Odean, 1998) and lower levels of portfolio diversification (Barber and Odean, 2000). Additionally, as predicted and shown in Ortoleva and Snowberg (2015a), more overprecise respondents will hold more extreme political ideologies. As for the socio-demographic determinants of overprecision, we find that years of education, age, and gross income reduce the overprecision of respondents, but do not detect any effect of gender on overprecision. Both the predictive power and the socio-demographic determinants are robust to a series of modifications, showing the appropriateness of our approach.

From a methodological point of view, our results show that the Reported Error Method produces a measure of overprecision that is consistent across different domains. For example, we use general knowledge history questions to measure the overprecision of respondents to then predict their political and financial behavior. The ability to cover different domains is a valuable advantage of the Reported Error Method as it allows researchers a high degree of freedom in the choice of their elicitation questions.

Overall, our work contributes to a large literature that tries to understand overconfidence, "the most significant of the cognitive biases" (Kahneman, 2013), and how it affects our lives. Because overconfidence can result in reckless behavior and lead to extreme political views of the world, our results and methodology should be of interest not only to economists and political scientists, but also to psychologists, financial researchers, and policymakers.

References

Alpert, M. and H. Raiffa (1982): "A progress report on the training of probability assessors," . Cited on pages 2 and 4.

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 22.

Bazerman, M. H. and D. A. Moore (2013): *Judgment in managerial decision making*, New York: Wiley, 8th edition ed. Cited on page 5.

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 17.

Brier, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly weather review*, 78, 1–3. Cited on page 8.

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

Cho, E. (2016): "Making Reliability Reliable:A Systematic Approach toReliability Coefficients," *Organizational Research Methods*, 19, 651–682. Cited on page 10.

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 16.

Cronbach, L. J. (1951): "Coefficient alpha and the internal structure of tests," *psychometrika*, 16, 297–334. Cited on page 11.

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 page 18.

Enke, B. and T. Graeber (2019): "Cognitive uncertainty," Tech. rep., 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 19.

Griffin, D. and L. Brenner (2004): "Perspectives on probability judgment calibration," *Blackwell handbook of judgment and decision making*, 177–199. Cited on pages 3 and 6.

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

Haran, U., D. A. Moore, and C. K. Morewedge (2010): "A simple remedy for overprecision in judgment," *Judgment and Decision Making*, 5, 467. Cited on pages 3 and 6.

Hayunga, D. K. and P. P. Lung (2011): "Explaining Asset Mispricing Using the Resale Option and Inflation Illusion," *Real Estate Economics*, 39, 313–344. Cited on pages 18 and 19.

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

Hong, H., J. Scheinkman, and W. Xiong (2006): "Asset Float and Speculative Bubbles," *The Journal of Finance*, 61, 1073–1117. Cited on page 18.

Hossain, T. and R. Okui (2013): "The binarized scoring rule," *Review of Economic Studies*, 80, 984–1001. Cited on page 8.

Johnson, D. D. (2009): *Overconfidence and war*, Harvard University Press. Cited on page 2.

Jose, V. R. R. and R. L. Winkler (2009): "Evaluating quantile assessments," *Operations research*, 57, 1287–1297. Cited on page 9.

Kahneman, D. (2013): *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 1st edition ed. Cited on pages 2 and 22.

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*. 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 page 4.

Merkle, C. (2017): "Financial overconfidence over time: Foresight, hindsight, and insight of investors," *Journal of Banking & Finance*, 84, 68–87. Cited on page 19.

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 21.

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, e12331. Cited on page 2.

Moore, D. A. and S. A. Swift (2011): "The three faces of overconfidence in organizations," in *Organization and management series. Social psychology and organizations*, Routledge/Taylor & Francis Group, 147–184. Cited on page 20.

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, 3, 4, 5, and 6.

Morera, O. F. and S. M. Sotkes (2016): "Coefficient α as a Measure of Test Score Reliability: Review of 3 Popular Misconceptions," *American Journal of Public Health*, 106, 458–461. Cited on page 11.

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, 17, 19, and 22.

Ortoleva, P. and E. Snowberg (2015a): "Are conservatives overconfident?" *European Journal of Political Economy*, 40, 333–344. Cited on pages 2, 4, 20, 21, 22, 42, and 44.

——— (2015b): "Overconfidence in political behavior," *American Economic Review*, 105, 504–35. Cited on pages 2, 3, 4, 5, 11, 14, 20, 21, 37, 40, 42, and 43.

Pennycook, G., Z. Epstein, M. Mosleh, A. Arechar, D. Eckles, and D. Rand (2020): "Shifting attention to accuracy can reduce misinformation online," Tech. rep., mimeo. 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 page 42.

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 16.

Richter, D., J. Schupp, et al. (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 5.

Scheinkman, J. A. and W. Xiong (2003): "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111, 1183–1220. Cited on pages 2, 5, and 18.

Stone, D. F. (2019): ""Unmotivated bias" and partisan hostility: Empirical evidence," *Journal of Behavioral and Experimental Economics*, 79, 12–26. Cited on pages 2, 4, 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 page 5.

Thaler, M. (2020): “The ”Fake News” Effect: Experimentally Identifying Motivated Reasoning Using Trust in News,” Tech. rep., mimeo. 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 17.

A Extra Figures

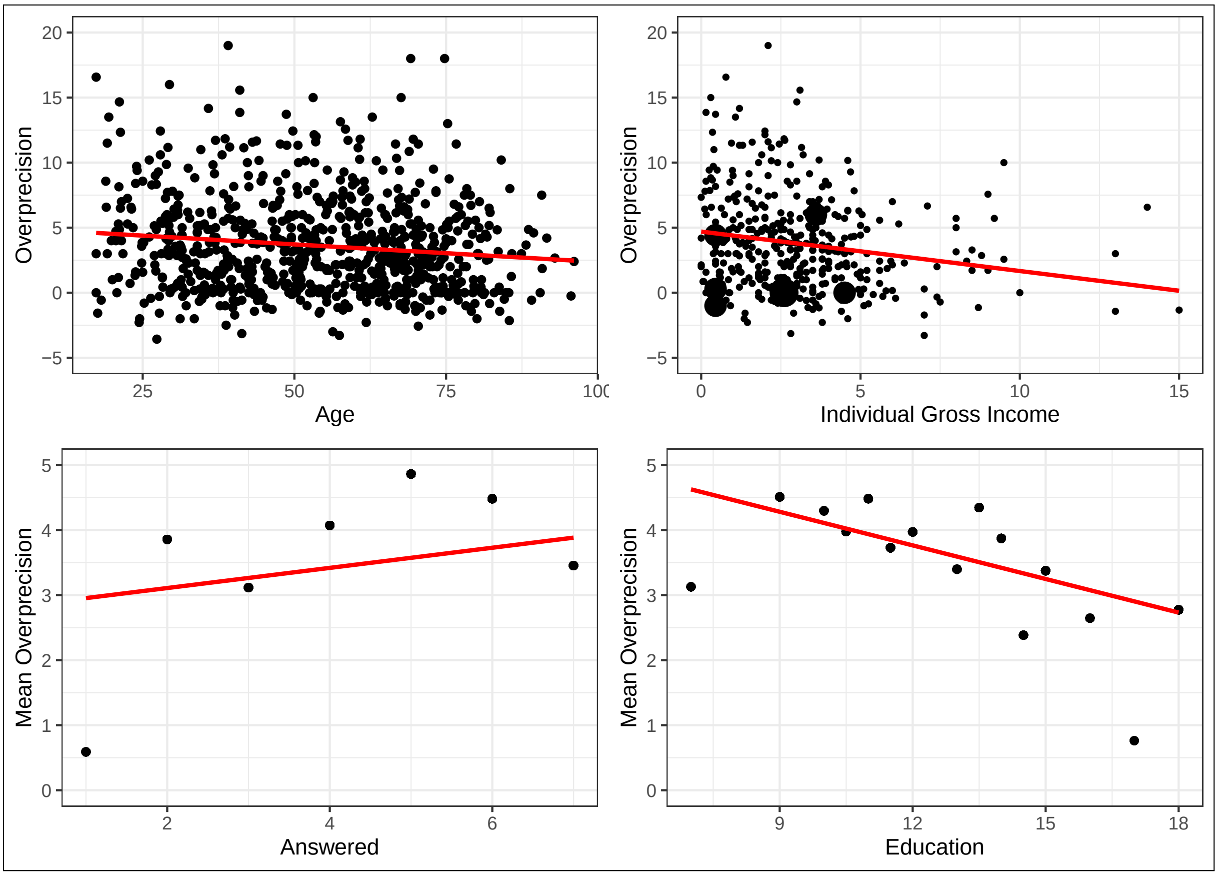


Figure A.1: Correlation of Overprecision. In the vertical axis of each panel we plot the overprecision (upper row) and mean overprecision across all groups which we plot in the horizontal axis (lower row). In all four cases the red line is the fitted linear regression. We dropped one individual outlier in all cases to make the graphs more readable.

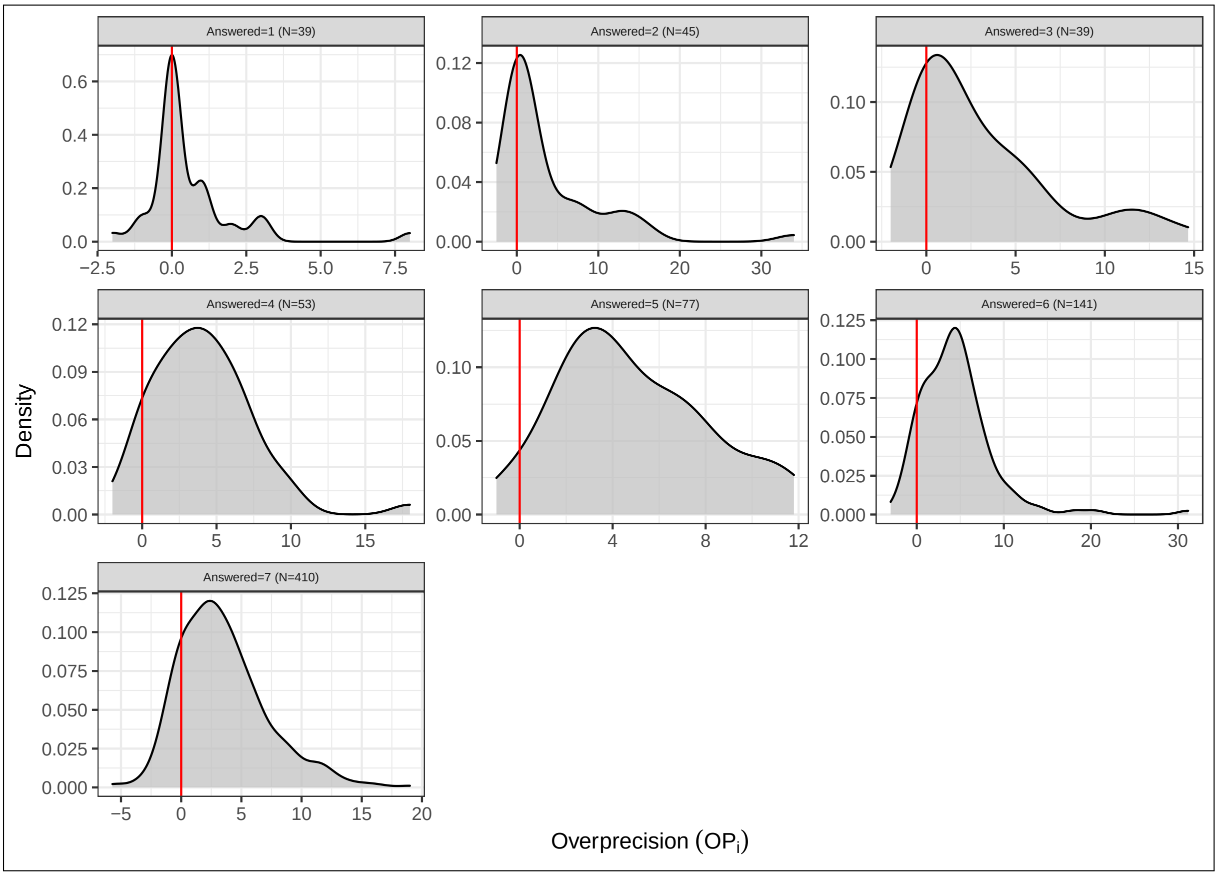


Figure A.2: Density of Overprecision (OP_i) for each of the subsets of questions answered. We plot from left to right the densities of OP_i for those respondents who answered from the minimum number of answers (1) to the maximum number of answers (7). In the title we report the number of respondents for each density. Notice that the scale of the Y-axis changes across panels.

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?	1983
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 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?	1983
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 questions in German language from the 2018 SOEP-IS

Variable	Definition
A Prediction error:	
err_dax	Absolute distance between one year-ahead prediction of the DAX realization and the actual realization over the horizon. Data from the first trading day of each month was used depending on the month of the interview. The data does not contain the Corona crash.
opt_dax	Difference between one year-ahead prediction of the DAX realization and the actual realization over the horizon. Positive values indicate an overestimation of the returns. Data from the first trading day of each month was used depending on the month of the interview. The data does not contain the Corona crash.
err_rent	Absolute distance between one year-ahead prediction of rental prices in Germany and the actual realization over the horizon. One year-ahead predictions were linearly derived from two year-ahead predictions. Quarterly data according to the month of the interview was used.
err_buy	Absolute distance between one year-ahead prediction of house prices in Germany and the actual realization over the horizon. One year-ahead predictions were linearly derived from two year-ahead predictions. Quarterly data according to the month of the interview was used.
B Diversification:	
std_divers	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 maximum attainable penalty score less actual penalty. The diversification measure is standardized to have mean 0 and standard deviation 1.

Variable	Definition
C Ideological Positioning:	
std_extreme	Absolute distance to center of an ideology scale from 0 (left) to 10 (right). Standardized to have mean 0 and standard deviation 1.
std_lr	Location on an ideology scale from 0 (left) to 10 (right). Standardized to have mean 0 and standard deviation 1.
D Voting Behavior:	
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.
gender	=1 if female.
east1989	=1 if living in East Germany in 1989.
std_risk	Location on risk scale from 0 (risk avers) to 10 (risk loving). Standardized to have mean 0 and standard deviation 1.
pgbilzt	Years of education.
pglabgro	Monthly gross labor income in thousands. Missings are coded with zero.
mispglabgro	=1 if missing pglabgro.
finlit	Share of correct answers to 6 questions related to financial knowledge.
owner	=1 if living in own property.
owner_rent	=1 if earning money from renting out property.
assets	=1 if owning financial assets.
std_narcis	Average narcissiscm measure over 6 items on scale from 1 to 6. Standardized to have mean 0 and standard deviation 1.
std_impuls	Location on impulsivity scale from 0 (not impulsive) to 10 (fully impulsive). Standardized to have mean 0 and standard deviation 1.
std_patient	Location on patience scale from 0 (not patient) to 10 (fully patient). Standardized to have mean 0 and standard deviation 1.
empl	=1 if employed.
unempl	=1 if unemployed.
nonwork	=1 if non-working.

Variable	Definition
matedu	=1 if on maternity, educational or military leave.
retire	=1 if retired.
answered	Number of questions answered for overprecision.
pol_int	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 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
Gender	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
East (1989)	0.186	(0.014)	0.198	(0.004)	0.012	0.404	24,591
Years Education	12.704	(0.098)	17.276	(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 sub-sample. This table shows the descriptives of selected personal characteristics of the respondents for 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) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	1.321**	0.014	0.081	2/43	yes/20	0.15	537
opt_dax	0.094***	0.010	0.068	3/43	yes/20	0.41	537
err_rent	0.335*	0.056	0.159	5/43	yes/2	0.01	624
err_buy	0.140	0.342	0.567	14/43	no/1	0.01	602
B Diversification:							
std_divers	-0.117***	0.002	0.016	4/43	yes/15	0.13	719
C Ideological Positioning:							
std_extreme	0.081**	0.048	0.179	7/44	yes/14	0.06	716
std_lr	-0.021	0.594	0.594	23/44	no/17	0.10	716
D Voting Behavior:							
non_voter	0.029**	0.015	0.073	3/44	yes/20	0.14	706

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

Table B.5: This table shows the estimation results of Section 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 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 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 4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) LASSO R^2	(7) N
A Prediction error:							
err_dax	1.011**	0.043	0.197	5/41	yes/22	0.18	573
opt_dax	0.096***	0.007	0.048	3/41	yes/11	0.40	573
err_rent	0.323*	0.075	0.209	7/41	yes/18	0.08	660
err_buy	0.158	0.279	0.480	9/41	no/0	0.00	634
B Diversification:							
std_divers	-0.126***	0.001	0.008	4/41	yes/19	0.14	762
C Ideological Positioning:							
std_extreme	0.084*	0.053	0.196	6/42	yes/14	0.04	705
std_lr	-0.001	0.980	0.980	20/42	no/15	0.08	705
D Voting behavior:							
non_voter	0.030**	0.016	0.092	3/42	yes/17	0.12	693

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

Table B.6: This table shows the estimation results of Section 4 including assets and home-ownership as controls. 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 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 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 4.1 along with the R^2 of the estimated model (Column (6)).

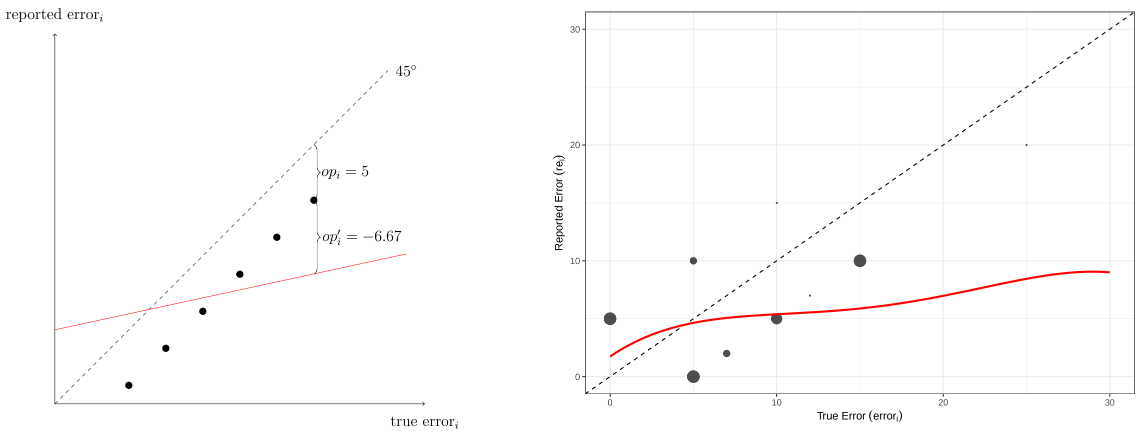
C Alternative Measures of Overprecision

To test the robustness of our overprecision measure, in Section C.1 we discuss four alternative measures of overprecision. In Section C.2 we use these alternative measures to test the robustness of our results from Section 3 regarding the socio-demographic characteristics and in Section C.3 the robustness of the predictions in Section 4.2.

C.1 Alternative Measures

Residual measure (op'_i): The first measure we construct, the *residual* measure, is a measure of overprecision similar to 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 the confidence of participants 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 error and computing quintiles. We then regress the respondents' confidence about the answer on a fourth-order polynomial of the true 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 reported error and the fitted fourth-order polynomial instead of the distance between the reported error and the *true error*. This approach comes with the caveat that, when computing the residuals for the seven questions, the measure classifies respondents as underconfident even when their true error is larger than their reported error, if their deviation is relatively smaller than that of the other respondents in the population (for an illustrative example see Figure C.1). Thus, for every measure of op'_i, the residual measure takes into account the relationship between the reported error and the true error for the entire *population* of respondents. In contrast, our approach focuses on the signal processing of the respondent only by comparing the true error to the reported error.



(a) Illustration **(b)** Data

Figure C.1: Misspecification of participants. This figure shows the difference between the Reported 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 Reported Error Method yields $op_i = error_i - re_i = 3$ 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 ($re_i = \alpha + \beta error_i + \epsilon_i$). In panel (b), the dots represent respondents for whom the Reported Error Method yields an overprecision of 3 and -3 respectively. The red line indicates the fitted line of the *residual* approach using a fourth order polynomial.

Relative measure (op''_i): To circumvent the above mentioned classification problem of the residual approach (op'_i), but still taking into account the relative distance between reported error and realized error, we compute a *relative* measure op''_i by dividing the absolute measure op_i in a specific question with the respective reported error re. Taking the relative distance into account has the advantage to make the measure more comparable across subjects as sketched in Figure C.2.

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

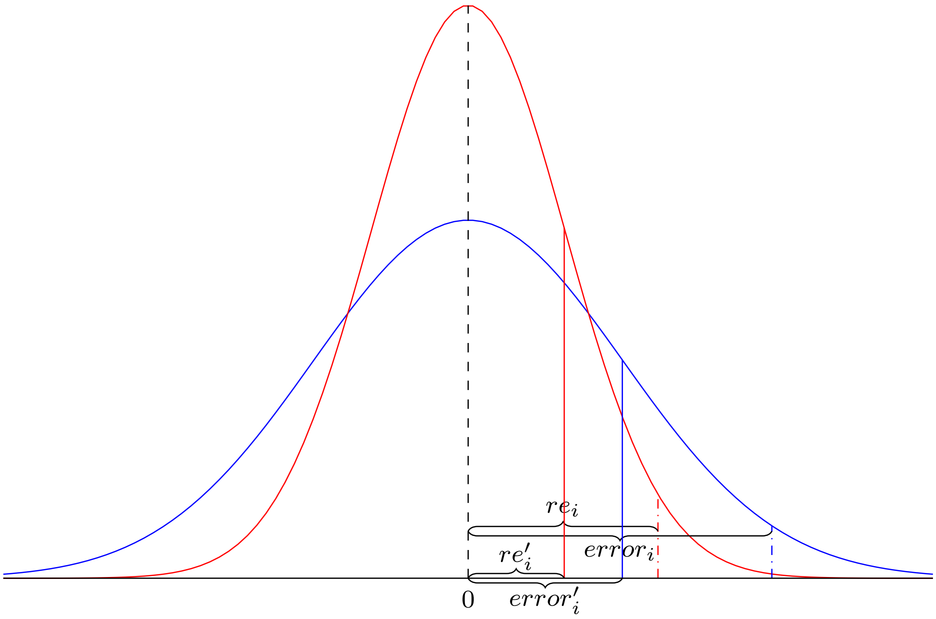


Figure C.2: This figure shows two distributions of the error. The blue curve shows the true distribution of the error with a standard deviation of 2 (precision of .25). The red curve shows the perceived distribution by an overprecise respondent with a standard deviation of 1.25 (precision of .64). The solid and dashed vertical lines indicate the reported errors re and the true errors $error$ resulting from respondents with two different ideas about the nature of the expected absolute error asked in the second question.

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.2, 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 (re and $error$) the absolute overprecision measure would yield 1.47 whereas for the respondent with having 68% confidence in mind (re' and $error'$) it would yield .75. Thus, the second respondent would incorrectly be classified as less overprecise.

The relative measure corrects for this inconsistency by scaling the absolute overprecision measure with the reported expected absolute error and, thus, makes the measure comparable across subjects. In the above example, the relative measure yields .6 in both cases which is exactly 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-question ranges from $\rho^{Spearman} = .91$ to $\rho^{Spearman} = .96$ which is consistent with the respondents having about the same idea on the question about the expected absolute error.[24] Therefore, using the absolute measure seems to be valid given the obvious drawback of this procedure that we lose observations of respondents whose reported error equals zero and, thus, that the sample further decreases when using this alternative measure. We again use the mean to aggregate across the seven questions.

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.

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 that took place within the last 100 years, respondents who lived during these events might be better calibrated. This becomes obvious in Figure C.3 where, for every question, we divide 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 in their answers than those born after the event. As a robustness test, we construct, for every respondent, an *age-robust* measure of overconfidence (op_i''''). This measure is constructed following the formulation described in Section 2.1, but using only the questions that ask about events that happened *after* the respondent was born. So, for example, the measure op_i'''' for a person born in 1992 would be composed using only the questions on events posterior to this year, which means using only three of the seven possible answers. The drawback of this approach is that we lose a substantial amount of information and give more weight to events that took place later in time. Taking fewer questions into account also comes at the risk that the aggregate measure is biased by one specific question.

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 not only

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

capture overprecision but also a miscalibration of the mean. To disentangle overprecision from such miscalibration, we construct a *centered* measure of overprecision. For each question, we subtract the sample mean of the true error and the reported error from the true error and reported error respectively before calculating our measure of overprecision. Hence, we correct for the difference in the means of the distributions and center the distribution of both metrics around zero. Any remaining systematic deviation of the reported error from the true error should be exclusively due to over- or underprecision.

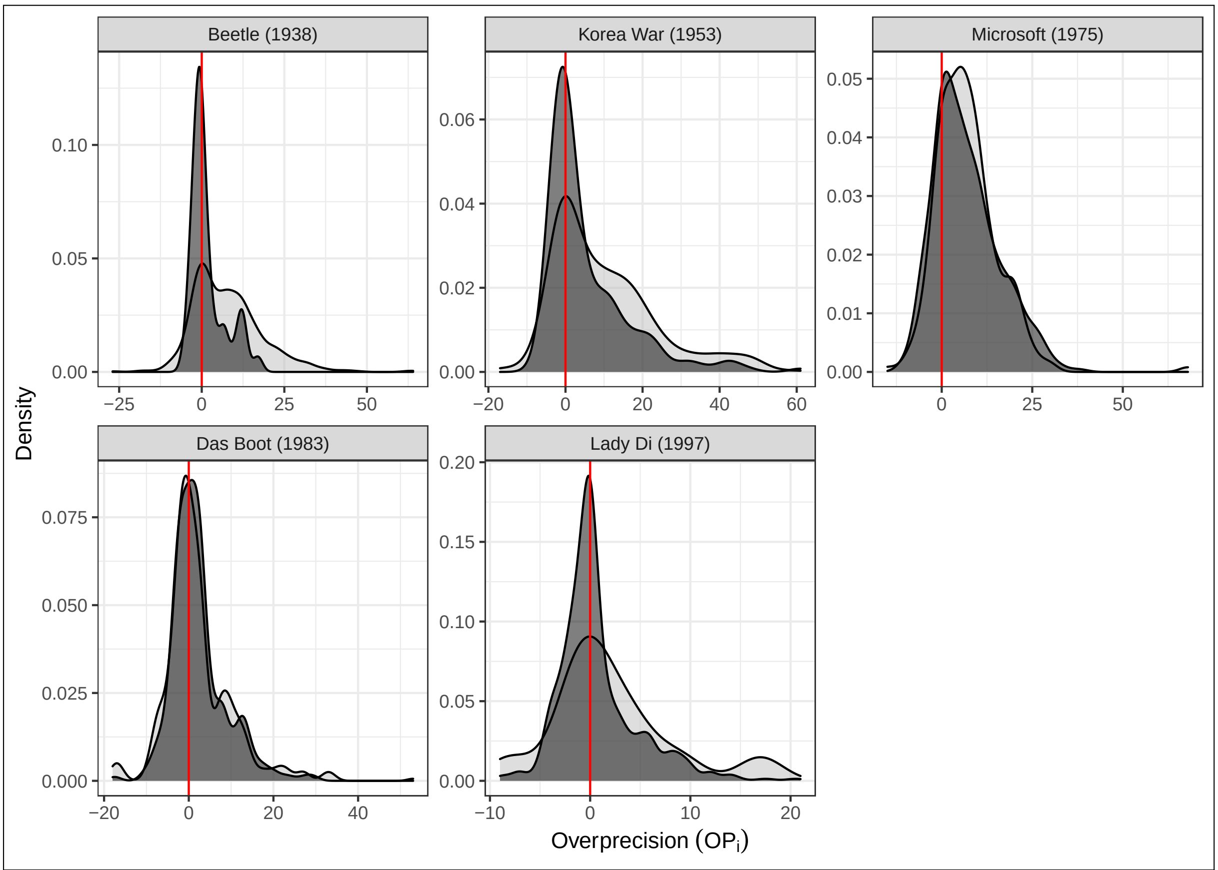


Figure C.3: Density of Overprecision (OP_{ij}) and Age. From left (less recent) to right (more recent) We plot the density of the measured overprecision ($op_{i,j}$) for each question j. In darker color, we plot the density of all respondents born at the year of the event or before. In lighter color, 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. Note that the scale of the vertical axis is different across the five plots. Questions with (correct) answers after 2000 are not included as there were no underage respondents.

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 *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.

Column (3) replicates the baseline estimations using the *relative* approach (op_i''), while Column (4) shows the results for the *standardized* measure (op_i'''). The results in both columns 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 (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) shows the results using the *centered* measure (op_i'''''). The results remain at large robust with the coefficient of gender becoming larger and the coefficient of answered turning insignificant.

Given the results in 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 when we test the predictive power of all of the robustness measures.

	Baseline (1)	Residual (2)	Relative (3)	Standardized (4)	Age robust (5)	Centered (6)
age	-0.007**	0.006**	0.002	-0.002	0.018***	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
gender	0.082	-0.204***	0.056	0.067	0.023	0.150**
	(0.072)	(0.071)	(0.080)	(0.073)	(0.071)	(0.072)
pgbilzt	-0.044***	-0.002	-0.015	-0.020	-0.015	-0.044***
	(0.014)	(0.014)	(0.016)	(0.014)	(0.014)	(0.014)
answered	0.070***	-0.107***	0.016	-0.042**	0.065***	-0.023
	(0.021)	(0.020)	(0.026)	(0.021)	(0.021)	(0.021)
pglabgro	-0.051**	0.002	-0.035	-0.051**	-0.041*	-0.045*
	(0.023)	(0.023)	(0.025)	(0.023)	(0.023)	(0.023)
mislabgro	-0.083	0.134	-0.139	-0.187	-0.193	-0.075
	(0.208)	(0.204)	(0.237)	(0.210)	(0.204)	(0.207)
_cons	0.502	0.361	0.085	0.822**	-0.620*	0.964**
	(0.377)	(0.370)	(0.426)	(0.381)	(0.371)	(0.376)
N	805	805	702	805	801	805
adj. R^2	0.083	0.117	0.028	0.060	0.123	0.088
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Employment Status Dummy	Yes	Yes	Yes	Yes	Yes	Yes

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 *residual* measure, the *relative* measure, the *standardized* measure, the *age-robust* measure, and the *centered measure* respectively. All include dummies for the labor force status (employed, unemployed, retired, maternity leave, non-working), and whether the respondent was a citizen of the GDR before 1989. 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*, and the *age-robust* measure of overprecision. Table C.2 shows the results from the predictions using the residual approach. As 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.

Table C.3 shows the results from the predictions using the *relative* measure of overprecision instead. The advantage is that it makes the measure more comparable across

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	0.014	0.977	0.977	17/38	no/15	0.14	578
opt_dax	-0.014	0.681	0.997	10/38	no/11	0.39	578
err_rent	-0.048	0.781	0.998	19/38	no/12	0.06	670
err_buy	-0.161	0.244	0.893	3/38	no/0	0.00	644
B Diversification:							
std_divers	-0.028	0.450	0.972	13/38	no/14	0.12	774
C Ideological Positioning:							
std_extreme	-0.044	0.274	0.894	8/39	no/9	0.04	716
std_lr	-0.009	0.811	0.964	17/39	no/11	0.07	716
D Voting Behavior:							
non_voter	-0.003	0.807	0.993	18/39	no/17	0.13	706

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

Table C.2: This table shows the estimation results of Section 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 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 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 4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

subjects. However, we lose those observations with a zero reported error due to mathematical reasons. The results, as compared to those in the baseline in Table 2, remain qualitatively similar.

Table C.4 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.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

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	1.072**	0.040	0.217	4/38	yes/20	0.17	530
opt_dax	0.112***	0.002	0.016	2/38	yes/12	0.38	530
err_rent	0.316*	0.091	0.317	2/38	yes/16	0.10	608
err_buy	-0.159	0.290	0.642	7/38	no/0	0.00	644
B Diversification:							
std_divers	-0.068*	0.078	0.334	24/38	yes/13	0.11	681
C Ideological Positioning:							
std_extreme	0.116***	0.005	0.034	2/39	yes/12	0.05	624
std_lr	-0.035	0.394	0.633	19/39	no/7	0.05	716
D Voting Behavior:							
non_voter	0.003	0.796	0.796	3/39	no/17	0.13	706

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

Table C.3: This table shows the estimation results of Section 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 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 as specified in Section 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 4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

to the LASSO estimations. The point estimates slightly decrease in size and significance. However, as pointed out above, this measure takes into account 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 *centered* measure of overprecision. Since the correlation between the centered and the baseline measure is .96, the results remain mostly unchanged.

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	0.782	0.112	0.378	8/38	yes/15	0.15	578
opt_dax	0.129***	0.000	0.000	2/38	yes/9	0.40	578
err_rent	0.253	0.137	0.357	2/38	yes/13	0.07	670
err_buy	0.124	0.369	0.602	15/38	no/0	0.00	644
B Diversification:							
std_divers	-0.102***	0.005	0.034	3/38	yes/19	0.13	774
C Ideological Positioning:							
std_extreme	0.093**	0.020	0.114	3/39	yes/12	0.05	716
std_lr	-0.020	0.610	0.610	17/39	no/11	0.07	716
D Voting Behavior:							
non_voter	0.026**	0.025	0.119	4/39	yes/19	0.14	706

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

Table C.4: This table shows the estimation results of Section 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 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 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 4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	-0.071	0.885	0.885	40/38	no/15	0.14	578
opt_dax	0.114***	0.001	0.008	2/38	yes/10	0.39	576
err_rent	0.132	0.434	0.819	7/38	yes/14	0.07	668
err_buy	-0.065	0.631	0.864	8/38	no/0	0.00	644
B Diversification:							
std_divers	-0.051	0.170	0.525	6/38	yes/15	0.12	770
C Ideological Positioning:							
std_extreme	0.067*	0.090	0.432	5/39	no/9	0.04	713
std_lr	-0.063	0.107	0.432	6/39	yes/12	0.07	713
D Voting Behavior:							
non_voter	0.023**	0.039	0.243	4/39	yes/19	0.14	702

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

Table C.5: This table shows the estimation results of Section 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 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 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 4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).

	(1) Point estimate	(2) Unadj. p-value	(3) SH p-value	(4) R^2 rank	(5) LASSO included	(6) R^2	(7) N
A Prediction error:							
err_dax	1.004**	0.047	0.175	2/38	yes/21	0.17	578
opt_dax	0.098***	0.005	0.034	2/38	yes/11	0.39	578
err_rent	0.327**	0.066	0.185	2/38	yes/16	0.08	670
err_buy	0.136	0.341	0.566	12/38	no/0	0.00	644
B Diversification:							
std_divers	-0.123***	0.001	0.008	3/38	yes/24	0.14	774
C Ideological Positioning:							
std_extreme	0.085**	0.047	0.175	5/39	yes/14	0.06	716
std_lr	0.000	0.997	0.997	17/39	no/13	0.08	716
D Voting Behavior:							
non_voter	0.026**	0.033	0.182	4/39	yes/17	0.13	706

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

Table C.6: This table shows the estimation results of Section 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 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 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 4.1 along with the number of control variables chosen by LASSO and the R^2 of the estimated model (Column (6)).