Incentives, Search Engines, and the Elicitation of Subjective Beliefs: Evidence From Representative Online Survey Experiments

Elisabeth Grewenig (ifo Institute)
Philipp Lergetporer (ifo Institute)
Katharina Werner (ifo Institute)
Ludger Woessmann (ifo Institute and LMU Munich)

Discussion Paper No. 146

March 11, 2019

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Incentives, Search Engines, and the Elicitation of Subjective Beliefs: Evidence from Representative Online Survey Experiments*

Elisabeth Grewenig, Philipp Lergetporer, Katharina Werner, and Ludger Woessmann†

Abstract

A large literature studies subjective beliefs about economic facts using unincentivized survey questions. We devise randomized experiments in a representative online survey to investigate whether incentivizing belief accuracy affects stated beliefs about average earnings by professional degree and average public school spending. Incentive provision does not impact earnings beliefs, but improves school-spending beliefs. Response patterns suggest that the latter effect likely reflects increased online-search activity. Consistently, an experiment that just encourages search-engine usage produces very similar results. Another experiment provides no evidence of experimenter-demand effects. Overall, results suggest that incentive provision does not reduce bias in our survey-based belief measures.

Keywords: beliefs, incentives, online search, survey experiment

JEL classification: D83, C83, C90

March 9, 2018

* For helpful comments and discussion, we would like to thank Lucas Coffman, Martin Kocher, Stefanie Stantcheva, and seminar participants at the World Meeting of the Economic Science Association in Berlin, the European Economic Association in Cologne, the Workshop on Subjective Expectations and Probabilities in Economics in Munich, and the CRC retreat in Tutzing. We are most grateful to Franziska Kugler and Sarah Kersten for their help in preparing the surveys. Financial support by the German Science Foundation (CRC TRR 190) and the Leibniz Competition (SAW-2014-ifo-2) is gratefully acknowledged.

† ifo Institute at the University of Munich; Emails: grewenig@ifo.de, lergetporer@ifo.de, werner.k@ifo.de, woessmann@ifo.de.

1. Introduction

Measuring people's subjective beliefs about economic facts is essential for understanding economic behavior and choices. For example, beliefs about earnings returns to education shape educational decisions (e.g., Boneva and Rauh (2017); Bleemer and Zafar (2018); Delavande and Zafar (2019)), beliefs about the size of government affect support for public spending (e.g., Lergetporer et al. (2018a); Roth and Wohlfart (2018b)), and beliefs about societal inequality are closely linked to preferences for redistribution (e.g., Kuziemko et al. (2015); Alesina et al. (2018b)).[1] To elicit such beliefs, economists often rely on survey questions which do not provide respondents with incentives for accurate answers. This raises concerns of systematic biases in unincentivized belief measures that might stem from lack of cognitive effort invested in truthful reporting or from socially desirable and self-serving answering behavior. Biased belief measures are not only uninformative about the target population's actual beliefs, but they can also bias estimates of the investigated relationship between subjective beliefs and economic decisions (e.g., Bertrand and Mullainathan (2001); Benitez-Silva et al. (2004)). While the effects of incentivizing belief accuracy have been studied extensively in the context of economic games conducted in the laboratory (e.g., Trautmann and de Kuilen (2014); Schlag et al. (2015)), little is known about incentive effects on survey-based belief measures about economic facts.

In this paper, we investigate experimentally whether incentive provision is necessary to retain unbiased measures of subjective beliefs in representative online surveys. Our two main experiments provide incentives for correct beliefs about two relevant economic facts. Results show that incentivization improves the accuracy of stated beliefs in one case but not in the other. Inspection of response spikes indicates that incentive effects may be related to the usage of online-search engines. Therefore, we conduct a third experiment that explicitly encourages online-search activity without providing incentives. The encouragement treatment produces very similar effects to the incentive treatment, suggesting that improved beliefs in the incentive treatment mainly reflect altered online-search behavior. In a fourth experiment, we show that inducing experimenter demand does not affect stated beliefs, suggesting that the incentive effect is unlikely to reflect experimenter-demand effects. Overall, our results support the validity of

[1] Other economically important dimensions of people's subjective beliefs include, for instance, beliefs about inflation (e.g., Armantier et al. (2013); Armantier et al. (2016)) and beliefs about the extent and consequences of immigration (e.g., Grigorieff et al. (2016); Haaland and Roth (2017); Alesina et al. (2018a)).

unincentivized belief elicitation in surveys and in fact warn about potential unintended consequences of belief incentivization in online surveys.

We conduct our two incentive experiments in a representative online survey in Germany (N>3,600). The elicited beliefs about economic facts refer to respondents' stated beliefs about (i) average earnings by professional degree and (ii) average public school spending.[2] In both experiments, we offer respondents in a randomly selected treatment group a monetary reward if their answer to the belief question is close to the true value. The control group answers the same questions without any incentives. If unincentivized belief questions are in fact biased (e.g., by lack of effort or self-serving answering behavior), we expect incentives to improve the accuracy of stated beliefs.

The first experiment indicates that incentivizing beliefs hardly affects stated beliefs about earnings by professional degree. We elicit beliefs about net average monthly earnings of full-time employed persons without any professional degree and of persons with a university degree. In the unincentivized control group, the 10-90 percentile range of earnings beliefs without a degree is 800 Euro to 1,500 Euro (true value: 1,400 Euro). For earnings beliefs about university graduates, the range is 1,500 Euro to 4,000 Euro (true value: 2,750 Euro). The majority of respondents underestimate current earnings levels: 82 percent of unincentivized beliefs about earnings without a degree and 57 percent of those about earnings with a university degree are below the true value. We find almost no incentive effects on earnings beliefs: While incentives marginally increase beliefs about average earnings with a university degree ($p<0.1$), they do not affect measures of belief accuracy or respondents' confidence about their beliefs. Although incentivized respondents take more time to answer the belief question, this does not translate into increased belief accuracy. Consistently, we also find no incentive effects on beliefs re-elicited in a follow-up survey about two weeks after the main survey.

The second experiment shows that, by contrast, the incentive treatment significantly improves the accuracy of stated beliefs about average annual public school spending per student. In the control group, school-spending beliefs are strongly dispersed with a 10-90 percentile range of 50 Euro to 15,000 Euro (true value: 7,400 Euro). The majority of

[2] Beliefs about earnings by professional degree are central to the economics literature on educational decisions (e.g., Arcidiacono (2004); Arcidiacono et al. (2012); Hoxby and Turner (2013); Kaufmann (2014); Belfield et al. (2019)). While this strand of research is often particularly interested in people's individual earnings expectations (which often cannot be verified and therefore not incentivized; see Manski (2004)), beliefs about population averages, the focus of this paper, have also been featured prominently in the literature (e.g., Wiswall and Zafar (2015); Bleemer and Zafar (2018); Lergetporer et al. (2018b)). Beliefs about public education spending are important for shaping public budgets since they determine the electorate's preferences for the size of government (e.g., Schueler and West (2016); Lergetporer et al. (2018a)).

unincentivized respondents (78 percent) underestimates the actual spending level. In contrast to the experiment on earnings beliefs, incentivization improves the accuracy of school-spending beliefs in the main survey. It also raises respondents' confidence about the accuracy of their beliefs. These effects persist in the follow-up survey about two weeks after the main survey. Again, incentive provision increases response time in the main survey.

To reconcile the different effects of incentives in the two experiments, we present further analyses on potential treatment-effect mechanisms. Closer inspection of the distribution of beliefs on public school spending shows that incentivizing belief accuracy increases the frequency of stated beliefs which correspond to top-listed online-search-engine results. Since information on average public school spending (which is published regularly by the German Statistical Office) is easier to find online than information on average earnings by professional degree (which we calculated ourselves for the purposes of this study), we suggest that differences in incentive effects across belief domains might be due to differences in the use of online-search engines.

To further investigate the role of search-engine usage, we conduct an additional experiment in a new representative sample (N>4,000). Instead of providing monetary incentives, in this experiment we encourage a randomly selected group of respondents to use online-search engines to inform their beliefs about public school spending. Strikingly, this encouragement produces treatment effects which are very similar in magnitude and significance to the effects of incentivizing belief accuracy. This similarity strongly suggests that our incentives for accurate beliefs about public school spending improve beliefs through altering respondents' online-search activity.

The investigation yields two main conclusions. First, the fact that monetary incentives do not improve belief accuracy about average earnings (i.e., information not easily found on the internet) suggests that the lack of incentives in commonly used survey-based belief measures does not yield reporting bias. Second, incentivizing belief accuracy in online surveys seems to induce respondents to use online-search engines for the purpose of improving their answers. While incentive effects on reported beliefs have been studied extensively in closely controlled laboratory environments, this result highlights potential limitations of applying incentives to less-controlled contexts such as online surveys. Since researchers are usually interested in unbiased measures of prior subjective beliefs about economic facts and not in beliefs updated after consulting external sources such as online-search engines, our results – taken together – underline the validity of using unincentivized belief measures. At the same time, using

monetary incentives or encouraging online-search activity when eliciting beliefs might be interesting for researchers who seek to shift survey respondents' information sets.

Finally, we present results of an additional experiment to assess whether the effects of the incentive treatment reflect experimenter-demand effects, as opposed to genuine incentive effects.[3] In a randomly selected treatment group, we emphasize that it is important for us that respondents provide a correct answer to the belief question about public school spending. While this wording explicitly states the experimenter demand, it does not affect stated beliefs or response time. This result suggests that the reported incentive effects are due to respondents adapting their answering behavior to earn the incentives, and not because the offer of incentives contains information about the importance of accurate beliefs to the experimenter.

Our paper contributes to several strands of the existing literature. At the most basic level, it adds to the large literature in economics which studies people's subjective beliefs about economic outcomes in large-scale surveys. One strand of this literature focuses on beliefs about future events, such as inflation expectations (e.g., Coibion and Gorodnichenko (2012); Armantier et al. (2016); Cavallo et al. (2017); Manski (2018)), future home prices (e.g., Fuster et al. (2018)), or educational expectations (e.g., Zafar (2011); Stinebrickner and Stinebrickner (2014); Boneva and Rauh (2017, 2018); Attanasio et al. (2018); Delavande and Zafar (2019)). Another strand studies beliefs about realized (contemporaneous or past) economic outcomes, such as beliefs about existing inequality (e.g., Alesina and La Ferrera (2005); Karadja et al. (2017); Alesina et al. (2018b); Lergetporer et al. (2018c)), immigration (e.g., Grigorieff et al. (2016); Haaland and Roth (2017); Alesina et al. (2018a)), or the size of government (e.g., Lergetporer et al. (2018a); Roth and Wohlfart (2018b)).[4] This literature mostly relies on unincentivized questions, often implemented in online surveys, to elicit respondents' beliefs. While some recent papers incentivize correct answers to belief questions (e.g., Grigorieff et al. (2016); Fuster et al. (2018); Roth and Wohlfart (2018a)), the effects of such incentives on stated beliefs have hardly been studied experimentally. Our first contribution is therefore to provide

[3] In our context, one concern might be that providing incentives for accurate beliefs signals to respondents that the surveyor's aim is to show that incentives in fact improve beliefs. Consequently, respondents might put more effort in providing a correct answer only because they want to please the experimenter. Similarly, the incentive treatment may induce respondents to state second-order beliefs about what they think the surveyor would want them to answer rather than their own true belief. Our experimental design to measure experimenter-demand effects is similar to de Quidt et al. (2018) and Mummolo and Peterson (2018).

[4] In addition, several studies investigate the connection between beliefs about future events and realized outcomes, e.g., by shifting beliefs about the future through providing information about contemporaneous (or past) outcomes (e.g., Amartier et al. (2013); Wiswall and Zafar (2015); Armona et al. (2016); Coibion et al. (2018)). Relatedly, Manski (2004) provides an exhaustive discussion of eliciting subjective expectations in surveys.

an in-depth assessment of the effects of incentivizing belief accuracy in a large-scale, representative online survey.

Relatedly, our paper is part of the smaller literature on the effectiveness of different belief-elicitation methods (e.g., Delavande et al. (2011); Ansolabehere et al. (2013)). Closest to our paper, Bullock et al. (2015) and Prior et al. (2015) show that monetary incentives for accurate answers can reduce partisan bias in stated beliefs about political facts.[5] We complement the existing literature by providing an extensive analysis of incentive effects, potential underlying mechanisms, and the role of experimenter-demand effects. In particular, our paper is, to the best of our knowledge, the first to study respondents' online-search activity in relation to incentive provision.

Finally, our paper contributes to the methodological literature on the opportunities and challenges of conducting experiments on the internet. Comparing experimental choices elicited in the laboratory and online, recent studies suggest that differences across these modes are modest (e.g., Clifford and Jerit (2014); Arechar et al. (2018)). Our finding that online-survey participants likely use search engines to improve stated beliefs points to thus far underappreciated challenges when conducting experiments in environments where the experimenter has limited control over the experimental setting.

The paper is structured as follows. Section 2 introduces our survey and the experimental setup of the incentive experiments. Section 3 shows the experimental results on belief incentivization. Section 4 presents the experiment on encouraging online-search activity. Section 5 provides experimental evidence on the role of experimenter-demand effects. Section 6 concludes.

2. Data and Empirical Strategy

This section describes the survey, the experimental design, and the econometric model and reports results of randomization tests.

2.1 The Survey

Our research is based on data from the ifo Education Survey, an annual representative opinion survey on education policy that we conduct in Germany. The main experiments of this paper were implemented in the 2017 survey wave and focus on 3,696 respondents who were

[5] Relatedly, there is an extensive literature using laboratory experiments to study effects of incentives on university-student participants' beliefs in economic games (see, for instance, Blanco et al. (2010); Gächter and Renner (2010); Wang (2011); Trautmann and de Kuilen (2015); Schlag et al. (2015)).

sampled and surveyed via an online platform.[6] The sample is weighted to match official statistics with respect to age, gender, federal state, school degree, and municipality size. The survey contains a total of 34 questions on different topics of education policy and collects information on respondents' sociodemographic characteristics (see Table 1). Median completion time is 17 minutes. Item non-response is very low, at less than one percent for the belief questions which measure our outcomes of interest. As we show below, treatment status does not predict item non-response on the outcome variables.

Sampling and polling was carried out by Kantar Public, a renowned survey company, in April and May 2017. As is standard for online surveys, respondents answer the survey autonomously on their own digital devices. For all respondents, survey completion is incentivized with 75 tokens (worth about 0.70 Euro), which they can exchange for gift vouchers of well-known retailers.

To investigate the persistence of potential treatment effects beyond the immediate survey horizon, respondents were invited to participate in a follow-up survey about two weeks after the main survey. The follow-up survey re-elicits some outcomes, but does not contain any experimental treatments and does not incentivize correct answers. Overall, 2,535 respondents (69 percent) participated in the follow-up survey.[7] The median lag to the main survey was 12 days with a range from 5 to 41 days.

2.2 Experimental Design

To test whether monetary incentives affect stated beliefs about economic facts, we devise two experiments that incentivize belief accuracy in randomly selected treatment groups. One experiment focuses on beliefs about average earnings by professional degree, the other on beliefs about average public school spending. All respondents participated in both experiments consecutively, and randomization was independent across experiments.[8]

[6] The overall sample comprised 382 additional respondents who do not use the internet and were therefore polled at their homes by trained interviewers. As it was not possible to incentivize their answers, we exclude these respondents from our analysis. Grewenig et al. (2018) show that our approach to weight online-survey observations to match the characteristics of the entire population yields representativeness for the entire population. Our results are qualitatively identical when using unweighted data (results available upon request).

[7] The take-up rate is relatively high compared to other recent studies. For instance, take-up in follow-up surveys was 14 percent in Kuziemko et al. (2015) and 24 percent in Alesina et al. (2018b).

[8] In the questionnaire, the experiment on school-spending beliefs preceded the experiment on earnings beliefs. One potential concern with running both experiments with all respondents is that incentive provision in the first experiment might affect answering behavior in the subsequent experiment. Reassuringly, treatment effects in the earnings experiment are robust to restricting the sample to those respondents who were assigned to the control group in the school-spending experiment. In our analysis, all regressions for the second experiment (beliefs on earnings) control for treatment status in the first experiment (beliefs on school spending), but results are qualitatively unchanged if the control is excluded (results available upon request).

2.2.1 The Incentive Treatment

In both experiments, we randomly assigned respondents to a treatment group that is offered additional tokens for accurate answers before eliciting their beliefs in the same way as in the unincentivized control group. In the incentive treatment, respondents can earn 35 tokens – in addition to the 75 tokens for participation – if their answer is "roughly correct". We define answers as "roughly correct" if they fall into the intervals around the true values that contain 50 percent of respondents. To calculate these intervals, we used data from the 2016 wave of the ifo Education Survey where we asked the same (unincentivized) belief questions.[9]

Our incentive scheme rewards answers within an interval around the true value with a fixed amount of tokens. Laboratory experiments with university students usually use more sophisticated methods to incentivize beliefs, such as the quadratic scoring rule (e.g., Trautmann and Kuilen (2015)). Since our representative sample comprises respondents from various educational backgrounds, including 36 percent with basic school degree or less (see Table 1), we deliberately opted for a much simpler incentive scheme to minimize the possibility of comprehension problems.[10]

In the incentive treatment, we use the following wording to inform about the possibility to earn additional tokens for accurate answers: *"On the following screen, you will be asked another guess question. If your guess is roughly correct, you will receive an additional reward worth about half of the reward you get for participating in this survey."*[11] To maximize salience of the treatment, treated respondents are also reminded of the incentives on the subsequent belief-elicitation screen. We kept the information about the incentive short to convey the idea that belief accuracy is rewarded as clearly and simply as possible. At the same time, treated respondents had the possibility to retrieve more information about the incentive scheme by clicking on an information button, which 28 percent (32 percent) chose to do in the earnings experiment (school-spending experiment).[12] Appendix Table A2 shows that older respondents

[9] While the ifo Education Survey is a repeated cross-sectional survey, some respondents participate in more than one survey wave. About 13 percent of respondents in the 2017 survey wave also participated in the 2016 survey wave. Excluding these respondents from the analysis yields virtually identical results (available upon request).

[10] The extent to which scoring rules make truthful reporting optimal usually depends on respondents' risk preferences (e.g., Trautmann and Kuilen (2015)). Interestingly, our heterogeneity analysis in section 3 reveals that incentive-treatment effects are not heterogenous with respect to respondents' general willingness to take risk.

[11] Appendix Table A1 presents the wording of all questionnaire items used in this paper.

[12] The following text appeared upon clicking on the information button: "*You will receive an additional reward worth about half of the reward you get for participating in this survey if your answer lies within the best 50 percent of answers from the previous year.*"

are more likely to acquire this information, and that information acquisition varies with parental education and employment status.

Our data suggest that the incentive treatments successfully increased respondents' effort to give a correct answer. This is not only apparent from the positive treatment effects on response time (see section 3), but also from descriptive survey evidence. After the experiments, we asked respondents in the treatment group whether the incentive increased their effort to give a correct answer. As Figure 1 shows, 58 percent "strongly" or "somewhat" agree with the statement that the prospect of earning more tokens encouraged them to put more effort in their answers. Appendix Table A3 depicts the relationship between these survey answers and respondents' sociodemographic characteristics. Younger respondents, males, those living together with a partner, those who do not have children, those without university-educated parents, and more risk-tolerant respondents are more likely to report that they are motivated by the incentives. Interestingly, respondents who acquired additional information about the scoring rule and those who were randomized into the incentive treatment in both (rather than just one) experiments do not state higher levels of incentive-induced motivation (result available upon request).

2.2.2 Eliciting Beliefs

In the earnings-beliefs experiment, we use the following question to elicit beliefs: "*Persons with a professional degree (apprenticeship) currently earn on average 1,850 Euro net per month (full-time position). What is your best guess, how much do the following groups with lower resp. higher educational attainment earn on average?*" Respondents are asked to enter their earnings beliefs about persons without a professional degree and about persons with a university degree into open numeric fields. Based on calculations using the German Microcensus, the median net monthly earnings of full-time employed persons in these groups are 1,400 Euro and 2,750 Euro, respectively.[13] Following our above definition of "roughly correct" answers, respondents in the treatment group whose stated earnings belief for those without a professional degree was between 1,008 Euro and 1,792 Euro *and* whose earnings belief for those with a university degree was between 1,980 Euro and 3,520 Euro received the incentive payment. This applied to 46 percent of respondents in the treatment group.

The experiment on beliefs about average public school spending follows the same structure. The question reads as follows: "*Based on your best guess, what is the average amount of money spent each year for a child in public schools in Germany?*" According to official

[13] Research Data Centres of the Federal Statistical Office and the statistical offices of the Länder, Microcensus, census year 2012.

statistics of the Federal Statistical Office, the current spending level is 7,400 Euro (Statistisches Bundesamt (2017)). 50 percent of respondents in the treatment group state a belief within the correct range (between 1,332 Euro and 13,468 Euro) and are therefore paid the incentive.

After each belief question, we elicit respondents' confidence that their stated belief is close to correct on a seven-point scale (from 1 = "very unsure" to 7 = "very sure"). To avoid potential treatment-effect spillovers across experiments, respondents are informed about whether their respective belief falls into the incentivized range only at the very end of the survey.

2.3 Econometric Model

We estimate the effects of the incentive treatment on stated beliefs using the following regression model:

$$y_i = \alpha + \beta_1 Incentive_i + \varepsilon_i \quad (1)$$

where y_i is the outcome variable of interest, $Incentive_i$ is a dummy variable indicating whether respondent i received the incentive treatment, and ε_i is the error term. Randomization ensures that the average treatment effect, captured by coefficient β_1, is identified.

As expected, adding the rich set of sociodemographic characteristics (as shown in Table 1) as control variables to the analyses does not substantively alter the estimated treatment coefficients. For ease of exposition, we therefore only report models without covariates throughout the paper. Detailed results of models with control variables are available upon request.

Throughout, we report treatment effects on seven different outcome variables. To test whether the incentive treatment affects average values of stated beliefs, we report effects on (1) respondents' beliefs relative to the correct value. We construct four different variables to measure our main outcome of interest, the accuracy of stated beliefs: (2) the absolute distance between the belief and the true value; as well as indicators of whether (3) the belief is in the incentivized range; (4) the belief is in the 10-percent interval around the true value; and (5) the belief takes an implausibly low value of below 100.

Apart from these belief measures, we are interested in treatment effects on (6) respondents' confidence about their beliefs and on (7) their response time. Since the instruction text on the screens is longer for the treatment group than for the control group, we apply the following two-step procedure to approximate response time. First, we regress response time in all other questions of the ifo Education Survey on a function of question characteristics and individual

fixed effects.[14] We use this model to predict each individual's expected response times for both belief-elicitation questions in the absence of incentive provision. Second, we take the difference between actual and predicted response time, interpreting this difference as the time the respondent invests in belief formation.

To analyze whether the effects of providing incentives are heterogeneous across different subgroups of respondents, we extend our basic regression model to:

$$y_i = \alpha + \beta_1 Incentive_i + \beta_2 Subgroup_i + \beta_3 Incentive_i \times Subgroup_i + \varepsilon_i \qquad (2)$$

where $Subgroup_i$ equals one if respondent i belongs to the respective subgroup and zero otherwise. Accordingly, β_1 indicates the incentive-treatment effect for the omitted baseline group of respondents and β_3 measures the additional incentive effect for the subgroup of interest.

2.4 Test of Randomization

To test whether the randomization successfully balanced respondents' observable characteristics between treatment and control groups, we run the following regression for each characteristic and both experiments:

$$Characteristic_i = \alpha + \gamma\, Treatment_i + \varepsilon_i \qquad (3)$$

Table 1 reports the coefficients γ from equation (3) for the earnings experiment and the school-spending experiment (columns 3 and 6, respectively), along with the corresponding means of the control and treatment groups (columns 1-2 and 4-5). Reassuringly, only one out of 42 estimated differences turns out statistically significant ($p<0.05$), which would be expected by pure change. Thus, the balancing tests suggest that randomization worked as intended. The tests also indicate that the incentive treatment does not affect item non-response, which is very low (below 0.5 percent) on both belief questions.

3. The Effects of Incentives on Belief Accuracy

This section presents our main results in three steps. We analyze incentive effects first on earnings beliefs and second on school-spending beliefs. Third, we provide an initial exploration of how to reconcile different incentive effects in the two experiments.

[14] Question characteristics include the number of words and characters, the number of screens for question presentation, and the type of question. We run an individual fixed-effects lasso including fourth-order polynomials of words, characters, and their interaction to determine the optimal inputs (details available upon request).

3.1 Beliefs about Earnings by Professional Degree

We start our analysis by investigating whether the incentive treatment affects stated beliefs about earnings by professional degree. Panel A of Figure 2 depicts the distribution of respondents' beliefs about earnings without a professional degree (left panel) and with a university degree (right panel) in the main survey. Green bars and transparent bars represent beliefs in the control group and in the incentive-treatment group, respectively. In general, beliefs are quite dispersed with a 10-90 percentile range of 800 Euro to 1,500 Euro for earnings without a degree (true value: 1,400 Euro) and 1,500 Euro to 4,000 Euro for earnings with a university degree (true value: 2,750 Euro) in the control group. Moreover, the majority of respondents, 82 percent and 57 percent, respectively, underestimate current earnings of those without a degree and those with a university degree. Comparing beliefs between the control and the treatment groups, graphical inspection does not reveal obvious differences in the distributions of beliefs by treatment status.

Table 2 presents treatment-effect estimates of incentive provision based on equation (1). Panel A depicts treatment effects on earnings beliefs without a degree and Panel B on earnings beliefs with a university degree.[15] Results indicate that coefficients on the treatment-group indicator are mostly insignificant, suggesting that the incentive treatment hardly affects earnings beliefs. In particular, the treatment does not affect any of the four measures of belief accuracy: the absolute distance between belief and the true value (column 2), the probability of reporting a belief within the incentivized range (column 3) or within the 10-percent interval around the true value (column 4), or the probability of stating a very low belief (column 5). Similarly, the treatment does not affect respondents' confidence that their belief is close to correct (column 6). While we find a small, marginally significant ($p<0.1$) incentive effect on relative beliefs about university graduates' earnings (see column 1 of panel B), the overall pattern suggests no incentive effects on earnings beliefs. Interestingly, however, the treatment significantly increases the response-time measure by 19.6 seconds. Thus, while incentives seem to induce respondents to spend more time on the question, this fails to translate into higher belief accuracy.

While incentives have no overall effect on the entire population's earnings beliefs, they might improve beliefs in certain subgroups of respondents. To explore this possibility, Appendix Table A4 estimates heterogeneous treatment effects on the accuracy of beliefs about

[15] Since we elicited both earnings beliefs on the same screen, confidence and response time were recorded only once for earnings beliefs in general.

earnings of those without a degree (column 1) and those with a university degree (column 2). This heterogeneity analysis is based on equation (2) and focuses on subgroups defined by the sociodemographic characteristics in Table 1. We find no effect heterogeneities on the accuracy of earnings beliefs for those without a degree. Similarly, treatment effects on the accuracy of earnings beliefs for those with a university degree are largely homogeneous across sociodemographic subgroups. The only exceptions are respondents with a middle school degree (respondents aged over 65), who exhibit a larger (smaller) treatment effect than persons with basic degree or less (persons below age 45). Overall, the heterogeneity analysis shows that the population-wide null effect of incentives on the accuracy of earnings beliefs reflects null effects in different sociodemographic subgroups, rather than opposing subgroup effects that cancel each other out.

To test for persistence of any treatment effects, we examine whether incentive provision affects respondents' stated beliefs in a follow-up survey about two weeks after the main survey. The follow-up survey is designed to evaluate the persistence of potential treatment effects beyond the immediate horizon of the main survey by eliciting beliefs without incentive provision. While participants in the follow-up differ from participants in the representative main survey in some sociodemographic characteristics, it is reassuring that follow-up participation does not relate to treatment status in the main survey (see Appendix Table A5). As a consequence, follow-up survey respondents' characteristics remain well balanced between control and treatment groups (see Appendix Table A6). This mitigates concerns about non-random selection into the follow-up and facilitates identification of persistent incentive effects.

Panel B of Figure 2 and Appendix Table A7 mostly confirm our results from the main survey by showing that incentive provision does not induce noteworthy improvements in the accuracy of stated earnings beliefs in the follow-up survey. Interestingly, the treatment effect on response time turns small and insignificant in the follow-up survey, suggesting that incentive provision in the main survey, which prolongs response time in the main survey, does not affect how much time respondents invest in answering the same question in the follow-up survey where no incentives are provided.

3.2 Beliefs about Public School Spending

The second experiment analyzes incentive effects on beliefs about average annual public school spending per student. Panel A of Figure 3 depicts respondents' beliefs in the main survey, separately for respondents in the unincentivized control group (green bars) and respondents in the incentivized treatment group (transparent bars). Beliefs about current

spending levels are very dispersed, with a 10-90 percentile range of 50 Euro to 15,000 Euro in the control group (true value: 7,400 Euro). Again, respondents tend to underestimate current spending levels on average, with 78 percent stating a belief below the true value. Visual inspection suggests that the treatment group's belief distribution has less density at the very left part of the distribution (representing very low belief values) and more density close to the true value, which suggests that incentive provision improves school-spending beliefs.

Table 3 presents regression results based on equation (1) which confirm the graphical inspection that incentivization improves the accuracy of school-spending beliefs. The incentive treatment strongly and significantly increases the probability of reporting a belief within the incentivized range from 37 percent to 50 percent (column 3) and within the 10-percent interval around the true value from 3 to 7 percent (column 4). Likewise, the treatment significantly decreases the probability of reporting an implausibly low value of below 100 Euro by 6 percentage points (column 5). While treatment-effect estimates on the relative belief (column 1) and on the absolute distance between belief and true value (column 2) are not statistically significant, the signs of the coefficients indicate that beliefs in the incentivized treatment group are closer to the true value. Furthermore, the treatment significantly increases respondents' confidence about their beliefs (column 6) and doubles the time respondents take to answer the question (column 7).

The specification in column 3 of Appendix Table A4 estimates heterogeneous incentive-treatment effects on school-spending beliefs across sociodemographic subgroups. Incentive-treatment effects are significantly larger for respondents with higher educational attainment than for those with lower education, and for those living in West Germany compared to those living in East Germany. They are also significantly smaller for older respondents, those living in large cities, and parents. Given that respondents' risk preferences have been theorized to affect the ability of incentive schemes to foster truthful reporting (e.g., Trautmann and Kuilen (2015)), it is interesting to note that treatment effects do not vary by answers to the general risk question.[16]

Results of the follow-up survey reveal that incentivizing belief accuracy improves school-spending beliefs persistently (Table 4 and panel B of Figure 3).[17] Similar to the immediate effects in the main survey, incentive provision increases the probability of reporting a belief within the incentivized range by 11 percentage points and within the 10-percent interval around

[16] See Falk et al. (2016) for an experimental validation of the general risk question.

[17] Again, Appendix Tables A5 and A6 show that selection into the follow-up survey is random with respect to treatment status in the main survey.

the true value by 6 percentage points in the follow-up survey. Furthermore, the negative incentive effect on the absolute distance between belief and true value is highly significant and even larger in the follow-up survey than in the main survey. In contrast to the main survey, incentive provision does not affect the probability of reporting very low beliefs of below 100 Euro, presumably because the control-group probability for reporting such low values is only 5 percent in the follow-up survey (compared to 16 percent in the main survey).

In the follow-up survey, the treatment group also continues to be more confident about their beliefs than the control group. By contrast, there is no persistent treatment effect on response time. This suggests that incentive provision in the main survey improves beliefs immediately, and respondents remember their improved beliefs when being resurveyed later (as opposed to permanent shifts in the time invested in forming beliefs).

3.3 Exploring Effect Differences: Online-Search Behavior as Potential Channel

Our results thus far show that incentive provision improves the accuracy of stated beliefs about average school spending, but not about earnings by professional degree. Still, incentives significantly increase the time respondents take to state their beliefs in both cases, which is consistent with an increase in respondents' effort to provide a correct answer in both experiments. In this section, we investigate increased use of online-search engines as a potential mechanism for the pattern of results described above.

Scrutiny of the stated school-spending beliefs in the treatment group reveals that the density of treatment-group beliefs spikes at 6,000, 6,300, 6,500, and 6,700 Euro. Quite strikingly, it turns out that these values are among the top results of online-search requests at Google for key phrases of our question wording.[18] Building on this observation, the first column of Table 5 regresses a dummy variable coded one if respondents state one of the online-search-engine results as their belief, and zero otherwise, on an incentive-treatment indicator. The treatment significantly increases the likelihood of stating one of the Google results as beliefs by about 12 percentage points from a control-group mean of 2 percent. That is, the incidence of reporting such values increased from 50 respondents in the control group to 277 respondents in the treatment group. While the probability of stating a belief corresponding to Google search results is certainly an imperfect measure of actual online-search activity, the analysis provides

[18] Appendix Figure A1 shows Google results from search requests for our question wording, namely "*Ausgaben pro Schüler pro Jahr an öffentlichen, allgemeinbildenden Schulen"* and *"Bildungsausgaben pro Schüler pro Jahr"*. We searched for these values shortly after the implementation of the ifo Education Survey 2017 to obtain results close to those which were available to the respondents and found the values 6,000, 6,300, 6,500, 6,700, and 7,400 Euro in the top search results.

suggestive evidence that the incentive effect on school-spending beliefs may operate through respondents' increased use of online-search engines.[19]

There are at least two potential reasons for why online-search activity might only improve school-spending beliefs, but not earnings beliefs. First, beliefs about average earnings by professional degree are less dispersed than school-spending beliefs, which limits the leeway for online-search-engine use to improve beliefs.[20] Second, accurate information on net average monthly earnings by professional degree seems to be relatively hard to find online.[21] Thus, the finding that incentive provision only improves school-spending beliefs is consistent with our proposition that incentive effects mainly operate through increased search-engine usage.

This descriptive analysis suggests that incentivizing belief accuracy in online surveys can have the (potentially unintended) effect that respondents resort to online-search engines to improve their stated beliefs – as opposed to increased effort to retrieve truthful beliefs from memory. To scrutinize this possibility further, the next section presents an additional experiment in which we encourage a randomly selected treatment group to use online-search engines before stating school-spending beliefs.

4. Encouraging Online-Search Activity

The results discussed in the previous section suggest that incentivized respondents may be more likely to search for correct answers online. To better understand how the use of online-search engines changes stated beliefs, we conduct an additional experiment in which we explicitly encourage respondents to search the internet for the correct answer. This experiment allows us to investigate whether the incentive-treatment effect described above can be reproduced by exogenously induced online-search activity. In what follows, we describe the

[19] Although we do not know which values the respondents ultimately found on the internet, we are confident that our approximation of online-search activities works reasonably well: In a subsequent experiment that explicitly encourages respondents to search for the correct value on the internet, the probability that a stated belief corresponds to the top Google-search results is highly correlated (correlation=0.64) with self-reported search engine usage (see section 4.2 below for details).

[20] We remain agnostic about the reasons for why earnings beliefs are relatively less dispersed. One reason might be that own experience makes monthly earnings a more tangible concept than public spending on schools. A complementary reason might be that the survey question that elicits beliefs about earnings of persons without a professional degree and persons with a university degree provides respondents with the anchor of earnings of persons with an apprenticeship degree (see section 2.2.2).

[21] As official statistics on earnings by professional degree are not published, we obtained the values from own calculations based on the German Microcensus. Appendix Figure A2 presents results from Google search requests of key phrases of the earnings-beliefs question. These Google results seem to correspond to minimum-wage earnings or to the earnings of workers in specific occupations rather than the German averages by professional degree. Further analyses show that 3 respondents in the control group and 10 respondents in the treatment group report a belief that corresponds to one of the results listed on the first page of Google results.

experimental setup, report results of the encouragement treatment on online-search activity and on belief accuracy, and compare the results to the effects of incentive provision.

4.1 Experimental Design

We conduct the encouragement experiment in a new representative sample of 4,046 respondents in the 2018 wave of the ifo Education Survey.[22] To investigate the impact of online-search activity on stated beliefs, we randomly assigned respondents to a control group and an encouragement-treatment group before eliciting school-spending beliefs. In both groups, beliefs about average public school spending are elicited using the same wording as in the 2017 survey wave.[23] Instead of being offered incentives for belief accuracy, members of the treatment group were encouraged to use online-search engines before stating their beliefs. The encouragement treatment is worded as follows: "*As an exception for this question, you can search the internet for the right answer to improve your guess, for example by using an internet-search engine*."[24]

After eliciting school-spending beliefs, we asked all respondents whether they searched for the correct answer on the internet in order to check whether encouraging online-search activity worked as intended. This question is worded as follows: "*To answer the preceding guess question, did you search for the correct value on the internet?*"

Appendix Table A8 (columns 1-3) presents sociodemographic characteristics for the control group and the encouragement-treatment group, indicating that the randomization successfully balanced respondents' observable characteristics across experimental groups.

4.2 The Effect of the Encouragement Treatment on Online-Search Activity

We start the analysis by investigating whether the encouragement treatment successfully increased the use of online-search engines. Table 6 reports results of regressing self-reported online-search activity on the treatment indicator. While search-engine use is very low in the control group at 1 percent, respondents in the treatment group are 14 percentage points more likely to state that they searched for the correct value on the internet. This manipulation check suggests that our encouragement treatment did in fact increase the usage of online-search engines.

[22] See Appendix for additional information about the 2018 wave of the ifo Education Survey.

[23] The true value in the 2018 survey wave is 7,500 Euro (Statistisches Bundesamt (2018)), which differs slightly from the previous year's spending level.

[24] To keep the structure of this experiment as similar as possible to the incentive experiment, the question text also informed respondents in the treatment group that they will receive feedback about the accuracy of their belief at the end of the survey.

To allow for direct comparison with the incentive experiment, we next construct the proxy for online-search activity introduced in section 3.3 for the search-engine experiment. Again, we code a dummy variable equal to one if respondents' stated beliefs correspond to one of the top Google search results for average public school spending.[25] As it turns out, the correlation of this dummy variable with the dummy variable for self-reported search-engine use is high (correlation=0.64), confirming its validity as a proxy for online-search activity.

Column 2 of Table 5 displays the encouragement-treatment effect on this proxy for online-search activity. The treatment significantly increases proxied search-engine usage by 9.1 percentage points, which is close to the 11.6 percentage-point treatment effect of incentive provision in the main experiment (see column 1). Thus, incentive provision and encouraging online-search activity seem to produce very similar effects on the probability to search the internet for the correct answer on the school-spending belief question.

4.3 The Effect of Encouraging Online-Search Activity on Belief Accuracy

The encouragement treatment strongly affects stated beliefs about average public school spending. Panel A of Figure 4 depicts the distribution of respondents' belief in the control (green bars) and treatment groups (transparent bars). In the control group, beliefs about spending levels are again very dispersed, with a 10-90 percentile range of 170 Euro to 15,000 Euro (true value: 7,500 Euro). The majority of 79 percent states a belief below the true value. This distribution of beliefs is remarkably similar to the unincentivized control group in the main experiment in the 2017 survey wave, where the 10-90 percentile range was 50 Euro to 15,000 Euro, and 78 percent of beliefs were below the true value. Comparing the distribution of beliefs between the control and the encouragement-treatment groups, the patterns are again very similar to the incentive experiment: Treated respondents are less likely to report very low belief values and more likely to report beliefs close to the true value.

Table 7 presents treatment-effect estimates based on equation (1). For most outcomes, the effects of encouraging search-engine usage on respondents' stated beliefs are very similar to the incentive effects in the main experiment (see Table 3 for comparison). The encouragement treatment increases average school-spending beliefs by 11 percent ($p<0.1$) (column 1). While the incentive-treatment effect on this outcome was not significant, the sign and magnitude of

[25] In 2018, the top results of Google searches for key phrases of our question wording ("*Ausgaben pro Schüler pro Jahr an öffenlichen, allgemeinbildenten Schulen"* and *"Bildungsausgaben pro Schüler pro Jahr"*) were 6,300, 6,500, 6,700, 6,900, 7,400, 7,500, and 8,900 Euro. Again, we searched for these values shortly after the field phase of the ifo Education Survey 2018 to obtain results close to those which the respondents would have found. Note that these values differ slightly from those found one year earlier. Screenshots are available upon request.

the two treatment effects are virtually identical. The negative sign on the treatment indicator in column 2 suggests that encouraging search-engine usage decreases the distance between stated belief and true value by 489 Euro. While this effect is shy of statistical significance, it again goes in the same direction as the incentive-treatment effect on this variable (which is larger and statistically significant). The encouragement treatment significantly increases the probability of reporting a belief within the incentivized range by 9 percentage points (column 3) and the probability of reporting a belief within the 10-percent interval around the true value by 5 percentage points (column 4).[26] Again, these effects are similar, in significance and magnitude, to the incentive-treatment effects.

The encouragement treatment also significantly increases respondents' confidence about their beliefs (column 6) and the response time (column 7). While treatment effects on confidence are very similar across the two experiments, the encouragement effect on response time is somewhat smaller than the incentive effect. The only outcome for which we observe fairly different treatment effects between the encouragement treatment and the incentive treatment is the probability to report very low belief values below 100 Euro (column 5). In contrast to the incentive treatment, encouraging search-engine usage does not affect reporting of implausibly low belief values, although (as in the follow-up survey of the incentive experiment) this is driven by a lower incidence in the control group. In general, the results confirm the visual impression that the effects of incentives and encouraging online-search activity are remarkably similar.

Appendix Table A9 shows results of a model that stacks the data of the incentive experiment and the encouragement experiment. Results confirm that the treatment effects differ significantly only on value estimates below 100 Euro. For all other outcome measures, the treatment effects of the encouragement experiment do not differ significantly from the treatment effects of the incentive experiment.

In sum, the similarity between the effects of the incentive treatment and the encouragement treatment, together with the inspection of response spikes in the incentive treatment in section 3.3, strongly suggest that at least part of the effect of incentive provision on the accuracy of school-spending beliefs is due to incentives increasing respondents' online-search activity to improve their stated beliefs.

[26] The incentivized interval in the search-engine experiment ranges from 1,332 Euro to 13,468 Euro. Even though no incentives were provided in this experiment, we report treatment effects on this measure to allow comparison to the findings from the main experiment. The 10-percent interval around the true value ranges from 6,750 Euro to 8,250 Euro.

5. Inducing Experimenter Demand

One major concern with the interpretation of treatment effects in experimental work is bias due to experimenter-demand effects, which refer to "changes in behavior by experimental subjects due to cues about what constitutes appropriate behavior" (Zizzo (2010), p. 75). For the case of incentive provision, the experimenter's willingness to pay out monetary incentives for accurate beliefs may signal to respondents that the incentivized belief question is of particular importance to the experimenter. If so, respondents might adapt their answering behavior not only to earn the incentive, but also to please the experimenter by being a "good" respondent. To address this concern, we report results from an additional experiment where we aim to induce experimenter-demand effects in order to assess their potential effect on stated school-spending beliefs. In what follows, we first describe the experimental design and then present the results.

5.1 Experimental Design

We conducted the demand experiment in another representative sample of 3,124 respondents in the 2016 wave of the ifo Education Survey.[27] Respondents were randomized into a control group or a demand-treatment group. For respondents in the demand treatment, we used the following wording to induce experimenter-demand effects before eliciting their beliefs about average school spending: "*As you might know, government institutions collect a variety of key statistics about schools. We are interested in discovering whether the public is familiar with these key statistics. On the next screen, we will ask you a question about such a key statistic, to which there are correct and incorrect answers. In order for your response to be informative for us, it is very important that you answer this question as accurately as possible.*"[28] On the next screen, beliefs about school spending were elicited in the same way as in the control group, using the same wording as in the 2017 and 2018 waves.[29]

Columns 4-6 of Appendix Table A8 show that respondents' sociodemographic characteristics are again well balanced across experimental groups, indicating that randomization was successful.

[27] See Appendix for additional information about the 2016 wave of the ifo Education Survey.

[28] Again, the question text also informed treated respondents that they will receive feedback about the accuracy of their belief at the end of the survey.

[29] In spirit, this demand treatment is very similar to recent papers by de Quidt et al. (2018) and Mummollo and Peterson (2018) which measure and bound experimenter-demand effects in the context of economic games and survey experiments, respectively.

5.2 The Effect of Inducing Experimenter Demand on Belief Accuracy

Panel B of Figure 4 depicts the results of the demand experiment on respondents' stated beliefs about average public school spending. The distribution of beliefs in the control group is similar to the previously reported survey waves, with a 10-90 percentile range of 150 Euro to 10,000, and 82 percent of respondents stating a belief below the true value (true value: 7,100 Euro; Statistisches Bundesamt (2016)). However, in contrast to the previously reported experiments on school-spending beliefs, graphical inspection does not yield any obvious difference in the distribution of beliefs between control group (green bars) and demand-treatment group (transparent bars).

The regression analysis in Table 8 confirms this impression. There is no indication that inducing experimenter demand improves the accuracy of stated beliefs. The small and insignificant coefficients on the treatment indicator in most regressions suggest that the experimenter-demand treatment does not affect stated beliefs about average school spending. The only marginally significant treatment effect ($p<0.1$) is that the treatment *increases* the absolute distance between stated belief and true value.

Consistently, column 3 of Table 5 shows that the demand treatment also does not affect our proxy for online-search activity.[30] Taken together, the evidence suggests that the effects of incentive provision on answering behavior reflect genuine incentive effects as opposed to experimenter-demand effects.

6. Conclusions

Subjective beliefs about economic facts are a key concept for explaining economically relevant behavior and choices. The fact that researchers usually measure these beliefs using unincentivized survey questions raises concerns about systematic biases in reporting that might undermine meaningful analysis of subjective beliefs. We conduct two experiments in a representative German online survey to study whether incentivizing belief accuracy affects stated beliefs about two important economic facts: average earnings by professional degree and average public school spending. We find that incentivization increases response time in both experiments but that this translates into more accurate beliefs only for average school spending.

[30] For this purpose, we again code a dummy variable equal to one if respondents report one of the top Google search results for average public school spending (2,200, 4,900, 5,600, 6,000, 6,200, 6,300, 6,500, 7,300, 8,000, and 8,100 Euro). Unlike in the 2017 and 2018 waves, we did not record the search results immediately after the implementation of the 2016 survey wave. However, when looking up the Google search results in autumn 2018, we restrict the search to display only results published until end of June 2016, the end of the field phase of the 2016 survey. Screenshots are available upon request.

To reconcile the heterogeneity of incentive effects on stated beliefs in the two experiments, we present evidence from comparisons of response spikes with search-engine results and from an additional experiment that encourages online-search activity which suggest that respondents resort to online-search engines in response to incentive provision. In another experiment that induces experimenter demand, we show that these results are unlikely to reflect experimenter-demand effects in our setting.

We draw two main conclusions from our analyses. First, the finding that monetary incentives fail to improve respondents' earnings beliefs suggests that unincentivized belief measures, which are heavily used in the literature, do not suffer from reporting bias due to self-serving or low-effort answering behavior. This does not rule out, however, that incentive provision can be an important design feature in certain research contexts.[31]

Second, providing monetary incentives in online surveys might increase respondents' use of external resources such as online-search engines to improve the accuracy of their stated beliefs. The extent to which respondents engage in this behavior will likely depend on a number of factors, such as the question at hand, the specific survey setting (e.g., whether it is possible to restrict or track respondents' online-search activity), or the design of the incentive scheme. Also, whether or not researchers want respondents to consult external sources to update beliefs might depend on the specific research question. In any case, our results raise a cautionary note that incentivizing belief accuracy might trigger unintended behavioral responses in online surveys and other contexts where experimenters' control over the experimental setting is limited.

For future research, it would be interesting to investigate which attributes of a belief question mediate the effects of incentive provision. As the two belief questions scrutinized in this paper – beliefs about average earnings and school spending levels – differ in various dimensions such as respondents' confidence about their beliefs, question complexity, respondents' familiarity with the elicited concepts, and the online availability of belief-improving information, further investigating the interplay between question attributes and incentive effects would be insightful.

[31] For instance, incentivizing beliefs might be important when researchers are interested in costly information-acquisition behavior (e.g. Fuster et al. (2018)) or when they want to mitigate the reporting of motivated beliefs (e.g. Bullock et al. (2015); Prior et al. (2015)).

References

Alesina, Alberto and Eliana La Ferrara, "Preferences for Redistribution in the Land of Opportunities," *Journal of Public Economics*, 89 (2005), 897–931.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva, "Immigration and Redistribution," NBER Working Paper 24733 (2018a).

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso, "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, 108 (2018b), 521–554.

Ansolabehere, Stephen, Marc Meredith, and Erik Snowberg, "Asking about Numbers: Why and How," *Political Analysis*, 21 (2013), 48–69.

Arcidiacono, Peter, "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 121 (2004), 343–375.

Arcidiacono, Peter, V. Joseph Hotz, and Songman Kang, "Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals," *Journal of Econometrics*, 166 (2012), 3–16.

Arechar, Antonio A., Simon Gächter, and Lucas Molleman, "Conducting Interactive Experiments Online," *Experimental Economics*, 21 (2018), 99–131.

Armantier, Olivier, Wändi B. d. Bruin, Simon Potter, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar, "Measuring Inflation Expectations," *Annual Review of Economics*, 5 (2013), 273–301.

Armantier, Olivier, Scott Nelson, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar, "The Price Is Right: Updating Inflation Expectations in a Randomized Price Information Experiment," *Review of Economics and Statistics*, 98 (2016), 503–523.

Armona, Luis, Andreas Fuster, and Basit Zafar, "Home Price Expectations and Behavior: Evidence from a Randomized Information Experiment," Federal Reserve Bank of New York Staff Report 798 (2016).

Attanasio, Orazia, Teodora Boneva, and Christopher Rauh, "Parental Beliefs about Returns to Different Types of Investments in School Children," HCEO Working Paper 2018-032 (2018).

Belfield, Chris, Teodora Boneva, Christopher Rauh, and Jonathan Shaw, "What Drives Enrollment Gaps in Further Education? The Role of Beliefs in Sequential Schooling Decisions," *Economica*, (2019), forthcoming.

Benítez-Silva, Hugo, Moshe Buchinsky, Hiu Man Chan, Sofia Cheidvasser, and John Rust, "How Large is the Bias in Self-reported Disability?," *Journal of Applied Econometrics*, 19 (2004), 649–670.

Bertrand, Marianne and Sendhil Mullainathan, "Do People Mean What They Say? Implications for Subjective Survey Data," *American Economic Review*, 91 (2001), 67–72.

Blanco, Mariana, Dirk Engelmann, Alexander K. Koch, and Hans-Theo Normann, "Belief Elicitation in Experiments: Is There a Hedging Problem?," *Experimental Economics*, 13 (2010), 412–438.

Bleemer, Zachary and Basit Zafar, "Intended College Attendance: Evidence from an Experiment on College Returns and Costs," *Journal of Public Economics*, 157 (2018), 184–211.

Boneva, Teodora and Christopher Rauh, "Socio-economic Gaps in University Enrollment: The Role of Perceived Pecuniary and Non-Pecuniary Returns," HCEO Working Paper 2017-080 (2017).

Boneva, Teodora and Christopher Rauh, "Parental Beliefs about Returns to Educational Investments – The Later the Better?," *Journal of the European Economic Association*, 16 (2018), 1669–1711.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber, "Partisan Bias in Factual Beliefs about Politics," *Quarterly Journal of Political Science*, 10 (2015), 519–578.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia, "Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments," *American Economic Journal: Macroeconomics*, 9 (2017), 1–35.

Clifford, Scott and Jennifer Jerit, "Is there a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies," *Journal of Experimental Political Science*, 1 (2014), 120–131.

Coibion, Olivier and Yuriy Gorodnichenko, "What Can Survey Forecasts Tell Us about Information Rigidities?," *Journal of Political Economy*, 120 (2012), 116–159.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar, "How Do Firms Form their Expectations? New Survey Evidence," *American Economic Review*, 108 (2018), 2671–2713.

Delavande, Adeline, Xavier Giné, and David McKenzie, "Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How Sensitive are Answers to Variations in Elicitation Design?," *Journal of Applied Econometrics*, 26 (2011), 479–497.

Delavande, Adeline and Basit Zafar, "University Choice: The Role of Expected Earnings, Non-Pecuniary Outcomes, and Financial Constraints," *Journal of Political Economy*, (2019), forthcoming.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde, "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," IZA Discussion Paper 9674 (2016).

Fuster, Andreas, Ricardo Perez-Truglia, and Basit Zafar, "Expectations with Endogenous Information Acquisition: An Experimental Investigation," NBER Working Paper 24767 (2018).

Gächter, Simon and Elke Renner, "The Effects of (Incentivized) Belief Elicitation in Public Goods Experiments," *Experimental Economics*, 13 (2010), 364–377.

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal, "Does Information Change Attitudes Towards Immigrants? Representative Evidence from Survey Experiments," IZA Discussion Paper 10419 (2016).

Haaland, Ingar and Christopher Roth, "Beliefs about Racial Discrimination: Representative Evidence," Working Paper (2017).

Hoxby, Caroline and Sarah Turner, "Expanding College Opportunities for High-Achieving, Low Income Students," Stanford Institute for Economic Policy Research Discussion Paper 12-014 (2013).

Karadja, Mounir, Johanna Mollerstrom, and David Seim, "Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Demand for Redistribution," *Review of Economics and Statistics*, 99 (2017), 201–212.

Kaufmann, Katja M., "Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns," *Quantitative Economics*, 5 (2014), 583–630.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva, "How Elastic are Preferences for Redistribution? Evidence from Randomized Survey Experiments," *American Economic Review*, 105 (2015), 1478–1508.

Lergetporer, Philipp, Guido Schwerdt, Katharina Werner, Martin R. West, and Ludger Woessmann, "How Information Affects Support for Education Spending: Evidence from Survey Experiments in Germany and the United States," *Journal of Public Economics*, 167 (2018a), 138–157.

Lergetporer, Philipp, Katharina Werner, and Ludger Woessmann, "Does Ignorance of Economic Returns and Costs Explain the Educational Aspiration Gap? Evidence from Representative Survey Experiments," CESifo Working Paper 7192 (2018b).

—— "Educational Inequality and Public Policy Preferences: Evidence from Representative Survey Experiments," CESifo Working Paper 7192 (2018c).

Manski, Charles F., "Measuring Expectations," *Econometrica*, 72 (2004), 1329–1376.

—— "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise," *NBER Macroeconomics Annual*, 32 (2018), 411–471.

Mummolo, Jonathan and Erik Peterson, "Demand Effects in Survey Experiments: An Empirical Assessment," *American Political Science Review*, 22 (2018), 1–13.

Prior, Markus, Gaurav Sood, and Kabir Khanna, "You Cannot Be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions," *Quarterly Journal of Political Science*, 10 (2015), 489–518.

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 108 (2018), 3266–3302.

Roth, Christopher and Johannes Wohlfart, "How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?," Working Paper (2018a).

—— "Public Debt and the Demand for Government Spending and Taxation," Working Paper (2018b).

Schlag, Karl H., James Tremewan, and Joël J. Van der Weele, "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs," *Experimental Economics*, 18 (2015), 457–490.

Schueler, Beth E. and Martin R. West, "Sticker Shock: How Information Affects Citizen Support for Public School Funding," *Public Opinion Quarterly*, 80 (2016), 90–113.

Statistisches Bundesamt, "Bildungsausgaben: Ausgaben je Schülerin und Schüler 2013," Wiesbaden: Statistisches Bundesamt (2016).

—— "Bildungsausgaben: Ausgaben je Schülerin und Schüler 2014," Wiesbaden: Statistisches Bundesamt (2017).

—— "Bildungsausgaben: Ausgaben je Schülerin und Schüler 2015," Wiesbaden: Statistisches Bundesamt (2018).

Stinebrickner, Ralph and Todd R. Stinebrickner, "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout," *Review of Economic Studies*, 81 (2013), 426–472.

Trautmann, Stefan T. and Gijs Kuilen, "Belief Elicitation: A Horse Race among Truth Serums," *Economic Journal*, 125 (2014), 2116–2135.

Wang, Stephanie W., "Incentive Effects: The Case of Belief Elicitation from Individuals in Groups," *Economics Letters*, 111 (2011), 30–33.

Wiswall, Matthew and Basit Zafar, "Determinants of College Major Choice: Identification using an Information Experiment," *Review of Economic Studies*, 82 (2015), 791–824.

Zafar, Basit, "How Do College Students Form Expectations?," *Journal of Labor Economics*, 29 (2011), 301–348.

Zizzo, Daniel J., "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 13 (2010), 75–98.

Appendix: Description of Additional Survey Waves

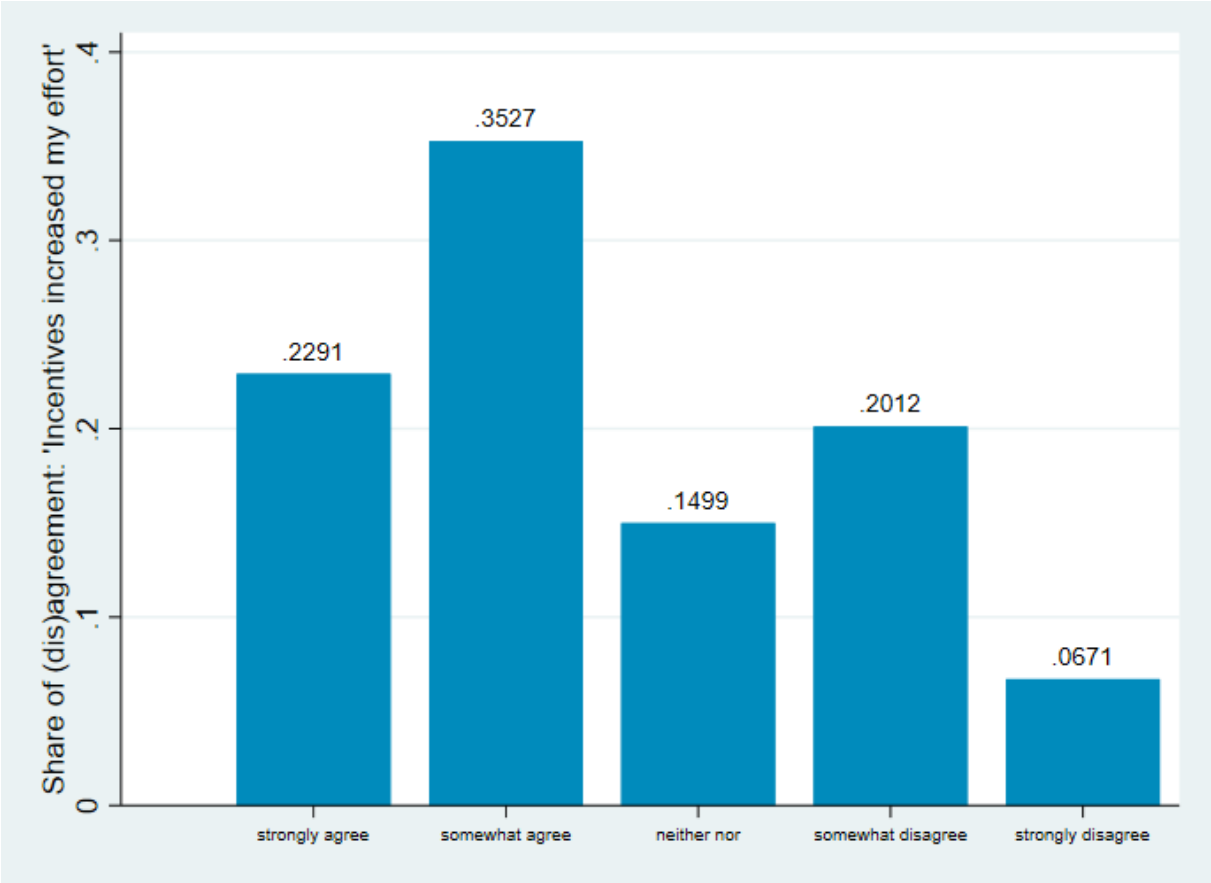
While our incentive experiments were implemented in the 2017 wave of the ifo Education Survey (described in section 2.1), the other two experiments were implemented in the 2016 and 2018 waves of the ifo Education Survey, respectively.

The ifo Education Survey 2018 was fielded by Kantar Public between May and June 2018. The sample consists of 4,046 respondents who are representative for the German voting-age population (18 years and older). Respondents completed the survey on their own digital devices. Median completion time was 15 minutes.

The ifo Education Survey 2016 was fielded by Kantar Public between April and June 2016. The sample consists of 3,302 respondents who are representative for the German voting-age population (18 years and older). In contrast to the data of the 2017 and 2018 waves employed in this paper, respondents were drawn in two strata. Respondents with access to the internet were recruited via an online platform and completed the survey on their own digital devices. Respondents without internet access were polled at their homes by trained interviewers. The interviewers equipped these respondents with tablet computers and asked them to fill out the survey on these devises. Median completion time was 18 minutes.

As with the 2017 data, both the 2016 and the 2018 analyses employ survey weights to achieve representativeness for the German population. The weights match characteristics of the overall German population with respect to age, gender, federal state, school degree, and municipality size. Performing the analyses without using the survey weights does not change the qualitative results.

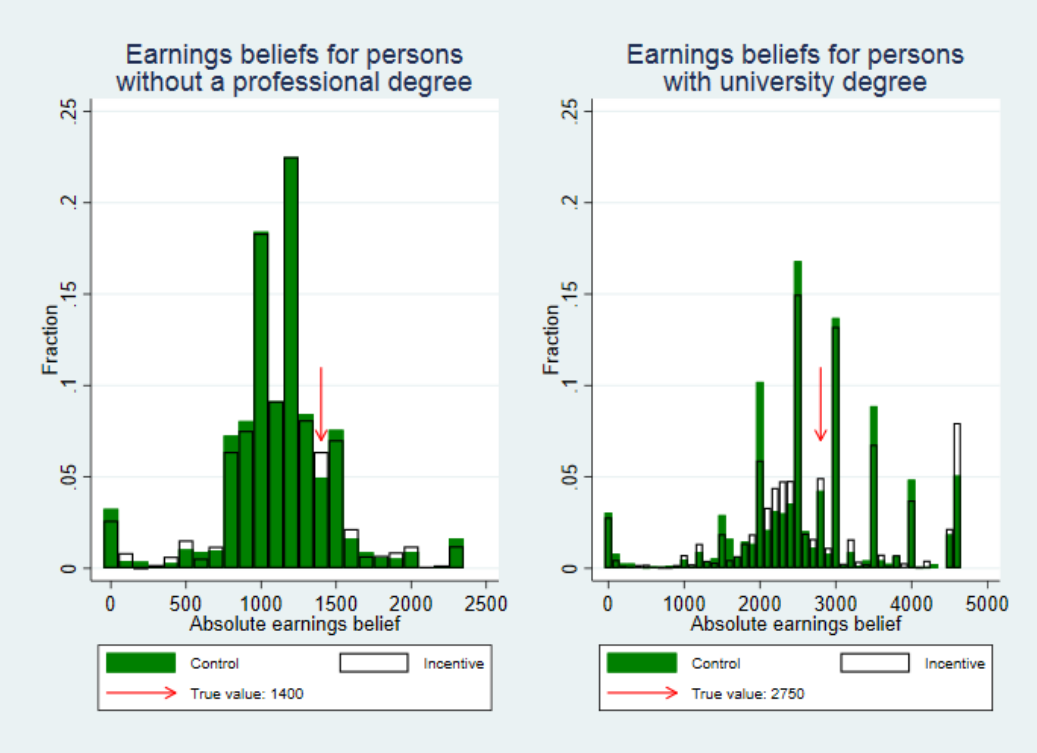
Figure 1: Survey responses on whether incentives increased effort



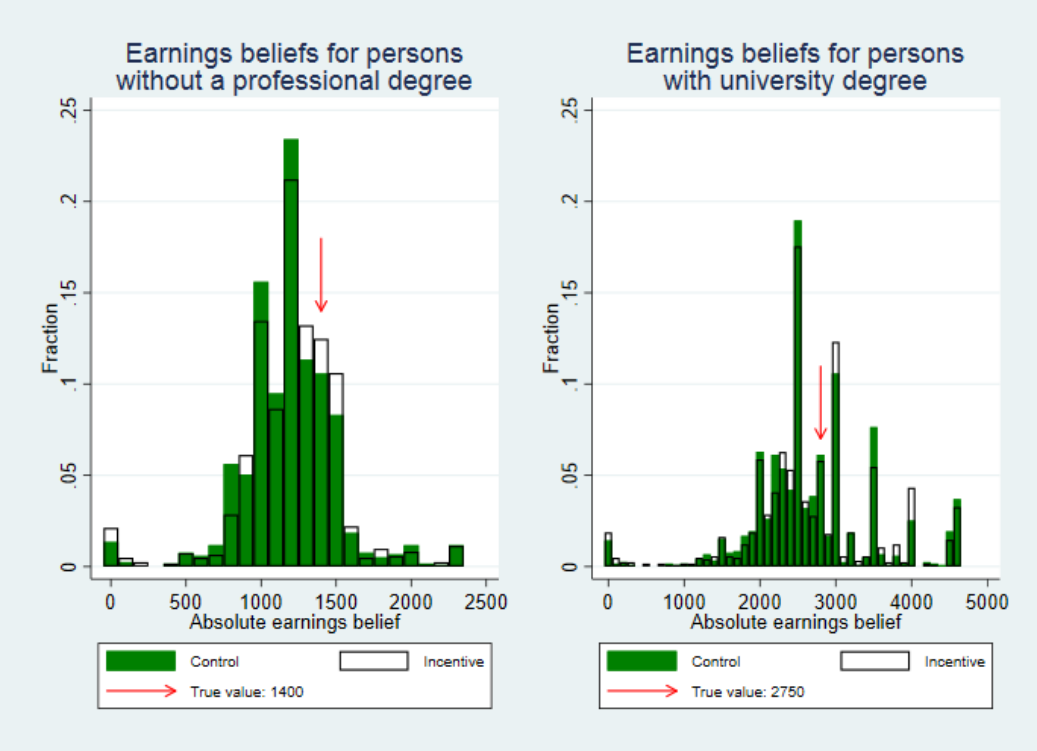
Notes: Wording of the question: "In the previous questions, you had the opportunity to earn additional tokens by providing a good guess. To what extent do you agree with the following statement (your answer will not affect your probability of winning): The prospect of receiving additional tokens has provided an incentive for me to put more effort in my guess." Sample: respondents incentivized in at least one belief question. Responses weighted by survey weights. Source: ifo Education Survey 2017.

Figure 2: Distribution of earnings beliefs with and without incentive provision

Panel A: Main survey



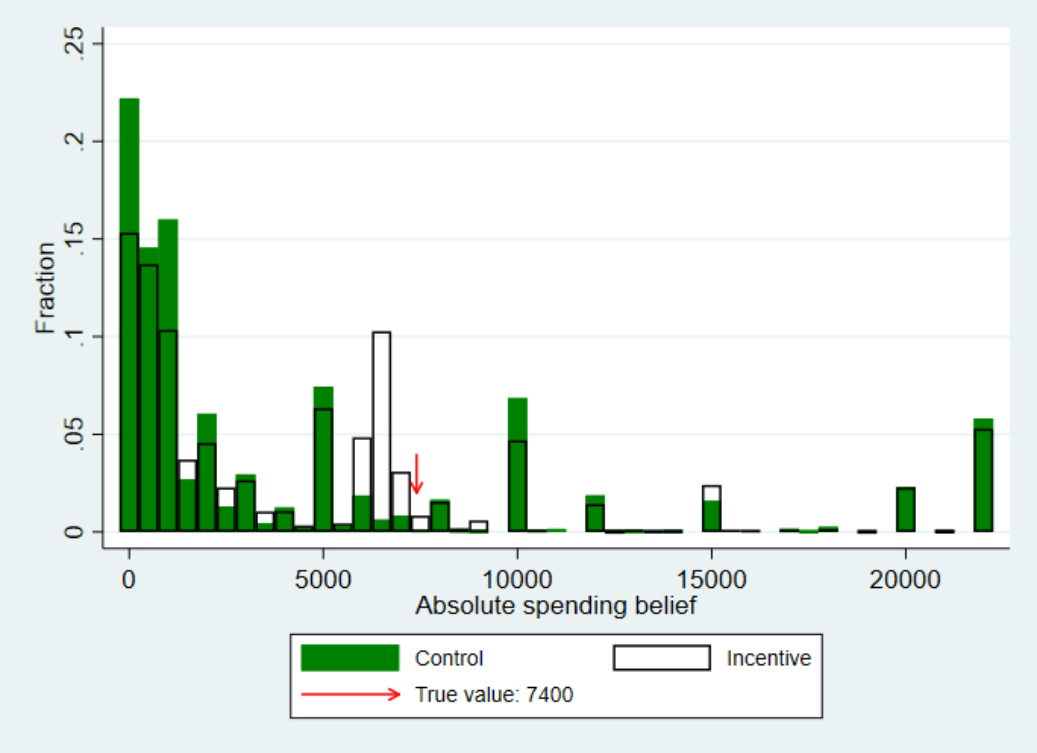
Panel B: Follow-up survey



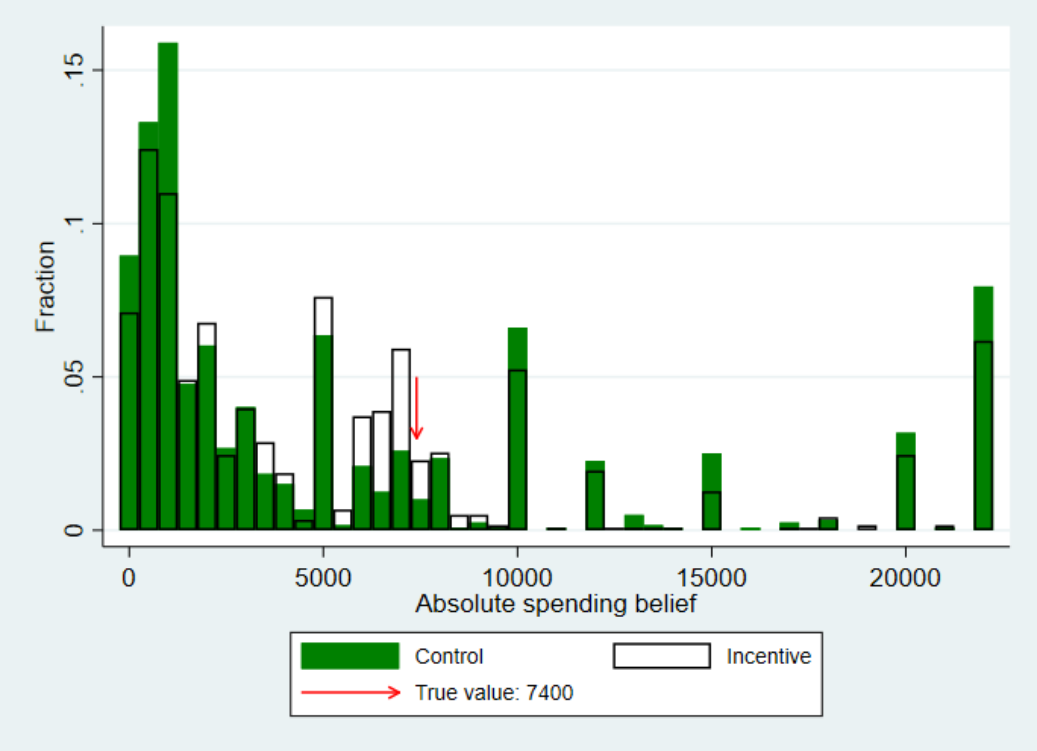
Notes: Histograms of respondents' stated beliefs about average earnings of persons without a degree (left panels) and with a university degree (right panels). Beliefs are winsorized at 2,380 Euro respectively 4,675 Euro to deal with outliers. Incentive: Belief accuracy incentivized. Control: Belief accuracy not incentivized. Panel A: Beliefs elicited in main survey. Panel B: Beliefs elicited in follow-up survey about two weeks later. Responses weighted by survey weights. Source: ifo Education Survey 2017.

Figure 3: Distribution of school-spending beliefs with and without incentive provision

Panel A: Main survey



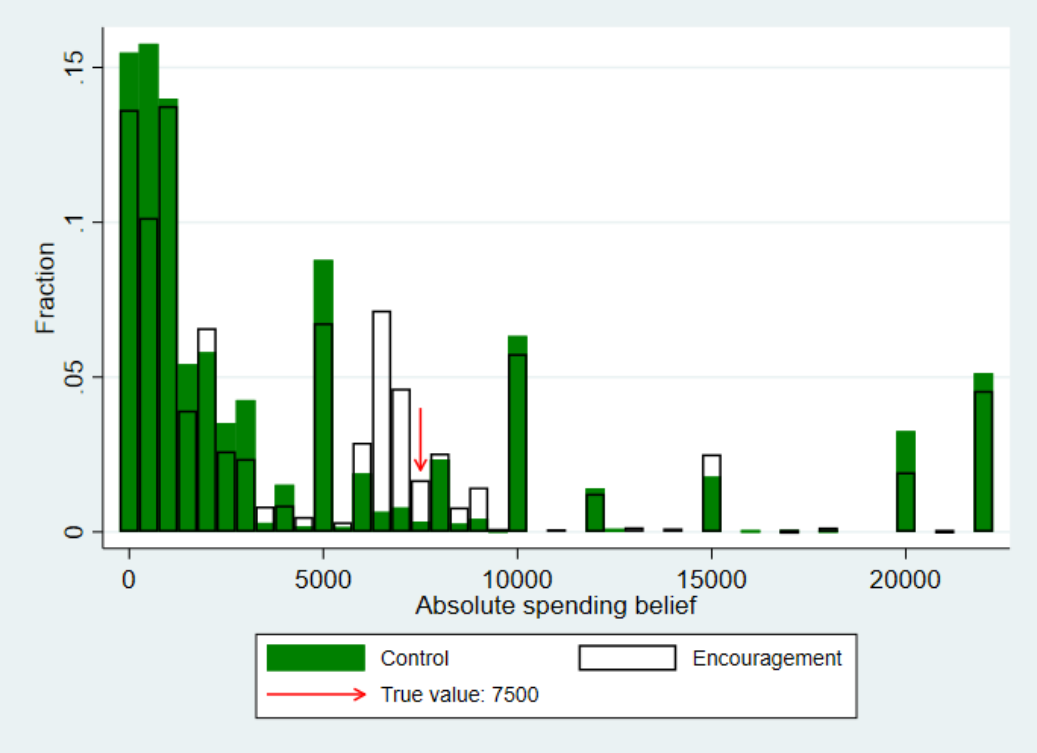
Panel B: Follow-up survey



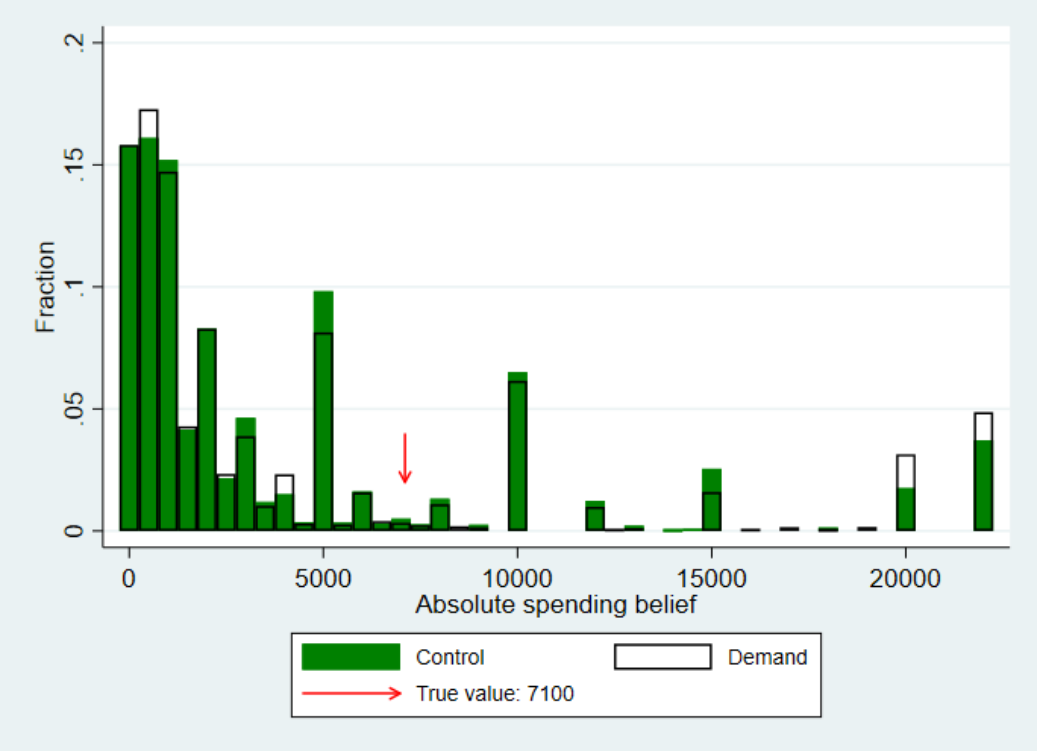
Notes: Histograms of respondents' stated beliefs about average annual public school spending per student. Beliefs are winsorized at 22,200 Euro to deal with outliers. Incentive: Belief accuracy incentivized. Control: Belief accuracy not incentivized. Panel A: Beliefs elicited in main survey. Panel B: Beliefs elicited in follow-up survey about two weeks later. Responses weighted by survey weights. Source: ifo Education Survey 2017.

Figure 4: Distribution of school-spending beliefs in additional experiments

Panel A: Experiment on encouraging online-search activity



Panel B: Experiment on inducing experimenter-demand effects



Notes: Histograms of respondents' stated beliefs about average annual public school spending per student. Panel A: Beliefs in the experiment on encouraging online-search activity. Panel B: Beliefs in the experiment on inducing experimenter-demand effects. Beliefs are winsorized at 22,200 Euro to deal with outliers. Encouragement/Demand: Belief in the respective treatment group. Control: Belief in the control group. Responses weighted by survey weights. Source: Panel A: ifo Education Survey 2018; Panel B: ifo Education Survey 2016.

Table 1: Summary statistics and balancing tests of incentive experiments

	Earnings experiment			School-spending experiment		
	Control mean	Treatment mean	Difference	Control mean	Treatment mean	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Highest educational attainment						
No degree/basic degree	0.360	0.403	0.043	0.386	0.378	-0.007
Middle school degree	0.307	0.297	-0.009	0.302	0.302	0.000
University entrance degree	0.333	0.300	-0.033	0.313	0.320	0.007
Age	49.010	49.855	0.845	49.235	49.663	0.428
Monthly household income (€)	2,323.947	2,282.017	-41.930	2,327.084	2,276.068	-51.016
Female	0.495	0.527	0.032	0.505	0.518	0.013
Born in Germany	0.948	0.966	0.018**	0.956	0.959	0.003
Partner in household	0.593	0.563	-0.030	0.576	0.580	0.004
Lives in West Germany	0.795	0.806	0.011	0.805	0.797	-0.008
City size ≥ 100,000	0.323	0.312	-0.011	0.307	0.328	0.020
Has children	0.579	0.576	-0.003	0.581	0.572	-0.009
Parent(s) with university degree	0.279	0.268	-0.011	0.277	0.269	-0.008
Employment status						
Full-time employed	0.348	0.327	-0.021	0.347	0.328	-0.019
Part-time employed	0.134	0.114	-0.020	0.126	0.121	-0.005
Self-employed	0.039	0.043	0.004	0.046	0.035	-0.011
Unemployed	0.045	0.041	-0.004	0.037	0.049	0.013
House wife/husband/retired/ill/student	0.053	0.072	0.019	0.058	0.067	0.009
Works in education sector	0.088	0.079	-0.010	0.089	0.077	-0.012
Risk tolerance (11-point scale)	4.389	4.258	-0.131	4.337	4.306	-0.031
Patience (11-point scale)	6.089	6.217	0.128	6.118	6.194	0.076
Item non-response: Beliefs	0.001	0.001	-0.001	0.003	0.003	0.000
Observations	1,852	1,844		1,856	1,840	

Notes: Weighted group means. "Difference" displays the difference in means between the respective control and treatment groups. Significance levels of "Difference" stem from linear regressions of the background variables on the respective treatment dummies. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table 2: Incentive effects on earnings beliefs

	Belief relative to true value	Distance to true value	Belief within incentivized range	Belief within 10-percent range	Extreme values below 100	Confidence	Response time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Beliefs on earnings without a degree							
Incentive	0.080	97.463	0.015	-0.001	-0.003		
	(0.064)	*(87.020)*	*(0.021)*	*(0.017)*	*(0.007)*		
Control mean	0.84	429.44	0.55	0.19	0.04		
Observations	3,694	3,694	3,694	3,694	3,694		
R^2	0.001	0.001	0.000	0.001	0.001		
Panel B: Beliefs on earnings with a university degree							
Incentive	0.057*	80.940	-0.011	-0.011	-0.005	0.033	19.609***
	(0.034)	*(83.067)*	*(0.020)*	*(0.021)*	*(0.007)*	*(0.067)*	*(5.886)*
Control mean	1.01	900.00	0.71	0.39	0.04	3.39	31.17
Observations	3,693	3,693	3,693	3,693	3,693	3,692	3,696
R^2	0.002	0.001	0.000	0.001	0.001	0.001	0.004

Notes: OLS regressions. Dependent variable: column (1): stated belief divided by true value; column (2): absolute distance between stated belief and true value; column (3): dummy variable coded one if stated belief is in the incentivized range; column (4): dummy variable coded one if stated belief is in the 10-percent interval around the true value; column (5): dummy variable coded one if stated belief is below 100; column (6): confidence about belief on seven-point Likert scale (1="very unsure", 7="very sure"); column (7): difference between actual and predicted response time. Randomized experimental treatment "Incentive": respondents offered monetary incentive for belief accuracy. Control mean: mean of the outcome variable for the control group. Panel A: beliefs about net average monthly earnings of full-time employed persons without any professional degree. Panel B: beliefs about net average monthly earnings of full-time employed persons with a university degree. Since both earnings beliefs were elicited on the same screen, confidence and response time were recorded only once for earnings beliefs in general. Regressions control for treatment status in the incentive experiment on school-spending beliefs. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table 3: Incentive effects on school-spending beliefs

	Belief relative to true value	Distance to true value	Belief within incentivized range	Belief within 10-percent range	Extreme values below 100	Confidence	Response time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentive	0.112	-720.657	0.132***	0.044***	-0.064***	0.768***	86.068**
	(0.078)	*(470.283)*	*(0.021)*	*(0.009)*	*(0.015)*	*(0.068)*	*(38.013)*
Control mean	0.77	7590.93	0.37	0.03	0.16	2.42	84.64
Observations	3,682	3,682	3,682	3,682	3,682	3,680	3,696
R^2	0.00	0.00	0.02	0.01	0.01	0.06	0.00

Notes: OLS regressions. Dependent variable: beliefs about average annual public school spending per student; column (1): stated belief divided by true value; column (2): absolute distance between stated belief and true value; column (3): dummy variable coded one if stated belief is in the incentivized range; column (4): dummy variable coded one if stated belief is in the 10-percent interval around the true value; column (5): dummy variable coded one if stated belief is below 100; column (6): confidence about belief on seven-point Likert scale (1="very unsure", 7="very sure"); column (7): difference between actual and predicted response time. Randomized experimental treatment "Incentive": respondents offered monetary incentive for belief accuracy. Control mean: mean of the outcome variable for the control group. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table 4: Incentive effects on school-spending beliefs in follow-up survey

	Belief relative to true value	Distance to true value	Belief within incentivized range	Belief within 10-percent range	Extreme values below 100	Confidence	Response time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentive	-0.094	-1354.549***	0.110***	0.056***	-0.008	0.355***	18.012
	(0.072)	*(441.580)*	*(0.020)*	*(0.012)*	*(0.008)*	*(0.064)*	*(15.376)*
Control mean	1.04	8008.43	0.47	0.06	0.05	2.90	12.34
Observations	2,377	2,377	2,377	2,377	2,377	2,378	2,392
R^2	0.001	0.004	0.012	0.010	0.000	0.013	0.001

Notes: OLS regressions. Dependent variable (recorded in follow-up survey conducted about two weeks after the main survey): beliefs about average annual public school spending per student; column (1): stated belief divided by true value; column (2): absolute distance between stated belief and true value; column (3): dummy variable coded one if stated belief is in the incentivized range; column (4): dummy variable coded one if stated belief is in the 10-percent interval around the true value; column (5): dummy variable coded one if stated belief is below 100; column (6): confidence about belief on seven-point Likert scale (1="very unsure", 7="very sure"); column (7): difference between actual and predicted response time. Randomized experimental treatment "Incentive": respondents offered monetary incentive for belief accuracy. Control mean: mean of the outcome variable for the control group in the follow-up survey. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table 5: Treatment effects on proxy for online-search activity

	Dependent variable: Proxy for online-search-engine use		
	Experiment on incentive provision	Experiment on encouraging online-search activity	Experiment on inducing experimenter-demand effects
	(1)	(2)	(3)
Treatment	0.116***	0.091***	-0.001
	(0.011)	*(0.010)*	*(0.007)*
Control mean	0.02	0.01	0.03
Observations	3,682	4,031	3,124
R^2	0.05	0.04	0.00

Notes: OLS regressions. Dependent variable: dummy variable coded one if respondents' belief corresponds to one of the top search-engine results. Randomized experimental treatment "Treatment": column (1): respondents offered monetary incentive for belief accuracy; column (2): respondents encouraged to search the internet for the correct answer; column (3): respondents informed that it is important for the experimenter that they provide a correct answer. Control mean: mean of the outcome variable for the control group. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: column (1): ifo Education Survey 2017; column (2): ifo Education Survey 2018; column (3): ifo Education Survey 2016.

Table 6: Encouragement-treatment effects on self-reported online-search activity

	Self-reported online-search activity
	(1)
Encouragement	0.137***
	(0.013)
Control mean	0.01
Observations	4,045
R^2	0.05

Notes: OLS regression. Dependent variable: dummy variable coded one if respondent reports to have searched for the correct answer to the school-spending-belief question online. Randomized experimental treatment "Encouragement": respondents encouraged to search the internet for the correct answer. Control mean: mean of the outcome variable for the control group. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2018.

Table 7: Encouragement-treatment effects on school-spending beliefs

	Belief relative to true value	Distance to true value	Belief within incentivized range	Belief within 10-percent range	Extreme values below 100	Confidence	Response time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Encouragement	0.113*	-488.511	0.087***	0.054***	0.002	0.615***	24.851***
	(0.064)	*(371.087)*	*(0.023)*	*(0.010)*	*(0.013)*	*(0.074)*	*(7.107)*
Control mean	0.70	6778.19	0.44	0.03	0.08	2.58	6.18
Observations	4,031	4,031	4,031	4,031	4,031	4,046	4,046
R^2	0.00	0.00	0.01	0.01	0.00	0.04	0.00

Notes: OLS regressions. Dependent variable: beliefs about average annual public school spending per student; column (1): stated belief divided by true value; column (2): absolute distance between stated belief and true value; column (3): dummy variable coded one if stated belief is in the incentivized range; column (4): dummy variable coded one if stated belief is in the 10-percent interval around the true value; column (5): dummy variable coded one if stated belief is below 100; column (6): confidence about belief on seven-point Likert scale (1="very unsure", 7="very sure"); column (7): difference between actual and predicted response time. Randomized experimental treatment "Encouragement": respondents encouraged to search the internet for the correct answer. Control mean: mean of the outcome variable for the control group. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2018.

Table 8: Demand-treatment effects on school-spending beliefs

	Belief relative to true value	Distance to true value	Belief within incentivized range	Belief within 10-percent range	Extreme values below 100	Confidence	Response time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Demand	0.073	559.019*	-0.016	-0.001	0.002	-0.040	0.974
	(0.059)	*(323.720)*	*(0.017)*	*(0.004)*	*(0.012)*	*(0.060)*	*(3.343)*
Control mean	0.68	6485.19	0.23	0.01	0.09	2.59	8.17
Observations	3,124	3,124	3,124	3,124	3,124	3,134	3,302
R^2	0.001	0.001	0.000	0.000	0.000	0.000	0.000

Notes: OLS regressions. Dependent variable: beliefs about average annual public school spending per student; column (1): stated belief divided by true value; column (2): absolute distance between stated belief and true value; column (3): dummy variable coded one if stated belief is in the incentivized range; column (4): dummy variable coded one if stated belief is in the 10-percent interval around the true value; column (5): dummy variable coded one if stated belief is below 100; column (6): confidence about belief on seven-point Likert scale (1="very unsure", 7="very sure"); column (7): difference between actual and predicted response time. Randomized experimental treatment "Demand": respondents informed that it is important for the experimenter that they provide a correct answer. Control mean: mean of the outcome variable for the control group. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2016.

Figure A1: Google search results: School spending

A. Search for "Bildungsausgaben pro Schüler pro Jahr"

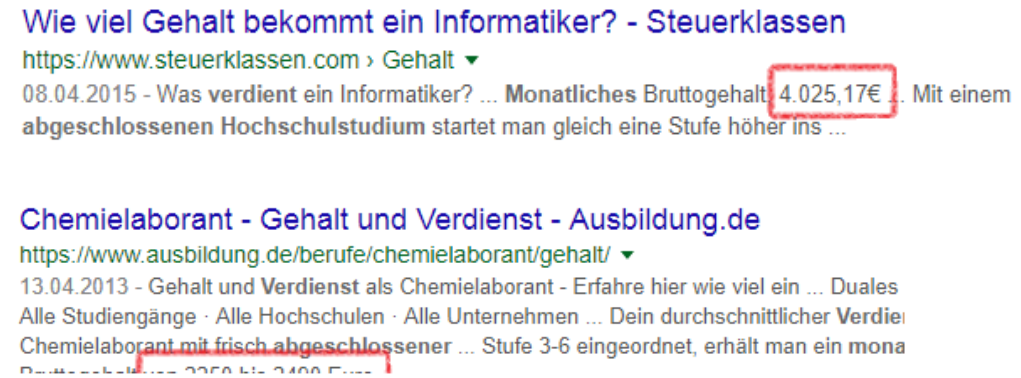
B. Search for „Ausgaben pro Schüler pro Jahr an öffenlichen. allgemeinbildenten Schulen"

Notes: Screenshots of Google search results of the indicated keywords.

Figure A2: Google search results: Earnings by professional degree

A. Earnings of persons without professional degree:
Search for „Monatlicher Verdienst Personen ohne abgeschlossene Berufsausbildung“

B. Earnings of persons with a university degree:
Search for „Monatlicher Verdienst Personen mit abgeschlossenem Hochschulstudium“



C. Earnings of persons with different professional degrees:
Search for „Monatlicher Verdienst Personen mit unterschiedlichem Bildungsabschluss“

Notes: Screenshots of Google search results of the indicated keywords.

Table A1: Wording of the survey questions

No.	Subgroup	Wording of question
Incentive experiments (2017 survey wave)		
10	Control	Based on your best guess, what is the average amount of money spent each year for a child in public schools in Germany?
	Treatment "Incentive"	On the following screen, you will be asked another guess question. If your guess is roughly correct, you will receive an additional reward worth about half of the reward you get for participating in this survey. (next screen) Based on your best guess, what is the average amount of money spent each year for a child in public schools in Germany?
11	Control	Persons with a professional degree (apprenticeship) currently earn on average 1,850 Euro net per month (full-time position). What is your best guess, how much do the following groups with lower resp. higher educational attainment earn on average?
	Treatment "Incentive"	On the following screen, you will be asked another guess question. If your guess is roughly correct, you will receive an additional reward worth about half of the reward you get for participating in this survey. (next screen) Persons with a professional degree (apprenticeship) currently earn on average 1,850 Euro net per month (full-time position). What is your best guess, how much do the following groups with lower resp. higher educational attainment earn on average?
Experiment on encouraging online-search activity (2018 survey wave)		
9	Control	Based on your best guess, what is the average amount of money spent each year for a child in public schools in Germany?
	Treatment "Encouragement"	As an exception for this question, you can search the internet for the right answer to improve your guess, for example by using an internet-search engine. Based on your best guess, what is the average amount of money spent each year for a child in public schools in Germany?
Experiment on inducing experimenter-demand effects (2016 survey wave)		
10	Control	Based on your best guess, what is the average amount of money spent each year for a child in public schools in Germany?
	Treatment "Demand"	As you might know, government institutions collect a variety of key statistics about schools. We are interested in discovering whether the public is familiar with these key statistics. On the next screen, we will ask you a question about such a key statistic, to which there are correct and incorrect answers. In order for your response to be informative for us, it is very important that you answer this question as accurately as possible. Based on your best guess, what is the average amount of money spent each year for a child in public schools in Germany?

Notes: Own translation from the German original. No.: consecutive ordering of the question in the ifo Education Survey of the respective wave. Subgroup: specific control or treatment group that received the respective question.

Table A2: Who acquires additional information about the incentive scheme?

	Information acquired in experiment on	
	Earnings beliefs	School-spending beliefs
	(1)	(2)
Middle school degree	-0.016 *(0.036)*	-0.021 *(0.038)*
University entrance degree	-0.029 *(0.040)*	-0.013 *(0.042)*
Age	0.008*** *(0.001)*	0.010*** *(0.001)*
Monthly household income (1000 €)	-0.005 *(0.011)*	0.002 *(0.011)*
Female	0.039 *(0.029)*	0.025 *(0.030)*
Born in Germany	-0.039 *(0.057)*	-0.026 *(0.061)*
Partner in household	0.013 *(0.033)*	-0.056* *(0.034)*
Lives in West Germany	0.014 *(0.036)*	0.012 *(0.039)*
City size ≥ 100,000	-0.014 *(0.031)*	-0.007 *(0.030)*
Has children	0.002 *(0.034)*	0.040 *(0.034)*
Parent(s) with university degree	0.068** *(0.032)*	-0.035 *(0.033)*
Full-time employed	-0.060** *(0.030)*	-0.078** *(0.032)*
Part-time employed	-0.029 *(0.045)*	0.040 *(0.045)*
Self-employed	-0.101 *(0.070)*	-0.026 *(0.070)*
Unemployed	-0.136** *(0.054)*	0.010 *(0.065)*
Works in education sector	-0.026 *(0.042)*	0.055 *(0.049)*
Risk tolerance (11-point scale)	-0.003 *(0.006)*	0.004 *(0.006)*
Patience (11-point scale)	-0.000 *(0.007)*	-0.001 *(0.006)*
Constant	-0.010 *(0.099)*	-0.073 *(0.113)*
Observations	1,837	1,827
R^2	0.12	0.17

Notes: OLS regressions. Sample: respondents assigned to the incentive-treatment group in the respective experiment. Dependent variable: dummy variable coded one if respondent clicked on the information button. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table A3: Who reports increased effort in response to incentive provision?

	"Incentives increased my effort"	
	(Strongly) agree	(Strongly) disagree
	(1)	(2)
Middle school degree	0.024 *(0.032)*	0.006 *(0.029)*
University entrance degree	0.019 *(0.035)*	0.026 *(0.032)*
Age	-0.002** *(0.001)*	0.000 *(0.001)*
Monthly household income (1000 €)	0.002 *(0.010)*	-0.014* *(0.008)*
Female	-0.063** *(0.026)*	0.026 *(0.023)*
Born in Germany	-0.075 *(0.058)*	0.076 *(0.052)*
Partner in household	0.057** *(0.029)*	-0.013 *(0.026)*
Lives in West Germany	0.019 *(0.032)*	0.007 *(0.029)*
City size ≥ 100,000	-0.008 *(0.026)*	0.004 *(0.024)*
Has children	-0.045 *(0.030)*	0.056** *(0.027)*
Parent(s) with university degree	-0.067** *(0.027)*	0.038 *(0.026)*
Full-time employed	-0.025 *(0.028)*	0.037 *(0.025)*
Part-time employed	-0.028 *(0.038)*	0.032 *(0.036)*
Self-employed	0.074 *(0.062)*	0.018 *(0.059)*
Unemployed	-0.055 *(0.053)*	0.056 *(0.052)*
Works in education sector	0.044 *(0.040)*	-0.012 *(0.034)*
Risk tolerance (11-point scale)	0.019*** *(0.005)*	-0.008* *(0.005)*
Patience (11-point scale)	0.008 *(0.006)*	0.000 *(0.005)*
Constant	0.640*** *(0.097)*	0.163* *(0.089)*
Observations	2,752	2,752
R^2	0.04	0.01

Notes: OLS regressions. Sample: respondents assigned to the incentive-treatment group in at least one experiment. Dependent variable: column (1): dummy variable coded one if respondent "strongly agrees" or "somewhat agrees" with the statement that the prospect of earning more tokens encouraged them to put more effort in their answers; column (2): dummy variable coded one if respondent "strongly disagrees" or "somewhat disagrees" with the statement. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table A4: Heterogeneous incentive effects by sociodemographic subgroups

	Dependent variable: Belief within 10-percent range		
	Earnings beliefs without a degree	Earnings beliefs with a university degree	School-spending belief
	(1)	(2)	(3)
Educational attainment: Baseline (no/basic degree)	-0.015 *(0.032)*	-0.070* *(0.039)*	0.020 *(0.015)*
Middle school degree	0.050 *(0.042)*	0.128** *(0.050)*	0.032 *(0.021)*
University entrance degree	-0.001 *(0.042)*	0.023 *(0.020)*	0.047** *(0.022)*
Age: Baseline (below 45)	0.023 *(0.023)*	0.031 *(0.028)*	0.086*** *(0.016)*
Aged 45-65	-0.055 *(0.034)*	-0.021 *(0.041)*	-0.043** *(0.020)*
Aged above 65	-0.010 *(0.053)*	-0.155** *(0.064)*	-0.106*** *(0.025)*
Household income: Baseline (below median)	-0.010 *(0.022)*	-0.036 *(0.028)*	0.047*** *(0.012)*
Income above median	0.022 *(0.034)*	0.057 *(0.041)*	-0.006 *(0.018)*
Gender: Baseline (male)	-0.012 *(0.025)*	0.004 *(0.028)*	0.052*** *(0.014)*
Female	0.027 *(0.033)*	-0.030 *(0.041)*	-0.015 *(0.018)*
Country of birth: Baseline (not born in Ger.)	0.027 *(0.033)*	0.018 *(0.080)*	0.003 *(0.072)*
Born in Germany	-0.001 *(0.082)*	-0.032 *(0.083)*	0.043 *(0.073)*
Partner in household: Baseline (no)	-0.007 *(0.026)*	-0.017 *(0.033)*	0.062*** *(0.013)*
Partner in household	0.011 *(0.034)*	0.010 *(0.042)*	-0.030* *(0.018)*
Area of residence: Baseline (East Germany)	0.044 *(0.036)*	-0.024 *(0.045)*	0.014 *(0.015)*
West Germany	-0.056 *(0.041)*	0.016 *(0.050)*	0.037** *(0.018)*
City size: Baseline (Population<100,000)	-0.003 *(0.021)*	-0.007 *(0.025)*	0.056*** *(0.012)*
Population >=100,000	0.004 *(0.034)*	-0.019 *(0.043)*	-0.036** *(0.018)*
Children: Baseline (Has no children)	0.005 *(0.025)*	0.019 *(0.029)*	0.075*** *(0.014)*
Has children	-0.009 *(0.034)*	-0.051 *(0.041)*	-0.053*** *(0.018)*
Parental education: Baseline (Parents w/o uni. degree)	0.003 *(0.020)*	-0.007 *(0.025)*	0.045*** *(0.010)*
Parent(s) w/ university degree	-0.013 *(0.035)*	-0.014 *(0.044)*	0.000 *(0.021)*
Works in education sector: Baseline (no)	0.005 *(0.017)*	-0.020 *(0.022)*	0.041*** *(0.009)*
Works in education sector	-0.063 *(0.065)*	0.095 *(0.070)*	0.044 *(0.031)*
Risk tolerance: Baseline (below median)	-0.000 *(0.021)*	-0.013 *(0.026)*	0.039*** *(0.012)*
Risk tolerance above median	-0.001 *(0.035)*	0.005 *(0.043)*	0.016 *(0.018)*
Patience: Baseline (below median)	0.000 *(0.020)*	-0.012 *(0.024)*	0.053*** *(0.011)*
Patience above median	-0.005 *(0.037)*	-0.004 *(0.045)*	-0.027 *(0.020)*
Information acquisition: Baseline (not clicked)	-0.003 *(0.018)*	0.015 *(0.023)*	0.038*** *(0.010)*
Clicked info button about incentive scheme	0.007 *(0.026)*	-0.079** *(0.031)*	0.016 *(0.016)*

Notes: OLS regressions. Dependent variable: dummy variable coded one if stated belief is in the 10-percent interval around the true value. Coefficients in baseline rows represent incentive-treatment effects for respective sociodemographic baseline group (coefficient β_1 of equation (2)). Coefficients in other rows represent interaction terms between treatment and respective subgroup indicator (coefficient β_3 of equation (2)). Earnings-beliefs regressions control for treatment status in school-spending experiment. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table A5: Participation in the follow-up survey

	Participation in follow-up survey		
	(1)	(2)	(3)
Treatment status in main survey			
Incentive in earnings experiment	0.020 *(0.018)*		0.035 *(0.022)*
Incentive in spending experiment		0.012 *(0.018)*	0.011 *(0.018)*
Incentives in earnings and spending experiment			-0.029 *(0.025)*
Covariates			
Middle school degree	-0.000 *(0.023)*	-0.001 *(0.023)*	-0.002 *(0.023)*
University entrance degree	-0.002 *(0.028)*	-0.004 *(0.028)*	-0.004 *(0.028)*
Age	0.006*** *(0.001)*	0.006*** *(0.001)*	0.006*** *(0.001)*
Monthly household income (1000 €)	0.001 *(0.008)*	0.001 *(0.008)*	0.001 *(0.008)*
Female	0.025 *(0.020)*	0.026 *(0.020)*	0.025 *(0.020)*
Born in Germany	0.038 *(0.048)*	0.040 *(0.048)*	0.038 *(0.048)*
Partner in household	-0.021 *(0.021)*	-0.022 *(0.021)*	-0.022 *(0.021)*
Lives in West Germany	-0.004 *(0.024)*	-0.004 *(0.024)*	-0.004 *(0.024)*
City size ≥ 100,000	-0.047** *(0.020)*	-0.047** *(0.020)*	-0.048** *(0.020)*
Has children	-0.001 *(0.023)*	-0.001 *(0.023)*	-0.001 *(0.023)*
Parent(s) with university degree	-0.013 *(0.022)*	-0.013 *(0.022)*	-0.013 *(0.022)*
Full-time employed	0.073*** *(0.023)*	0.072*** *(0.023)*	0.073*** *(0.023)*
Part-time employed	0.055** *(0.027)*	0.054** *(0.027)*	0.055** *(0.027)*
Self-employed	0.019 *(0.050)*	0.020 *(0.050)*	0.020 *(0.050)*
Unemployed	0.091** *(0.042)*	0.089** *(0.042)*	0.090** *(0.042)*
Works in education sector	-0.053 *(0.035)*	-0.053 *(0.035)*	-0.052 *(0.035)*
Risk tolerance (11-point scale)	-0.014*** *(0.004)*	-0.014*** *(0.004)*	-0.014*** *(0.004)*
Patience (11-point scale)	0.006 *(0.004)*	0.006 *(0.004)*	0.006 *(0.004)*
Constant	0.373*** *(0.077)*	0.376*** *(0.077)*	0.368*** *(0.077)*
Observations	3,673	3,673	3,673
R^2	0.06	0.06	0.06

Notes: OLS regressions. Dependent variable: dummy variable coded one if respondent participated in the follow-up survey. Treatment status in the main survey: column (1): incentive-treatment in earnings experiment; column (2): incentive-treatment in spending experiment; column (3): incentive treatment in both experiments and their interaction. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table A6: Summary statistics and balancing tests of incentive experiments: Follow-up survey

	Earnings experiment			School-spending experiment		
	Control mean	Treatment mean	Difference	Control mean	Treatment mean	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Highest educational attainment						
No degree/basic degree	0.211	0.241	0.030	0.225	0.228	0.003
Middle school degree	0.404	0.389	-0.015	0.403	0.389	-0.014
University entrance degree	0.385	0.370	-0.015	0.372	0.383	0.011
Age	46.568	47.272	0.704	46.936	46.913	-0.022
Monthly household income (€)	2,403.523	2,386.787	-16.736	2,438.049	2,351.376	-86.674
Female	0.509	0.512	0.002	0.508	0.513	0.004
Born in Germany	0.950	0.963	0.013	0.954	0.959	0.005
Partner in household	0.600	0.586	-0.015	0.606	0.579	-0.027
Lives in West Germany	0.752	0.780	0.028	0.777	0.756	-0.022
City size ≥ 100,000	0.363	0.357	-0.006	0.345	0.376	0.032
Has children	0.553	0.540	-0.013	0.555	0.538	-0.017
Parent(s) with university degree	0.312	0.290	-0.021	0.306	0.296	-0.010
Employment status						
Full-time employed	0.406	0.401	-0.005	0.429	0.379	-0.050**
Part-time employed	0.138	0.125	-0.012	0.128	0.135	0.007
Self-employed	0.046	0.037	-0.008	0.046	0.037	-0.008
Unemployed	0.054	0.045	-0.009	0.046	0.054	0.008
House wife/husband/retired/ill/student	0.059	0.064	0.005	0.062	0.061	-0.001
Works in education sector	0.082	0.090	0.008	0.090	0.081	-0.009
Risk tolerance (11-point scale)	4.362	4.328	-0.034	4.335	4.355	0.020
Patience (11-point scale)	6.216	6.283	0.067	6.199	6.301	0.102
Item non-response: Beliefs	0.002	0.002	0.000	0.010	0.006	-0.005
Observations	1,251	1,284		1,275	1,260	

Notes: Group means. "Difference" displays the difference in means between the respective control and treatment groups. Significance levels of "Difference" stem from linear regressions of the background variables on the respective treatment dummies. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table A7: Incentive effects on earnings beliefs in follow-up survey

	Belief relative to true value	Distance to true value	Belief within incentivized range	Belief within 10-percent range	Extreme values below 100	Confidence	Response time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Beliefs on earnings without a degree							
Incentive	0.018	3.501	0.031	0.049***	0.010*		
	(0.036)	*(49.583)*	*(0.019)*	*(0.019)*	*(0.006)*		
Control mean	0.87	333.55	0.66	0.28	0.02		
Observations	2,418	2,418	2,418	2,418	2,418		
R^2	0.001	0.000	0.002	0.003	0.001		
Panel B: Beliefs on earnings with a university degree							
Incentive	0.009	53.055	-0.023	-0.006	0.008	0.078	6.278
	(0.032)	*(82.548)*	*(0.017)*	*(0.020)*	*(0.006)*	*(0.058)*	*(6.463)*
Control mean	1.01	706.83	0.79	0.44	0.02	3.77	52.50
Observations	2,418	2,418	2,418	2,418	2,418	2,417	2,419
R^2	0.000	0.000	0.001	0.001	0.001	0.002	0.000

Notes: OLS regressions. Dependent variable (recorded in follow-up survey conducted about two weeks after the main survey): column (1): stated belief divided by true value; column (2): absolute distance between stated belief and true value; column (3): dummy variable coded one if stated belief is in the incentivized range; column (4): dummy variable coded one if stated belief is in the 10-percent interval around the true value; column (5): dummy variable coded one if stated belief is below 100; column (6): confidence about belief on seven-point Likert scale (1="very unsure", 7="very sure"); column (7): difference between actual and predicted response time. Randomized experimental treatment "Incentive": respondents offered monetary incentive for belief accuracy. Control mean: mean of the outcome variable for the control group in the follow-up survey. Panel A: beliefs about net average monthly earnings of full-time employed persons without any professional degree. Panel B: beliefs about net average monthly earnings of full-time employed persons with a university degree. Since both earnings beliefs were elicited on the same screen, confidence and response time were recorded only once for earnings beliefs in general. Regressions control for treatment status in the incentive experiment on school-spending beliefs. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017.

Table A8: Summary statistics and balancing tests of additional experiments

	Experiment on encouraging online-search activity			Experiment on inducing experimenter-demand effects		
	Control mean (1)	Treatment mean (2)	Difference (3)	Control mean (4)	Treatment mean (5)	Difference (6)
Highest educational attainment						
No degree/basic degree	0.368	0.381	0.014	0.401	0.380	-0.021
Middle school degree	0.300	0.308	0.008	0.293	0.315	0.022
University entrance degree	0.332	0.311	-0.022	0.306	0.305	-0.001
Age	50.202	50.908	0.706	50.702	50.321	-0.381
Monthly household income (€)	2,451.623	2,466.528	14.905	2,120.379	2,200.880	80.501
Female	0.505	0.517	0.011	0.530	0.497	-0.033
Born in Germany	0.958	0.962	0.004	0.944	0.954	0.010
Partner in household	0.589	0.575	-0.014	0.556	0.570	0.014
Lives in west Germany	0.684	0.651	-0.033	0.803	0.796	-0.008
City size ≥ 100,000	0.307	0.323	0.016	0.328	0.315	-0.013
Has children	0.586	0.592	0.007	0.592	0.601	0.009
Parent(s) with university degree	0.286	0.291	0.005	0.243	0.243	0.000
Employment status						
Full-time employed	0.390	0.376	-0.013	0.343	0.334	-0.009
Part-time employed	0.160	0.156	-0.003	0.122	0.128	0.006
Self-employed	0.053	0.050	-0.002	0.030	0.031	0.001
Unemployed	0.027	0.032	0.005	0.051	0.060	0.009
House wife/husband/retired/ill/student	0.371	0.385	0.014	0.454	0.448	-0.006
Works in education sector	0.107	0.108	0.001	0.078	0.079	0.001
Risk tolerance (11-point scale)	4.391	4.238	-0.152	4.312	4.252	-0.060
Patience (11-point scale)	6.215	6.336	0.121	5.961	5.941	-0.020
Item non-response: Beliefs	0.005	0.003	-0.002	0.081	0.073	-0.008
Observations	1,979	2,067		1,642	1,660	

Notes: Weighted group means. "Difference" displays the difference in means between the respective control and treatment groups. Significance levels of "Difference" stem from linear regressions of the background variables on the respective treatment dummies. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: columns (1)-(3): ifo Education Survey 2018; columns (4)-(6): ifo Education Survey 2016.

Table A9: Stacked estimation of incentive and encouragement effects on school-spending beliefs

	Belief relative to true value	Distance to true value	Belief within incentivized range	Belief within 10-percent range	Extreme values below 100	Confidence	Response time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.112	-720.657	0.132***	0.044***	-0.064***	0.768***	86.068**
	(0.078)	*(470.255)*	*(0.021)*	*(0.009)*	*(0.015)*	*(0.068)*	*(38.010)*
Treatment x Wave 2018	0.001	232.146	-0.045	0.010	0.066***	-0.153	-61.217
	(0.101)	*(595.175)*	*(0.031)*	*(0.014)*	*(0.019)*	*(0.101)*	*(38.662)*
Wave 2018	-0.074	-812.739**	0.077***	0.007	-0.082***	0.162**	-78.467***
	(0.068)	*(384.815)*	*(0.022)*	*(0.008)*	*(0.014)*	*(0.064)*	*(5.495)*
Control mean	0.74	7175.06	0.41	0.03	0.12	2.50	44.46
Observations	7,713	7,713	7,713	7,713	7,713	7,726	7,742
R^2	0.00	0.00	0.02	0.01	0.01	0.05	0.02

Notes: OLS regressions. Dependent variable: beliefs about average annual public school spending per student; column (1): stated belief divided by true value; column (2): absolute distance between stated belief and true value; column (3): dummy variable coded one if stated belief is in the incentivized range; column (4): dummy variable coded one if stated belief is in the 10-percent interval around the true value; column (5): dummy variable coded one if stated belief is below 100; column (6): confidence about belief on seven-point Likert scale (1="very unsure", 7="very sure"); column (7): difference between actual and predicted response time. Randomized experimental treatment "Treatment": wave 2017: respondents offered monetary incentive for belief accuracy; wave 2018: respondents encouraged to search the internet for the correct answer. Control mean: mean of the outcome variable for the control group. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** p<0.01. ** p<0.05. * p<0.10. Data source: ifo Education Survey 2017 and 2018.