

# Limited Self-Knowledge and Survey Response Behavior

*Armin Falk, Thomas Neuber, Philippe Strack*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# Limited Self-Knowledge and Survey Response Behavior

## Abstract

We study response behavior in surveys and show how the explanatory power of self-reports can be improved. First, we develop a choice model of survey response behavior under the assumption that the respondent has imperfect self-knowledge about her individual characteristics. In panel data, the model predicts that the variance in responses for different characteristics increases in self-knowledge and that the variance for a given characteristic over time is non-monotonic in self-knowledge. Importantly, the ratio of these variances identifies an individual's level of self-knowledge, i.e. the latter can be inferred from observed response patterns. Second, we develop a consistent and unbiased estimator for self-knowledge based on the model. Third, we run an experiment to test the model's main predictions in a context where the researcher knows the true underlying characteristics. The data confirm the model's predictions as well as the estimator's validity. Finally, we turn to a large panel data set, estimate individual levels of self-knowledge, and show that accounting for differences in self-knowledge significantly increases the explanatory power of regression models. Using a median split in self-knowledge and regressing risky behaviors on self-reported risk attitudes, we find that the  $R^2$  can be multiple times larger for above- than below-median subjects. Similarly, gender differences in risk attitudes are considerably larger when restricting samples to subjects with high self-knowledge. These examples illustrate how using the estimator may improve inference from survey data.

JEL-Codes: C830, D830, C910, D910, J240.

*Armin Falk*

*briq, University of Bonn / Germany*

*armin.falk@briq-institute.org*

*Thomas Neuber*

*University of Bonn / Germany*

*thomas.neuber@uni-bonn.de*

*Philipp Strack*

*Yale University / New Haven / CT / USA*

*philipp.strack@yale.edu*

We thank Roland Bénabou, Philipp Eisenhauer, Botond Köszegi, and participants at various conferences and seminars for helpful comments. We thank Markus Antony for excellent administrative support. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866. Funding by the DFG through CRC TR 224 (Project A01) is gratefully acknowledged. Philipp Strack was supported by a Sloan fellowship.

# 1 Introduction

Survey evidence is a major source of knowledge in the social sciences, including economics. With growing interest in measuring cognitive and non-cognitive skills—such as economic preferences, beliefs, attitudes, and values—survey evidence is gaining increasing relevance in economics (Heckman, Stixrud, and Urzua, 2006; Almlund et al., 2011; Falk et al., 2018). This paper provides a method to improve the explanatory power of subjective survey data. The method is derived from a simple model of survey response behavior that allows identifying more vs. less informative respondents based only on *patterns* of their response behavior. Hence, this paper makes two main contributions: it offers a framework for modeling and understanding survey response behavior in general and it derives a method to empirically identify more or less reliable answers, which in turn helps to improve the explanatory power of survey measures.

As a first step, we derive a simple choice model of survey response behavior. In the model, we are serious about the idea that when being asked to report an individual characteristic such as a preference, belief, or some non-cognitive skill, a respondent has to make herself the object of her own self-assessment and makes a *choice*. We assume that there exists a true type (level of each characteristic) but that the respondent is not perfectly aware of her true type. This *limited self-knowledge* is modeled as an imperfect signal that the respondent receives about her true type. Differences in self-knowledge can capture the fact that individuals vary in their capacity to retrieve or memorize relevant information about themselves, engage more or less in reflecting who they are, or that some people simply lack life experience in the domain of interest. We further assume that the respondent wants to minimize the squared distance between her true type and her report, i.e., the interests of the respondent and the researcher are aligned. Conditional on the informativeness of the signal, our agent’s Bayesian optimal report is a weighted sum of the population mean of the respective characteristic and her signal. The more informative the signal, the greater the weight placed on the signal relative to the population mean. We analyze the expected variance of respondents’ answering behavior conditional on the informativeness of the signal, both over time and between characteristics. We find that the variance *between* characteristics increases in the informativeness of the signal, which mirrors the fact that the more confident a respondent is about her answer, the more she deviates in expectation from the population mean. In contrast, the *within* variance—the variance of responses for a given characteristic over time—is non-monotonic in the signal precision. The intuition is that response behavior is stable over time if a person knows herself either very well or not at all. This result cautions against the use of simple stability to measure the accuracy of signals and reports. Importantly, we show that the ratio of the variance between characteristics and the variance over time (for given characteristics) is equal to the informativeness of the signal. This key result implies that we can use

observed variances to estimate individual differences in self-knowledge or the accuracy of respective reports.

We provide several extensions of the model and discuss their implications for expected response behavior. Our first extension relaxes the assumption of exogenous signals and explores the consequences of endogenous precision. We derive an expression for the choice of signal precision and discuss implications for how the quality of survey responses reacts to incentives. Second, we relax the assumption that respondents are perfectly aware of the signal strength, i.e., how well they know themselves. Instead, we allow for subjective levels of self-knowledge that are higher or lower than actual self-knowledge. While subjective beliefs about self-knowledge affect the distribution of responses, we show that they do not impede the identification of differences in self-knowledge, simply because they cancel out. Third, we allow for individual-specific scale use, i.e., a tendency to report either rather extreme or moderate answers. Again, we show that scale use affects responses but that the identification of self-knowledge remains unchanged. Finally, we relax the assumption that respondents want to report their type truthfully. Instead, we allow for response biases arising from motives such as social desirability or image effects. We study the implications of such motives and show that respondents act similarly as in the case of subjective scale use.

The second step in the paper is to use the theoretical results in empirical applications—especially the insight that the precision of signals about types can be inferred from the ratio of the between- and the within-variance. We first show that self-knowledge can be estimated using a closed-form estimator before discussing results from a laboratory experiment designed to test the main predictions of the model. Subsequently, we analyze representative panel data to show how accounting for signal precision affects empirical results and explained variance.

To derive an estimator of signal precision—or self-knowledge—from panel data, we essentially consider the ratio between two sample variances, namely the between-variance (the variance of responses between items) and the within-variance (the variance for a given item over time). These are the sample analogs to our theoretically derived variances. We study the asymptotic properties of the estimator and formally show its consistency as well as unbiasedness. Using simulations, we illustrate the performance of the estimator for realistic sample sizes. We study various combinations of the number of respondents, survey items, and waves (periods), respectively. The estimator generally performs well. For example, for 100 respondents, 15 items, and three waves, the rank correlation between the estimated and the true level of self-knowledge is 0.76.

To empirically test the main predictions of the model, it is crucial to observe responses and compare them with respondents' *true* types. However, this is difficult—if not impossible—with typical survey data. Therefore, we ran a laboratory experiment that creates a panel data set with types that are imperfectly known to subjects but perfectly

known to the researcher. In particular, subjects in the experiment were paid to accurately report the sizes of 60 male figures shown to them on separate computer screens. This setup allows us to observe subjects' reports and compare them with the respective true types. Results from the experiment confirm the main predictions derived from the model. First, subjects' reports are biased towards the mean, i.e., small sizes are, on average, overestimated, and large sizes are typically underestimated. Second, subjects who are estimated to be more informative actually provide more accurate reports. Based on the estimates, we split the sample and regress reports on true types. We find that the regression coefficient for the above-median sample is about 2.5 times as large as the respective coefficient for below-median subjects and that the explanatory power in terms of  $R^2$  is about five times as large. Third, we use the experiment to create random variation in signal precision. For this purpose, we randomized subjects into one of two treatments: a Long-treatment in which they saw the figures for 7.5 seconds each, and a Short-treatment in which each figure was presented only for 0.5 seconds. We show that we can use our empirical estimates to predict subjects' treatment status, i.e., we are able to predict whether subjects were assigned to the treatment condition with high or with low signal precision.

Finally, we apply our estimator to a large representative panel data set, the German Socio-economic Panel (SOEP). We provide several examples to illustrate how the suggested estimates of self-knowledge can help to increase the explanatory power of regressions based on self-reports. In particular, we use a fifteen-item Big Five personality inventory from multiple waves of the SOEP to estimate self-knowledge. Using these estimates, we form two sub-samples: one with above- and one with below-median values of estimated self-knowledge, respectively. As an illustrative example, we choose the context of risk attitudes, which has received a lot of attention in the literature. We study both determinants and consequences of risk attitudes, measured on an eleven-point Likert scale. To illustrate, we find that the gender effect on the general willingness to take risks is substantially larger for the above-median sample than for the below-median sample. Moreover, the difference in  $R^2$  between the two sub-samples amounts to 36%. Likewise, when we regress the likelihood of receiving performance pay as part of one's compensation on the willingness to take risks, the explained variance ( $R^2$ ) is 238% higher in the above-median sample than in the below-median sample.

Our paper is related to multiple strands of the literature. As we take the informational constraints of the agent seriously and study their choice implications, we relate to the work on rational inattention (Sims, 1998, 2003; Caplin and Dean, 2015; Matějka and McKay, 2015; Caplin et al., 2020). This literature focuses on flexible information acquisition and studies what type of information is acquired in a single-agent setting. Our goal is different, and we analyze how to identify agents' levels of information in a situation with many agents who share a common prior. Our framework enables analyzing the provision of incentives in surveys as studied, e.g., in Prelec (2004) and Cvitanić et al. (2017) as

well as how contextual factors such as social desirability affect survey responses (see, e.g., Bénabou et al., 2020; Chen et al., 2020). The notion of limited self-knowledge and its economic consequences for the labor market has been studied in Falk, Huffman, and Sunde (2006a, 2006b). The model is also related to work on preferences for consistency, as modeled and tested in Falk and Zimmermann (2017) and applied to survey methodology in Falk and Zimmermann (2013).

Moreover, the paper contributes to the literature on measurement error in surveys (for an overview, see Bound, Brown, and Mathiowetz, 2001). For the case of classical measurement error—where deviations in answers are independent of the respective true value—, instrumental variables techniques are capable of removing bias. Recently, Gillen, Snowberg, and Yariv (2019) have suggested measuring duplicate instances of control variables and using them as mutual instruments. Hyslop and Imbens (2001) consider a model that is related to ours where an agent observes a Normal signal and reports his best estimate of an underlying variable of interest. They analyze the effect of the resulting non-classical measurement error on regression coefficients but do not consider remedies. The focus of our paper is to estimate the precision of the agent’s signal, which allows placing higher weight on subjects with better self-knowledge.

Drerup, Enke, and Gaudecker (2017) estimate a structural model of stock market participation that identifies individuals for whom relevant preferences and beliefs have increased explanatory power. Alternative approaches to deal with measurement error in subjective survey data use structural estimation techniques to recover underlying primitives and choice models, finding that accounting for measurement error yields greater predictive power (Kimball, Sahm, and Shapiro, 2008; Beauchamp, Cesarini, and Johannesson, 2017).<sup>1</sup> Despite not referring to qualitative survey measures, a related contribution comes from Beauchamp et al. (2020), who analyze how accounting for the “compromise effect”—whereby subjects’ answers tend towards the center of the provided scale—, can improve estimates for risk preference.

The remainder of the paper proceeds as follows. Section 2 develops the model with its basic framework and extensions. Building upon its insights, Section 3 introduces the estimator, presents its theoretical properties, and explores its performance in finite samples. Section 4 presents the stylized laboratory experiment. In Section 5, we apply the estimator to a large and representative panel and explore its implications for improving estimates. Finally, Section 6 concludes.

---

<sup>1</sup>In the psychology literature, processes that underlie response behavior have been studied under the label of *cognitive aspects of survey methodology* (see Sudman, Bradburn, and Schwarz, 1996; Bradburn, Sudman, and Wansink, 2004; Schwarz, 2007). Broadly, our paper is also related to classical test theory and item response theory (see, e.g., Edwards, 2009; Kyllonen and Zu, 2016; Bolsinova, Boeck, and Tijmstra, 2017).

## 2 Model

In this section, we first introduce a simple framework to model the answering process in surveys, based on limited self-knowledge. Second, we derive how patterns in answering behavior reveal the informational content of responses, providing the intuition for how we later estimate self-knowledge. Finally, we present various extensions of the baseline model to study further important aspects of the answering process and show the robustness of our identification approach.

**Introspection and Self-knowledge.** The context that we are interested in is a simple survey situation. A researcher asks a respondent (or agent) a question about a specific characteristic, e.g., some preference, personality trait, or belief.<sup>2</sup> The agent’s true type is denoted by  $\theta$ , and we assume that it is normally distributed in the population with mean  $\bar{\theta}$  and variance  $\sigma^2$ . Agents act upon their true types but vary with respect to how well they know their type. Hence, when asked about her type  $\theta$ , the respondent does not perfectly know herself but instead engages in a process of introspection. The outcome of this process is an informative but noisy signal  $x$  about her true type. The signal is normally distributed with a mean equal to the agent’s type  $\theta$  and variance  $\sigma^2/\tau$ . The parameter  $\tau > 0$  hence indicates the precision of the signal relative to the variance in the population. The higher the value of  $\tau$ , the more precise the signal that an individual receives about herself. We refer to  $\tau$  as *self-knowledge*.

**Response Behavior.** After reflecting on her true type  $\theta$ , the respondent reports her answer. We assume that she seeks to provide a response  $r$  that is as precise as possible, i.e., the interests of the researcher and respondent are perfectly aligned.<sup>3</sup> Formally, the respondent uses her signal  $x$  to provide a response  $r$  that minimizes the expected quadratic distance to her unknown true type, i.e.,

$$u_{\theta}(r) = -(r - \theta)^2. \quad (1)$$

Hence, she reports her best guess of her type  $r = \mathbb{E}[\theta | x]$ . The respondent’s prior equals the distribution of types in the population with mean  $\bar{\theta}$ . Substituting for the expected value of her posterior belief about her type, we obtain by Bayes’ Rule that

$$r = \frac{\bar{\theta} + \tau x}{1 + \tau}. \quad (2)$$

---

<sup>2</sup>For example, the researcher may ask the respondent to state her willingness to take risks, her level of agreeableness or conscientiousness, or her belief about her internal or external locus of control.

<sup>3</sup>For many interview situations, we think that this is a valid assumption. However, there are contexts in which respondents may want to strategically signal a specific type that is actually different from their belief about their true type for reputational or “social desirability” reasons. For a discussion, see Section 2.2.4.



Intuitively, the higher her self-knowledge  $\tau$ , the more precise the respondent's signal, and the more weight she puts on her signal relative to the population mean  $\bar{\theta}$ . In the limit, if she knows nothing about herself, her best estimate is to report the mean of her prior, whereas if she knows herself perfectly, she disregards the prior completely.

This concludes our basic framework. The model defines a mapping from true types to distributions over observable responses, taking into account the notion of limited self-knowledge. In the next subsection, we study how response patterns can be used to identify differences in self-knowledge.

## 2.1 Response Patterns

We now explore the implications of limited self-knowledge for response patterns. We are particularly interested in the variances in reports, both unconditional and conditional on an agent's type. These variances will allow us to identify differences in self-knowledge. In Section 3, we will build on these insights when we derive an estimator for an individual's level of self-knowledge in panel data.

**Expected Report.** It follows from Equation (2) that the expected report conditional on the true type  $\theta$  equals

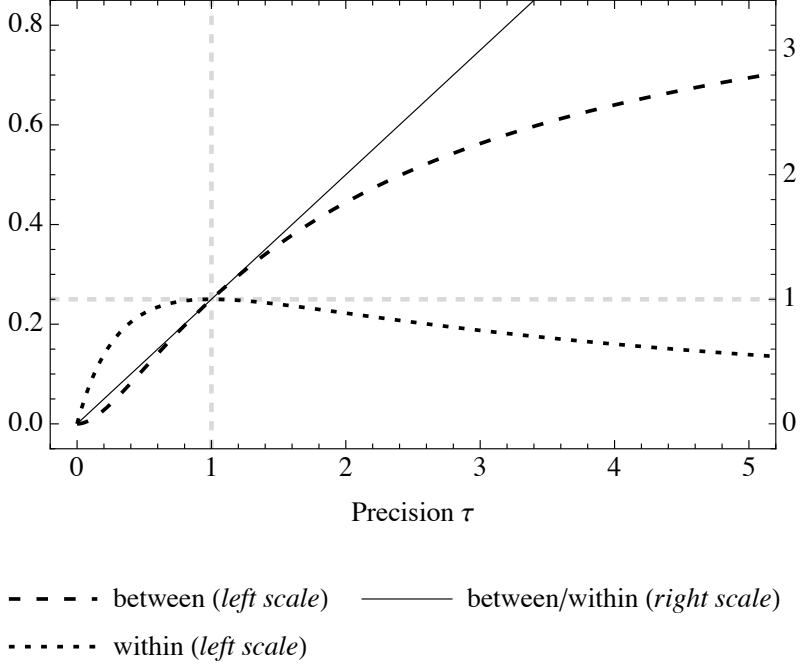
$$\mathbb{E}[r \mid \theta] = \frac{\bar{\theta} + \tau \theta}{1 + \tau}. \quad (3)$$

For low values of self-knowledge  $\tau$ , the expected report is close to the population mean  $\bar{\theta}$ , irrespective of the true type  $\theta$ . For large values of  $\tau$ , the expected report converges to the true type  $\theta$ .

**Between-variance.** Consider now the variance of conditional expected reports. In the context of panel data, one can think of this theoretical quantity as an approximation of the variance in average reports concerning different characteristics. Following this interpretation (as the variance between different characteristics), we refer to it as the *between-variance*. It is given by

$$\begin{aligned} \sigma_{\text{between}}^2 &:= \text{var}(\mathbb{E}[r \mid \theta]) = \text{var}\left(\frac{\bar{\theta} + \tau \theta}{1 + \tau}\right) \\ &= \left(\frac{\tau}{1 + \tau}\right)^2 \text{var}(\theta) = \left(\frac{\tau}{1 + \tau}\right)^2 \sigma^2. \end{aligned} \quad (4)$$

The between-variance is strictly increasing in self-knowledge  $\tau$ . This reflects the fact that agents with high levels of self-knowledge put relatively little weight on their prior. Instead, they provide reports that tend to deviate from the population mean.



*Note:* Variances  $\sigma_{\text{between}}^2$  and  $\sigma_{\text{within}}^2$  as functions of  $\tau$  (values on the left axis). The solid line shows the ratio of the two variances, which is equal to  $\tau$  (values on the right axis).

Figure 1: Theoretical variances

**Within-variance.** Now consider the variance conditional on an agent’s type. This theoretical quantity can be thought of as the variation in responses of an agent responding multiple times to questions about the *same* characteristic. We call this variation the *within-variance* of the agent’s reports. It is given by

$$\begin{aligned}
 \sigma_{\text{within}}^2 &:= \text{var}(r \mid \theta) = \text{var}\left(\frac{\bar{\theta} + \tau x}{1 + \tau} \mid \theta\right) \\
 &= \left(\frac{\tau}{1 + \tau}\right)^2 \text{var}(x \mid \theta) = \frac{\tau}{(1 + \tau)^2} \sigma^2.
 \end{aligned} \tag{5}$$

The relationship between self-knowledge  $\tau$  and the within-variance is non-monotonic. For very low levels of  $\tau$ , the variance is low, simply because the respondent refers to her prior. As  $\tau$  increases, the variance increases as more weight is placed on the noisy signal. However, as  $\tau$  further increases, the variance decreases because the signal about the true type becomes increasingly precise. From a researcher’s perspective, this pattern implies that stable responses—i.e., similar responses regarding the same characteristics over time—do not necessarily indicate high levels of self-knowledge and precision. The most stable responses come from respondents who know themselves perfectly—or who do not know themselves at all.

Figure 1 illustrates the relationship between the two variances and self-knowledge. It plots the between-variance (long dashes) and the within-variance (short dashes) as

functions of self-knowledge  $\tau$ . As  $\tau$  goes to zero, both variances converge to zero. This means that the respondent provides the same answer (equal to the prior) to any question. As  $\tau$  increases, the respondent places higher weight on her signal, which increases both the within- and between-variance. At  $\tau = 1$ , i.e., when the signal  $x$  is exactly as informative as the respondent’s prior knowledge about the population, the within-variance reaches its maximum and is equal to the between-variance. Beyond this point, the between-variance further increases and ultimately converges to the variance of true types in the population,  $\sigma^2$ . At the same time, the within-variance strictly decreases and converges to zero, because a respondent with perfect self-knowledge will always provide exactly the same report for a given characteristic.

Both the between- and within-variance contain information about the respondent’s level of self-knowledge  $\tau$ . While a large between-variance is always “good news,” indicating high levels of  $\tau$ , a low within-variance can reflect either high or low levels of  $\tau$ , respectively. However, considering both variances *jointly* perfectly reveals the level of self-knowledge. In fact, the ratio of the between- and within-variance equals the degree of self-knowledge:

$$\frac{\sigma_{\text{between}}^2}{\sigma_{\text{within}}^2} = \frac{\left(\frac{\tau}{1+\tau}\right)^2 \sigma^2}{\frac{\tau}{(1+\tau)^2} \sigma^2} = \tau. \quad (6)$$

The respective relationship is also shown in Figure 1 where, for each level of  $\tau$ , the thin solid line plots the ratio of the two variances.

Our paper builds on this insight. We show that the relationship between the variances and self-knowledge is robust to various extensions of the model, construct a finite sample estimator based on this relationship, and show that this estimator indeed predicts the informativeness of subjects’ responses both in lab and field data.

## 2.2 Extensions

In this subsection, we study four extensions of the basic framework. The purpose of this exercise is twofold. The first aim is to show that the framework enables integrating additional important aspects of the survey response process in a meaningful way. In particular, we consider the role of costly introspection, deviations of subjective self-knowledge from actual self-knowledge, subjective scale use, and social desirability issues. Second, we show that for the extensions studied, the result that self-knowledge  $\tau$  can be inferred as the ratio of the between- and within-variance is robust.

### 2.2.1 Endogenous Precision

Our first extension considers endogenous precision. So far, we have modeled the process of introspection as receiving an exogenous signal with a fixed relative precision  $\tau$ . However, the cognitive process of introspection requires mental effort, and a respondent has to decide

how much effort to invest. For example, the agent chooses how long and intensively she engages in recollecting past behaviors to extract her type and how carefully she evaluates and maps information into a response. We assume that the variance of the signal  $x$  is no longer fixed at a given level of  $\sigma^2/\tau$ . Instead,  $\tau$  is chosen by the agent at a cost  $\tau/a$  for some ability  $a \in \mathbb{R}_+$ . The utility function (corresponding to Equation (1)) equals

$$u_\theta(r, \tau) = -m(r - \theta)^2 - \frac{\tau}{a}.$$

Here,  $m \in \mathbb{R}_+$  measures the motivation of the respondent to provide an accurate answer, and it can be thought of as either extrinsic or intrinsic motivation.<sup>4</sup> Assume that  $ma > \sigma^{-2}$ , as, otherwise, incentives are too weak to motivate any effort and a precision of zero is optimal.

**Lemma 1.** *The respondent’s optimal precision is given by  $\tau^* = \sqrt{ma}\sigma - 1$ .*

The chosen signal precision  $\tau^*$  is increasing in both incentives  $m$  and ability  $a$ , i.e., a higher level of incentives or ability generates more precise signals. The proof of Lemma 1 is provided in Appendix A.

In the presence of endogenous effort, subjects giving high- vs. low-quality answers can be distinguished by the exact same response patterns as for the case with exogenous self-knowledge. However, the interpretation changes, as differences may now reflect differences in motivation  $m$  or ability  $a$ . In fact, the model predicts that the higher the incentives, the more reliable and informative the responses. This is exactly the rationale for paying subjects in economic experiments (Smith, 1976; Camerer and Hogarth, 1999) and similar attempts to incentivize survey responses as, e.g., Prelec’s Bayesian Truth Serum (Prelec, 2004). In addition, differences may reflect motivational dispositions (e.g., mood, fatigue, boredom) or fundamental differences in “introspection ability”  $a$ , such as cognitive skills, memory, and recollection capabilities.

### 2.2.2 Subjective Self-knowledge

The basic framework assumes that the respondent knows the relative precision  $\tau$  of her signal  $x$ . In other words, she perfectly knows how well she knows herself and weighs her signals accordingly. However, a large body of evidence has shown that individuals often misperceive their own knowledge and skills (Camerer and Lovallo, 1999; Malmendier and Tate, 2005). Applied to our context, respondents may be over-confident and place too much weight on their signal  $x$ , or they are under-confident and place too much weight on the prior. In either case, this will result in a wedge between the optimal and the actual response, again potentially complicating inference about respondents’ true types.

---

<sup>4</sup>The former could reflect, e.g., monetary or social approval incentives, while the latter may capture motives such as a desire to respond truthfully and accurately or simply an interest in (promoting) research.

To model potential biases in perceived self-knowledge, we introduce subjective self-knowledge  $\tilde{\tau}$ . A respondent has correct beliefs about her self-knowledge if  $\tilde{\tau} = \tau$ , she is under-confident if  $\tilde{\tau} < \tau$ , and she is over-confident if  $\tilde{\tau} > \tau$ . We assume that the agent is naive and that when determining her survey response, she applies relative weights according to her subjective self-knowledge  $\tilde{\tau}$ . Equation (2) changes as follows:

$$r = \frac{\bar{\theta} + \tilde{\tau} x}{1 + \tilde{\tau}}$$

Corresponding to Equation (4), the between-variance becomes

$$\sigma_{\text{between}}^2 = \text{var}(\mathbb{E}[R | \theta]) = \left( \frac{\tilde{\tau}}{1 + \tilde{\tau}} \right)^2 \sigma^2.$$

Hence, the variability in answers between different items reflects the respondent's subjective self-knowledge but is independent of self-knowledge itself. Intuitively, as the between-variance is based only on the expected response, which is independent of the true precision of the agent's signal  $\tau$ , the variance is also independent of the true precision of the agent's signal.

This is different for the within-variance, corresponding to Equation (5).

$$\sigma_{\text{within}}^2 = \text{var}(r | \theta) = \left( \frac{\tilde{\tau}}{1 + \tilde{\tau}} \right)^2 \frac{\sigma^2}{\tau}.$$

The latter depends on both subjective self-knowledge as well as actual self-knowledge. Intuitively, the within-variance of responses is affected by the respondent's subjective self-knowledge  $\tilde{\tau}$  through the weight that she places on her signal and by her self-knowledge  $\tau$  through the variance of the signal.<sup>5</sup>

Importantly, the result from Equation (6) about the ratio of the two variances still holds.

$$\frac{\sigma_{\text{between}}^2}{\sigma_{\text{within}}^2} = \frac{\left( \frac{\tilde{\tau}}{1 + \tilde{\tau}} \right)^2 \sigma^2}{\left( \frac{\tilde{\tau}}{1 + \tilde{\tau}} \right)^2 \frac{\sigma^2}{\tau}} = \tau$$

Hence, while deviations from correct beliefs about the precision of one's signals affect expected response behavior in general, inference about  $\tau$  remains feasible.

### 2.2.3 Subjective Scale Use

Empirical research typically assumes that individuals who want to express the same level of agreement or disagreement with respect to a particular survey item will respond in the exact same way. For example, two respondents intending to express the exact same willingness to take risks on a Likert scale would be expected to choose the exact same answer

---

<sup>5</sup>Observe that only for  $\tilde{\tau} \rightarrow \infty$ , the model predicts classical measurement error.

category. However, if response scales are subjectively interpreted, responses may differ. Hence, the mapping from an intended response to some scale may depend on individual-specific notions of how to express a given level of agreement or disagreement. We suggest a simple way how to model this kind of subjective scale use and show that it affects responses in general but not the estimation approach for  $\tau$  suggested by Equation (6).

In particular, assume that an agent has arrived at her *intended* report and now needs to map it to an *actual* report  $r$  on an answering scale. This mapping may be individual-specific in the sense that some agents may use more “extreme” answers while others use more “moderate” answers to express the same information. For a given intended response, therefore, two agents may come up with different actual responses. We assume that the agent’s response is scaled away from some point  $c \in \mathbb{R}$ , e.g., the center of the scale, by a factor  $\phi \in (0, 1]$ . The report and its expected value (corresponding to Equations (2.2.3) and (3), respectively) are then given by

$$r = (1 - \phi) c + \phi \left( \frac{\bar{\theta} + \tau x}{1 + \tau} \right) \quad \text{and} \quad \mathbb{E}[r \mid \theta] = (1 - \phi) c + \phi \left( \frac{\bar{\theta} + \tau \theta}{1 + \tau} \right).$$

Depending on  $\phi$ , actual responses may thus be pushed towards the center of the scale, rendering the interpretation of responses more difficult. This holds in particular if  $\phi$  is systematically correlated with underlying types (such as preferences) or group characteristics under study (such as gender or socioeconomic status).

The between-variance (corresponding to Equation (4)) becomes

$$\begin{aligned} \sigma_{\text{between}}^2 &= \text{var}(\mathbb{E}[R \mid \theta]) = \text{var} \left( (1 - \phi) c + \phi \frac{\bar{\theta} + \tau \theta}{1 + \tau} \right) \\ &= \phi^2 \left( \frac{\tau}{1 + \tau} \right)^2 \text{var}(\theta) = \phi^2 \left( \frac{\tau}{1 + \tau} \right)^2 \sigma^2, \end{aligned}$$

and the within-variance (corresponding to Equation (5)) becomes

$$\begin{aligned} \sigma_{\text{within}}^2 &= \text{var}(r \mid \theta) = \text{var} \left( (1 - \phi) c + \phi \frac{\bar{\theta} + \tau x}{1 + \tau} \mid \theta \right) \\ &= \phi^2 \left( \frac{\tau}{1 + \tau} \right)^2 \text{var}(x \mid \theta) = \phi^2 \frac{\tau}{(1 + \tau)^2} \sigma^2. \end{aligned}$$

We see that both variances increase quadratically in the scale use parameter  $\phi$ . However, for the ratio of the two, the effect of scale use cancels out, and it still holds that the ratio equals  $\tau$ .

$$\frac{\sigma_{\text{between}}^2}{\sigma_{\text{within}}^2} = \frac{\phi^2 \left( \frac{\tau}{1 + \tau} \right)^2 \sigma^2}{\phi^2 \frac{\tau}{(1 + \tau)^2} \sigma^2} = \tau$$

### 2.2.4 Social Desirability Effects

In some situations, respondents might not want to truthfully report their type but rather provide an answer that is deemed socially desirable. These contexts are likely to arise if the interview situation is not anonymous (audience effects) and/or if items are image relevant. For example, it is plausible that a respondent feels more comfortable reporting that she is an honest rather than a dishonest person. Such concerns can be integrated into our framework by adding a desirable answer  $d \in \mathbb{R}$ . Respondents' objective now is to minimize the weighted sum of the squared distances to their type and the desirable answer, respectively. The utility function is thus

$$u_{\theta,d}(r) = -(1 - \psi)(r - \theta)^2 - \psi(r - d)^2,$$

where  $\psi \in [0, 1]$  measures the intensity of the preference to report  $d$ . The optimal report of a respondent equals the weighted sum of the best guess of her type  $\theta$  and the desirable answer

$$r = (1 - \psi) \left( \frac{\bar{\theta} + \tau x}{1 + \tau} \right) + \psi d.$$

The respondent thus acts as if subject to subjective scale use, as introduced in Section 2.2.3. The main difference between subjective scale use and desirability arises in the context of multiple agents and characteristics: while the scale use parameters  $(\phi, c)$  are naturally agent-specific, the desirability parameters  $(\psi, d)$  are naturally specific to the characteristic.

## 3 Estimator

In this section, we derive an estimator for an individual's level of self-knowledge that is based on the insights from Section 2. We consider a panel data set comprising  $I > 1$  agents and  $T > 1$  waves. In each wave  $t$ , each agent  $i$  answers an identical set of  $K > 1$  questions about distinct, time-invariant characteristics, traits, or beliefs. We denote by  $\theta_{ik}$  the value of the  $k^{\text{th}}$  characteristic for agent  $i$  and assume that characteristics are independently normally distributed in the population with mean  $\bar{\theta}$  and variance  $\sigma^2$ . In contemplating the answer to question  $k$  in wave  $t$ , agent  $i$  generates a signal  $x_{ikt}$  that she uses to form her answer  $r_{ikt}$ . The signal  $x_{ikt}$  is normally distributed with mean  $\theta_i$  and variance  $\sigma^2/\tau_i$ , independent of all other signals, such that the optimal response is given by

$$r_{ikt} = \frac{\bar{\theta} + \tau_i x_{ikt}}{1 + \tau_i}.$$

Given the  $K \times T$  answers observed for each agent  $i$ , the objective of a researcher is to estimate agents' levels of self-knowledge  $\tau_i$ . In Section 2, we have shown that  $\tau$  equals the (the-

oretical) variance among expected answers to different questions (between-variance) divided by the (theoretical) variance among answers to the same questions (within-variance). To construct an estimator  $\hat{\tau}_i$ , we use the sample variance between average answers for different characteristics as an approximation of the true between-variance and the average sample variance of answers for a given characteristic as an approximation of the true within-variance. Denote agent  $i$ 's average answer to question  $k$  by  $\bar{r}_{ik} = \frac{1}{T} \sum_{t=1}^T r_{ikt}$  and her average answer over all questions by  $\bar{r}_i = \frac{1}{K} \sum_{k=1}^K \bar{r}_{ik}$ . Our estimator  $\hat{\tau}_i$  for the self-knowledge of agent  $i$  is given by

$$\hat{\tau}_i = \frac{\frac{1}{K-1} \sum_{k=1}^K (\bar{r}_{ik} - \bar{r}_i)^2}{\frac{1}{K(T-1)-2} \sum_{k=1}^K \sum_{t=1}^T (r_{ikt} - \bar{r}_{ik})^2} - \frac{1}{T}. \quad (7)$$

The numerator in the first summand of the expression captures the variation *between* the average answers of an agent for different characteristics, while the denominator expresses the average variation in answers *within* characteristics. Since the expected value of the ratio of two random variables is not the same as the ratio of their respective individual expected values, the denominator is adjusted by a constant factor relative to the unbiased estimator of the within-variance<sup>6</sup> and a correction term of  $1/T$  is subtracted from the ratio. These two adjustments are necessary to ensure that the estimator is unbiased.

The following theorem establishes that  $\hat{\tau}_i$  is a consistent, unbiased estimator of self-knowledge  $\tau_i$  and describes its properties.

**Theorem.** *For every  $K, T$  that satisfy  $K(T-1) > 4$ .*

1. *The estimator  $\hat{\tau}_i$  satisfies*

$$\hat{\tau}_i = \left( \tau_i + \frac{1}{T} \right) \frac{K(T-1)-2}{K(T-1)} F_i - \frac{1}{T} \quad (8)$$

*for some random variable  $F_i$  that is  $F$  distributed with  $K-1, K(T-1)$  degrees of freedom for every fixed vector of parameters  $\tau_i, \sigma, \bar{\theta}$ .*

2.  *$\hat{\tau}_i$  is an unbiased estimator for  $\tau_i$ , i.e.,  $\mathbb{E}[\hat{\tau}_i | \tau_i] = \tau_i$ .*

3. *The standard error of the estimator  $\hat{\tau}_i$  is given by*

$$\sqrt{\mathbb{E}[(\hat{\tau}_i - \tau_i)^2 | \tau_i]} = \left( \tau_i + \frac{1}{T} \right) \sqrt{\frac{2((K-1) + K(T-1) - 2)}{(K-1)(K(T-1) - 4)}}. \quad (9)$$

4.  *$\hat{\tau}_i$  is a consistent estimator and converges to  $\tau_i$  at the rate  $1/\sqrt{K}$  in the number of attributes, and for all  $K > 4$  it satisfies the following upper bound independent of*

---

<sup>6</sup>An unbiased estimator of the within-variance is given by  $\frac{1}{K(T-1)} \sum_{k=1}^K \sum_{t=1}^T (r_{ikt} - \bar{r}_{ik})^2$ .



the number of repeated observations  $T$ :

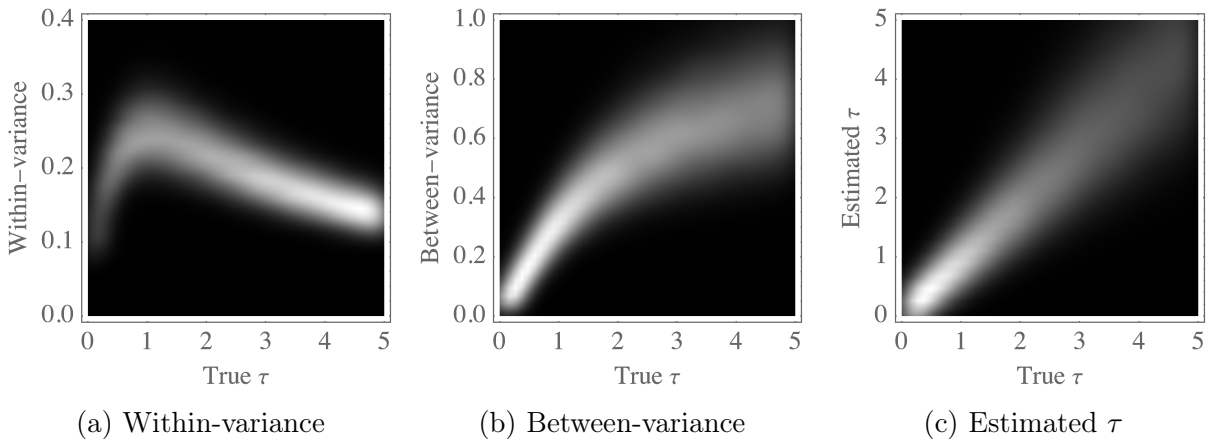
$$\sqrt{\mathbb{E}[(\hat{\tau}_i - \tau_i)^2 | \tau_i]} \leq \frac{2\tau_i + 1}{\sqrt{K - 4}}$$

The proof of the theorem is provided in Appendix A. Part 4 of the theorem shows that for retrieving precise estimates, additional questions are more valuable than additional waves. This is the case because, intuitively, having additional questions adds to the precision of estimating both the between as well as the (average) within-variance, whereas additional waves only improve the precision of the estimated within-variance. Therefore, as  $K$  goes to infinity, the estimator converges to the true value even for just two waves, while the precision of the estimator is always limited for a finite number of questions.

**Remark.** *As we show in the proof of the theorem in Appendix A, the properties of the estimator extend unchanged to the model with endogenous effort, subjective self-knowledge, and subjective scale use. We state the properties here without these extensions for ease of exposition.*

Next, we illustrate our model and the behavior of the estimator using numerical simulations. For all illustrations, agents' levels of self-knowledge  $\tau_i$  are drawn from a uniform distribution with support  $[0.1, 5]$ , and we abstract from subjective scale use and subjective self-knowledge. The true average value of characteristics  $\bar{\theta}$  is set to 5 and the true population variance  $\sigma^2$  equals 1.

Figure 2 displays the joint distribution of the true level of self-knowledge  $\tau_i$  and the sample within-variance, the sample between-variance, and estimated self-knowledge  $\hat{\tau}_i$ , respectively. For the within-variance, we observe the expected non-monotonic, hump-shaped relationship with the true level of self-knowledge (Figure 2a). The estimates for



*Note:* Kernel-density estimates, where lighter shading corresponds to a higher estimated density. Each panel is based on the same 100 simulations, each with  $I = 1,000$  hypothetical individuals, for whom reports about  $K = 50$  characteristics are observed  $T = 3$  times. The panels use Gaussian kernels with bandwidth selection according to Silverman's rule.

Figure 2: Simulations

Table 1: Accuracy of estimates for different sample sizes

	(1)	(2)	(3)	(4)	(5)
$I$ (respondents)	100	10,000	100	100	100
$K$ (characteristics)	15	15	50	15	50
$T$ (waves)	3	3	3	10	10
Correlation	0.68	0.68	0.87	0.76	0.91
Rank correlation	0.76	0.77	0.90	0.82	0.93
Median split	80%	80%	88%	83%	90%

the between-variance increase in the true level of self-knowledge, but heavily “fan out” for higher levels of true self-knowledge (Figure 2b). Our proposed estimator for self-knowledge is strongly concentrated around the 45-degree line and thus highly informative about agents’ true levels of self-knowledge (Figure 2c).

In Table 1, we illustrate how the estimator performs for various sample sizes. We consider 100 or 10,000 agents, 15 or 50 characteristics, and 3 or 10 waves, respectively. For each scenario, we run 10,000 simulations and report the average value of three measures for the quality of the estimates: Pearson’s correlation and Spearman’s rank correlation between estimated and true self-knowledge and the proportion of simulated agents correctly identified as having a level of self-knowledge above or below the value of the median. If our estimator had no informational value at all, we would expect a correlation and rank correlation of zero and 50% of correctly-assigned agents in the median split.

The values of the correlation and the rank correlation coefficients of 0.68 and 0.76 shown in Column 1 for  $I = 100$ ,  $K = 15$ , and  $T = 3$  suggest that the estimator is already informative about self-knowledge for modest sample sizes. This is confirmed by 80% of hypothetical agents being assigned to the correct half of the sample in terms of self-knowledge. In Column 2, the number of hypothetical agents is increased to 10,000. The quality of predictions remains almost exactly unchanged, reflecting the fact that our estimator does not use population information. However, as can be seen from Column 3, estimates strongly benefit from a larger number of characteristics (50 instead of 15), in line with Part 4 of the theorem. Relative to these increases, the increase in performance from a higher number of answers per characteristic in Column 4 (ten instead of three) is not quite as large (in line with Part 4 of the theorem, which shows that the standard error does not vanish in  $T$ ). Column 5 combines the number of characteristics from Column 3 with the number of waves from Column 4, reaching the best performance, with correlation coefficients above 0.9 and a median split result of 90%. In sum, we find that the estimator performs reasonably well with a modest number of fifteen characteristics and three waves, and its performance can be increased, in particular, by a larger number of characteristics.

## 4 Experimental Evidence

This section presents experimental evidence to provide an empirical test of the model’s main predictions. The idea of the experiment is to create a choice environment where the researcher observes subjects’ reports (allowing to estimate  $\tau$ ) while at the same time *knowing the true state*  $\theta$ . Accordingly, we can study whether our estimator is successful in identifying subjects whose reports are relatively more informative than those of others. In addition, we *exogenously* vary the quality of the signals that subjects receive about true types. In particular, we run two treatments with either high or low signal quality and test whether our estimator of  $\tau$  is capable of predicting subjects’ treatment status, i.e., whether a subject received high- or low-quality signals. Such tests are difficult—if not impossible—with non-experimental data, where true states are unknown to the researcher and the precision of signals cannot be exogenously varied.

### 4.1 Design of the Experiment

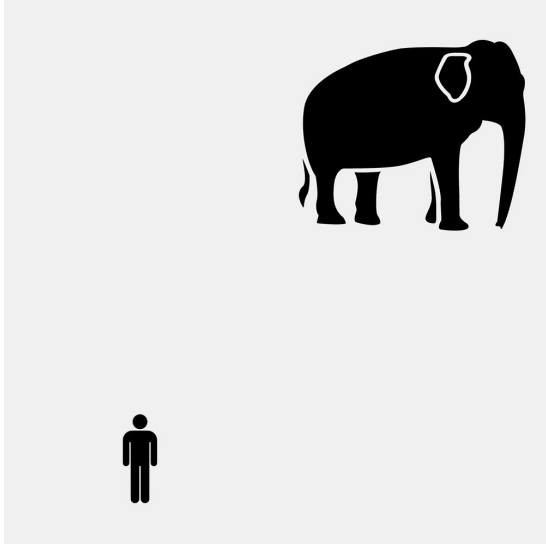
To create a choice environment with known types  $\theta$  and an exogenous variation in knowledge  $\tau$ , the experiment exposed subjects to a simple, repeated, and incentivized estimation task. The setup mimics a panel data set where respondents are repeatedly asked to respond to a set of different questions.

**Types.** The requirement that the researcher knows true types implies that we cannot work with individual characteristics such as personality traits, preferences, or IQ, simply because these cannot be known with certainty. To implement types known to the researcher ( $\theta_i$ ), we thus presented subjects a series of abstract figures. In particular, subjects saw a total of 60 screens, each showing a stylized male figure of varying size (see Figure 3). On each screen, the figure was randomly located at one of four different parts of the screen, i.e., at either the upper left, the upper right, the lower left, or the lower right part of the screen, respectively. The sizes of the figures were drawn from a normal distribution that closely matches the actual height distribution of men in Germany (based on data from the Socio-economic Panel, SOEP). In particular, sizes were matched into eleven size categories (in meters) with likelihoods as shown in Table 2. For example, Category 3

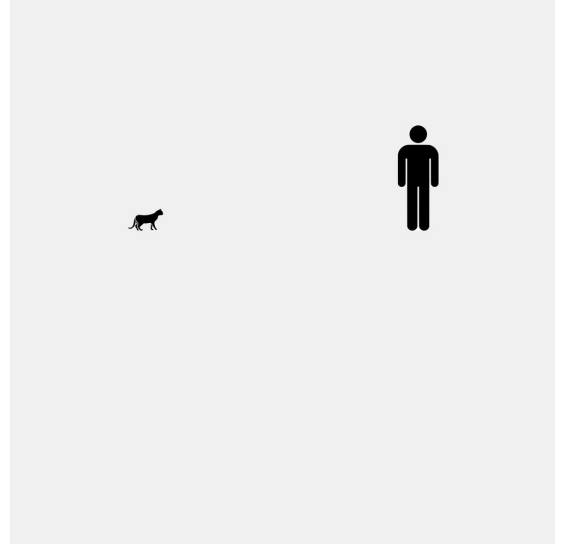
Table 2: Choice categories

0	1	2	3	4	5	6	7	8	9	10
<1.56	1.56– 1.60	1.61– 1.65	1.66– 1.70	1.71– 1.75	1.76– 1.80	1.81– 1.85	1.86– 1.90	1.91– 1.95	1.96– 2.00	>2.00
0.1%	0.8%	3.8%	11.1%	21.1%	26.1%	21.1%	11.1%	3.8%	0.8%	0.1%

*Note:* top row: categories; middle row: sizes (in meters); bottom row: respective likelihoods.



(a) Size within Category 2



(b) Size within Category 8

*Note:* The panel on the left shows a male figure with a height of 1.63m along with the elephant, which is 3.50m tall. The male figure on the right corresponds to 1.93m, and the cat has a height of 40cm. Animal pictograms adapted with permission from Storey (2016).

Figure 3: Example screens

represents male persons of sizes between 1.66 and 1.70 meters, occurring with a likelihood of 11.1%. Subjects received a handout showing this distribution and the corresponding figures underneath (see the instructions in the Online Appendix).

Subjects were informed that a total of 15 distinct sizes were independently drawn from the eleven categories and shown four times. Specifically, subjects saw four blocks, each comprising these 15 distinct sizes. This procedure hence implements a panel structure, i.e., for every subject  $i$ , we observe a total of 60 reports for  $K = 15$  characteristics in  $T = 4$  periods. The location of the male figures was randomly determined for each screen.

To facilitate the estimation task and vary the presentation style of the screens, next to the male figure, subjects also saw a “reference category,” i.e., either an elephant or a cat (see Figure 3). Subjects were informed that—unlike for the male figures—the size of the two animals was always exactly the same. The height of the elephant was 3.50 meters, and it was 0.40 meters for the cat. Conditional on the randomly determined location of the male figures, the location and type of the reference category (elephant or cat) were also randomly drawn for each screen.

**Payoff Function.** Subjects had an incentive to estimate the shown size of the male figure as precisely as possible. The payoff function,  $\pi$ , implements a quadratic loss function and corresponds exactly to Equation (1) in the model, with

$$\pi(r) = -(r - \theta)^2,$$

where  $\theta$  indicates the true type (size of the male figure) and  $r$  a subject's report. For the payoff, one of the 60 screens was randomly selected. For the selected screen and respective report, subjects received €10 minus the product of €0.10 and the squared difference between the true type and the report. For example, if a subject was shown a male figure of size Category 1 (1.56 meters – 1.60 meters) and estimated a size according to Category 8 (1.91 meters – 1.95 meters), the subject received  $\text{€}10 - (1 - 8)^2 \times \text{€}0.10 = \text{€}5.10$ . Note that we chose an endowment of €10 to rule out losses even if the difference between the true and the estimated type was maximal.

**Signal Precision and Treatments.** To exogenously vary the precision  $\tau$  of the signal, we ran two treatments that only differed in terms of how long subjects saw each of the 60 screens. In the treatment *Long*, subjects saw each screen for 7.5 seconds, in contrast to treatment *Short*, where they saw each screen only for 0.5 seconds. Treatments were randomly assigned within each lab session. Each subject participated in one treatment condition only.

**Procedural Details.** 199 subjects—mostly undergraduate university students from all majors—took part in the experiment, 101 subjects in the treatment *Long* and 98 in the treatment *Short*. We used z-Tree as the experimental software (Fischbacher, 2007). Subjects were recruited using the software hroot (Bock, Baetge, and Nicklisch, 2014). At the beginning of an experimental session, participants received detailed information about the rules and the structure of the experiment. In all treatments, the experiment only started after all participants had correctly answered several control questions. The experiments were run at the BonnEconLab in May 2019. For participation, subjects received a show-up fee of €5.

## 4.2 Hypotheses and Results

Our experimental data are well suited for testing several hypotheses derived from our model:

**Hypothesis 1.** *Average reports are linear in true types and biased towards the population average of the true types, i.e., towards five.*

The first hypothesis follows from an optimal report being the weighted sum of the population average  $\bar{\theta}$  and the received signal  $x$  (see Equation (2)). It can only be tested because, in our experiment, we know the true type. Graphically, we would expect average reports for different true types to lie on a straight line that is rotated clockwise around the point (5, 5), i.e., we would expect upward bias for small true values, no bias for average true values, and downward bias for large true values.

Hypothesis 1 uses that knowledge  $\tau$  is finite for any subject. Hypotheses 2–4 additionally exploit individual-specific information about  $\tau$ , either in terms of treatment differences (Short vs. Long) or using the estimator introduced in Section 3.

**Hypothesis 2.** *Estimates  $\hat{\tau}$  are larger for the subjects in the Long-treatment than for those in the Short-treatment.*

An implication of Hypothesis 2 is that the estimates for  $\tau$  should have reasonable power for predicting subjects’ treatment status. Thus, we expect that we can blindfold ourselves regarding the treatment status and be able to tell only from the patterns in answers to which treatment a given subject was assigned.

Regardless of which approach is used to make inferences about  $\tau$  (the treatment status or the estimator), the following further hypothesis should hold.

**Hypothesis 3.** *The lower subjects’ level of knowledge  $\tau$ , the stronger the reports’ bias towards the average value of the characteristic, i.e., five.*

This hypothesis is a refinement of Hypothesis 1. It states that when estimating figure sizes, subjects realize and take into account their *individual-specific* level of  $\tau$ , which may reflect ability or treatment status.

**Hypothesis 4.** *The higher the level of  $\tau$  in a given population, the stronger the predictive power of reports for true types.*

Hypothesis 4 is our main hypothesis. It states, in particular, that using reports of subjects for whom we have high values of  $\hat{\tau}$  yields higher explanatory power of reports in comparison to using either all subjects or subjects with low levels of  $\hat{\tau}$ .

Figure 4a provides a visual test of Hypothesis 1. It plots true types against observed reports, pooled for both treatments. Gray bubbles represent average reports for given true types, with their sizes reflecting the respective number of observations (which is largely determined by the sampling distribution). Relative to the dotted 45-degree line, the fitted ordinary least squares (OLS) line is rotated clockwise around the point (5, 5). Its slope of 0.279 is significantly smaller than one, i.e., answers are biased towards the population average (see Column 1 of Table 3 below for details).

To test the further hypotheses, we apply the estimator from Equation (7) to our experimental data. Recall that a given subject saw each of the sizes that were drawn for her exactly four times. Therefore, we treat the respective four answers given by a subject as referring to the same characteristic. We hence observe  $K = 15$  characteristics and  $T = 4$  waves.<sup>7</sup> Figure 4b shows the distribution of  $\hat{\tau}$ , separately for the Short and the Long-treatment (gray and transparent, respectively).

---

<sup>7</sup>The estimator uses the information that, e.g., signals 3, 18, 33, and 48 showed the same true type, but it does *not* use the information what that type was.

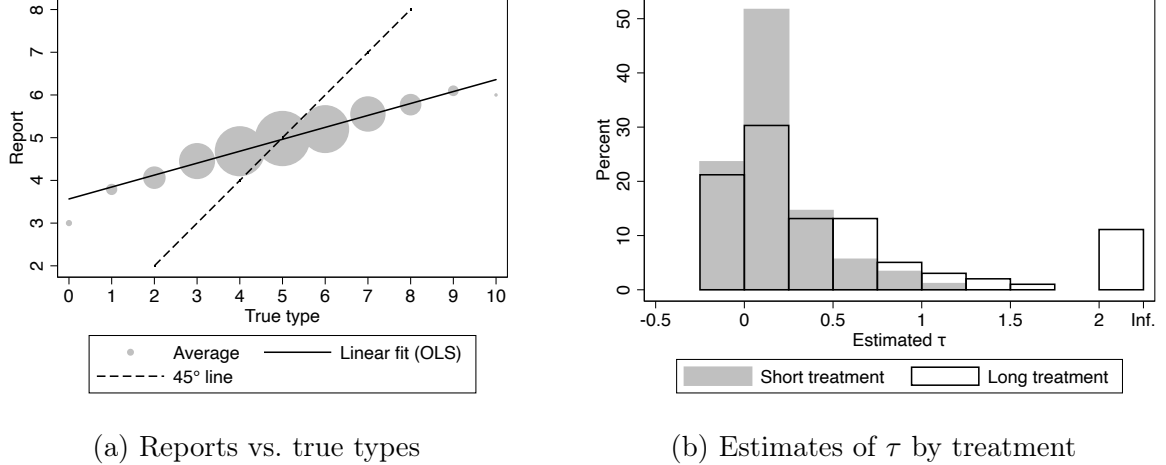


Figure 4: Results from the experiment

In support of Hypothesis 2, estimates of  $\tau$  are higher for subjects in the Long-treatment than for those in the Short-treatment ( $p < 0.001$ , Mann–Whitney  $U$  test). Conversely, this implies that our estimator predicts subjects’ treatment status. A simple probit regression of an indicator variable for the Long-treatment on our estimates for  $\tau$  yields a significant positive coefficient value (average marginal increase in the predicted probability = 0.37;  $p < 0.001$ , two-sided).

For the tests of Hypotheses 3 and 4, we turn to Table 3. Column 1 corresponds to the fitted line shown in Figure 4a, regressing reports on true types within the full sample. Columns 2 and 3 replicate Column 1 separately for the two treatments, Short and Long. In comparison with the pooled sample, the slope is flatter for the Short and steeper for the Long-treatment. The three possible pairwise differences in slopes (full sample, Short-treatment, and Long-treatment) are all statistically significant ( $p < 0.001$ , two-sided). This is in line with a successful treatment manipulation of  $\tau$  and with Hypothesis 3. In Columns 4 and 5, we split the sample by  $\hat{\tau}$ . As predicted, for subjects with above-median values of  $\hat{\tau}$ , the estimated coefficient for the relationship between reports and true types is larger than for below-median subjects (Column 4) and the whole sample (Column 1). Again, all three possible pairwise differences are statistically significant ( $p < 0.001$ , two-sided).<sup>8</sup>

To test Hypothesis 4, which states that the predictive power of reports for true types should increase in a population’s level of  $\tau$ , we again draw on Table 3 and compare the  $R^2$ -values within the two pairs of sub-samples (Columns 2–5). The data confirm our hypothesis: relative to the Short-treatment, the value of  $R^2$  in the Long-treatment is more than doubled (comparison of Columns 2 and 3;  $p < 0.001$ , two-sided). For the two sub-

<sup>8</sup>Note that the difference between the sub-samples in Columns 4 and 5 is more pronounced than the one between Columns 2 and 3: the coefficient in the low- $\tau$  sub-sample (Column 4) is smaller than the one for the Short-treatment in Column 2 ( $p = 0.060$ , two-sided), and the high- $\tau$  coefficient in Column 5 is larger than the Long-treatment coefficient in Column 3 ( $p = 0.045$ , two-sided).

Table 3: Relationship between reports and true types

Subjects	<i>Dependent variable: Report</i>				
	<i>all</i>	<i>by treatment</i>		<i>by <math>\hat{\tau}</math></i>	
		<i>Short</i>	<i>Long</i>	<i>low</i>	<i>high</i>
	(1)	(2)	(3)	(4)	(5)
True type	0.279*** (0.0167)	0.190*** (0.0169)	0.366*** (0.0260)	0.162*** (0.0138)	0.404*** (0.0250)
Constant	3.565*** (0.0799)	3.973*** (0.0859)	3.177*** (0.122)	4.100*** (0.0741)	3.012*** (0.116)
Observations	11940	5880	6060	5640	5640
Clusters	199	98	101	94	94
$R^2$	0.134	0.0793	0.189	0.0491	0.243
$\Delta R^2$		139%, $p < 0.001$		394%, $p < 0.001$	

*Note:* The table reports OLS estimates. The sample underlying Columns 4 and 5 excludes eleven subjects for whom there exists no variation in answers and, therefore, no estimates for  $\tau$  are available. The  $p$ -values for the respective sizes of  $\Delta R^2$  are each based on 10,000 permutations (Heß, 2017). Standard errors clustered at the subject level in parentheses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

samples based on the estimator  $\hat{\tau}$ , the difference is even larger (comparison of Columns 4 and 5): the  $R^2$  for the above-median sample is about five times as large as the respective  $R^2$  for the below-median subjects ( $p < 0.001$ , two-sided). In addition to supporting our hypothesis, these comparisons show that the estimates  $\hat{\tau}$  are more informative than knowledge about subjects' treatment status. This is remarkable, given that our estimator only uses the pattern of subjects' responses. We conclude this section with a discussion about two further analyses (i) using individual-level data and (ii) using survey items on the quality of answers.

**Individual-level Data.** Recall that each subject in the experiment made 60 estimation decisions. This means that we can run regressions of these 60 reports on the respective true states *separately for each individual*. The resulting individual-specific value of  $R^2$  is informative about how well a subject is able to discriminate between different true states, and it is therefore informative about  $\tau$ . Moreover, the individual slope parameter reveals how much weight is assigned to signals, and thus it is informative about the level of subjective knowledge, or confidence,  $\tilde{\tau}$ . Several observations can be made. First, in individual-level regressions, the values of  $R^2$  and the slope parameters are strongly positively related, with a rank correlation of 0.83 ( $p < 0.001$ , two-sided,  $N = 188$ ).<sup>9</sup> This positive correlation supports the central assumption of the model that agents with more knowledge (measured in terms of  $R^2$ ) place more weight on their signals (measured in

<sup>9</sup>As in Table 3, the eleven subjects for whom there exists no variation in answers (all of them always chose the answer “5”) have to be excluded.



terms of coefficients). Second, the individual-level values of  $R^2$  allow us to further test the validity of our estimator from Section 3, the latter *not* using information about the true types. We find that the individual values of  $R^2$  are strongly correlated with the values of  $\hat{\tau}$ : the rank correlation is 0.83 ( $p < 0.001$ , two-sided,  $N = 188$ ). In fact, this relationship can be analyzed even more thoroughly. In light of our model, the  $R^2$ -values can be transformed into alternative estimates of  $\tau$  according to the formula  $\hat{\tau}_{\text{alt.}} = R^2 / (1 - R^2)$ .<sup>10</sup> The Pearson correlation between the alternative estimates and our main estimates  $\hat{\tau}$  is 0.98 ( $p < 0.001$ , two-sided,  $N = 188$ ). This finding is not mechanic, since the identification approaches behind the two estimators rest on entirely different information in the data: the  $R^2$ -based measure uses the information about true states, while our main estimator only uses information about which of the states are identical across the four waves.

**Survey Items on Self-knowledge.** We have argued that accounting for differences in (self-)knowledge can help to improve estimates, and we have suggested an estimator based on the pattern of behavior. An alternative to using this estimator could be to simply ask respondents directly how accurate or reliable they think their responses are. The use of such survey items appears to be fairly common. At the end of the experiment, we asked two such items and can compare their discriminatory power to that of our estimates  $\hat{\tau}$ . In particular, we asked subjects “how difficult” they thought the estimation task had been and “how sure” they were about their answers. The answers to both questions were provided on seven-point Likert scales. Reassuringly, responses to these two items are strongly negatively correlated ( $\rho = -0.59$ ;  $p < 0.001$ , two-sided). To obtain a single measure, we take the first principal component of these two items. The rank correlation between this measure of *self-reported precision* and our estimate of knowledge  $\hat{\tau}$  is only 0.05 and statistically insignificant ( $p = 0.46$ , two-sided,  $N = 188$ ). However, the rank correlation between self-reported precision and the individual-level values of  $R^2$  is also just 0.08 ( $p = 0.29$ , two-sided,  $N = 188$ ), i.e., very small and, in particular, much smaller than the respective correlation of 0.83 between  $R^2$  and  $\hat{\tau}$ . These results suggest that—in contrast to our estimator—the survey items of self-reported precision contain only very limited information.

## 5 Applications

In this section, we apply our estimator to data from the German Socio-economic Panel (SOEP), a large, representative panel data set. The main objective is to show that by using estimates of self-knowledge,  $\hat{\tau}$ , we can increase the explanatory power of regressions that involve self-reports. In particular, we estimate  $\tau$  using answers to the Big Five personality inventory from multiple waves and split the relevant samples by the respective

---

<sup>10</sup>For the derivation, see the last paragraph of Appendix C.1.

median levels of  $\hat{\tau}$ , exactly as it was done in Section 4 for the data from the experiment (see Table 3). We illustrate differences in explanatory power ( $R^2$ ) between the resulting sub-samples in the context of risk preferences, using self-reported measures of individual willingness to take risks. Following recent work on consequences and determinants of risk preferences, we use the preference measures to explain economic outcomes (with risk measures on the right-hand side) and explore determinants such as gender (with risk measures on the left-hand side).

## 5.1 Data and Measures

Our measure of self-knowledge is constructed using the fifteen-item Big Five inventory that was included in the 2005, 2009, 2013, and 2017 waves of the SOEP (Gerlitz and Schupp, 2005). The respective questions are particularly suitable for our purposes since they are meant and designed to cover independent traits that are stable over time (see, e.g., Cobb-Clark and Schurer, 2012). We use the maximum number of waves available for a given respondent, i.e., two waves for 47.4%, three waves for 22.1%, and four waves for 30.4% of the respondents ( $N = 21,157$ ). The estimator introduced in Section 3 assumes that types are identically distributed for different characteristics. Empirically, however, the means and variances of answers might differ for different characteristics. Therefore, we add the following modification to our estimation procedure:

1. We construct normalized responses  $n_{ikt}$  as the difference between agent  $i$ 's response  $r_{ikt}$  and the average response  $\bar{r}_k$ , divided by the standard deviation  $s_k$  of agents' average responses  $\bar{r}_{ik}$  for the given characteristic  $k$ .

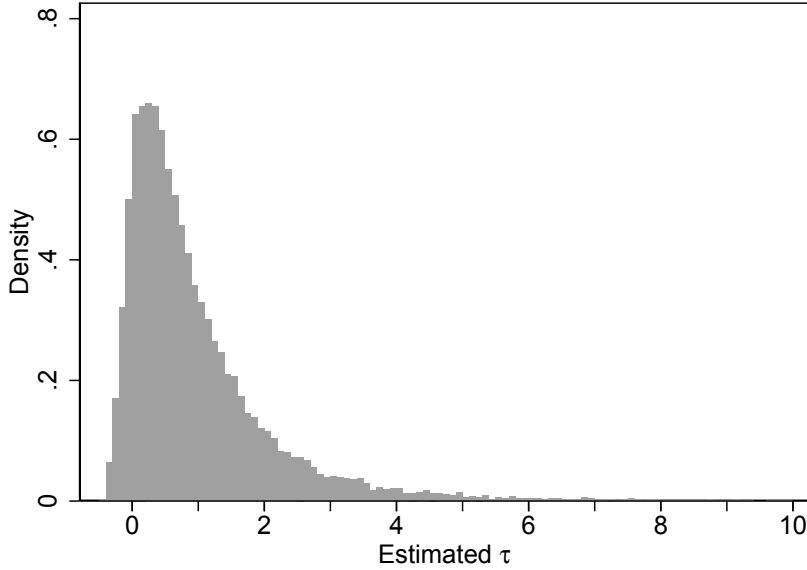
$$n_{ikt} = \frac{r_{ikt} - \bar{r}_k}{s_k}, \quad \text{with} \quad s_k = \sqrt{\frac{1}{I-1} \sum_{i=1}^I (\bar{r}_{ik} - \bar{r}_k)^2}$$

2. Analogous to Equation (7), we use the standardized answers to apply the following estimator.

$$\hat{\tau}_i^{POP} = \frac{\frac{1}{K-1} \sum_{k=1}^K (\bar{n}_{ik} - \bar{n}_i)^2}{\frac{1}{K(T-1)-2} \sum_{k=1}^K \sum_{t=1}^T (n_{ikt} - \bar{n}_{ik})^2} - \frac{1}{T} \quad (10)$$

As we show in Appendix B.1, for  $I \rightarrow \infty$ , this population-based estimator retains the properties that were stated in the theorem in Section 3. In Appendix B.2, we also consider the case that characteristics are correlated, with results showing that the estimator remains informative.

Figure 5 shows the empirical distribution of  $\hat{\tau}$  in the SOEP sample. We see considerable variation in these estimates, suggesting substantial heterogeneity in latent self-knowledge. The median value is 0.64, and for about 66% of respondents, the estimate  $\hat{\tau}$  is smaller than one.



*Note:* Distribution of  $\hat{\tau}$  in the German SOEP. Estimates that are larger than ten (48 out of 21,157) are not displayed.

Figure 5: Distribution of  $\hat{\tau}$  in the SOEP

The main focus of this section is to show how empirical relationships between non-cognitive skills and economic outcomes are attenuated due to limited self-knowledge. However, the concept of self-knowledge might also be interpreted as an individual trait, i.e., an interesting object in itself: high or low self-knowledge can be thought of as an integral part of one's personality, reflecting individual differences in life experience, cognitive skills, or parental influence. Before turning to the main analyses, we therefore briefly consider potential determinants of  $\tau$ , treating it as an individual trait.

In Table 4, we present results from regressions of estimated self-knowledge  $\hat{\tau}$  on a set of plausibly exogenous determinants, in particular gender and age, as well as education. As shown in Column 1, self-knowledge is very weakly correlated with gender, with an  $R^2$  of virtually zero. With respect to age, Column 2 reveals a hump-shaped relationship with self-knowledge. Descriptively, the latter increases until the age of about 43 years and then declines. However, the coefficients and the values of  $R^2$  are fairly small. Given that self-knowledge might reflect differences in cognitive skills, we also consider an association with education (see Column 3). The correlation is significant and indicates that one more year of education is, on average, associated with an increase of about 0.06 in the level of self-knowledge. In Column 4, we regress estimated self-knowledge simultaneously on all of the previously considered variables. Education seems to dominate, as becomes apparent when comparing the values of  $R^2$  between the columns. However, even the combined  $R^2$  of 0.033 is fairly low, suggesting that the estimates of self-knowledge contain much information above and beyond socio-demographic characteristics.

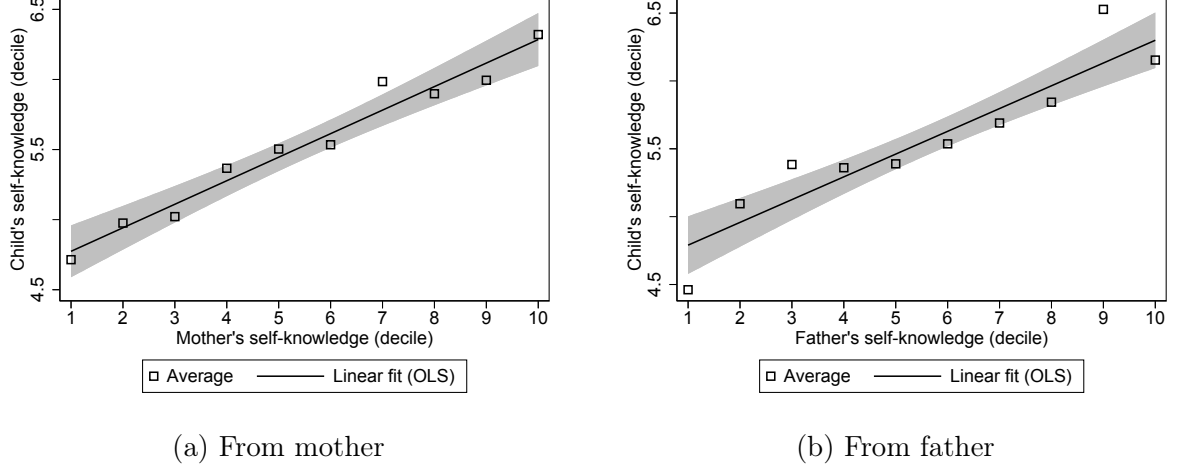
If self-knowledge is a trait, it might be intergenerationally transmitted, similar to,

Table 4: Correlations with  $\tau$ 

	<i>Dependent variable: <math>\hat{\tau}</math></i>			
	(1)	(2)	(3)	(4)
Female	-0.0286* (0.0139)			-0.00105 (0.0151)
Age (in '11)		0.0126*** (0.00215)		0.00822** (0.00269)
Age <sup>2</sup> (in '11)		-0.000148*** (0.0000207)		-0.0000971*** (0.0000254)
Edu. years (in '11)			0.0623*** (0.00292)	0.0598*** (0.00296)
Constant	0.926*** (0.0103)	0.696*** (0.0523)	0.139*** (0.0359)	0.0338 (0.0747)
Observations	20946	20946	16158	16158
$R^2$	0.000	0.004	0.031	0.033

*Note:* The table reports OLS estimates. Individuals for whom  $\hat{\tau}$  lies above the 99<sup>th</sup> percentile are excluded. Age as well as years of education refer to the year 2011, i.e., the center of the relevant time interval (2005–2017). Heteroskedasticity-robust standard errors in parentheses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

e.g., risk aversion, trust, patience, and social preferences (Dohmen et al., 2012; Kosse and Pfeiffer, 2012; Alan et al., 2017; Kosse et al., 2020). Such transmission could come from various sources, e.g., imitation, exposure to similar social environments, or genetic dispositions. Among the SOEP participants for whom we have estimates of self-knowledge, we can match 3,573 respondents to their mothers and 2,964 respondents to their fathers. Figure 6 scrutinizes the relationship between parents' estimated levels of self-knowledge and the respective estimates for their children. For this purpose, each survey respondent is assigned to the respective decile in the distribution of  $\hat{\tau}$  in the full sample. The figure depicts the average deciles for children conditional on the respective parental deciles, separately for mothers (left panel) and fathers (right panel). As the corresponding regression lines indicate, parents' and children's estimated levels of self-knowledge are positively related ( $p < 0.001$ , two-sided, separately for both cases). In terms of the precise underlying values of  $\hat{\tau}$  (i.e., not in terms of deciles), the rank correlation between children and their parents is 0.17 in the case of mothers and 0.16 for fathers.



*Note:* Graphs plot average deciles to which children's  $\hat{\tau}$  belong for the deciles of  $\hat{\tau}$  for mothers (left) and fathers (right). Deciles refer to the full sample, i.e., thresholds are the same for parents and children. The shaded areas around the regression lines represent 95% confidence intervals. Standard errors are clustered at the level of the respective parents.

Figure 6: Intergenerational transmission of self-knowledge

## 5.2 Predicting Outcomes

To illustrate how accounting for individual estimates of  $\hat{\tau}$  can increase explanatory power in the context of non-cognitive skills, we study the effect of risk attitudes on various economic outcomes. Similar to the analysis of the experiment in Table 3, we split the respective samples of SOEP respondents into two groups: individuals with either low self-knowledge (below the median value of  $\hat{\tau}$ ) or high self-knowledge (above the median level of  $\hat{\tau}$ ). This way, we refrain from imposing any functional form assumptions about how self-knowledge affects the estimates. In light of the model and the experimental results, we would expect to see larger explanatory power for the above-median sample than for the below-median sample, reflected in larger values of  $R^2$ .

Table 5 presents empirical results for three different economic outcomes related to risk attitudes: holding risky financial securities, receiving performance-related pay, and smoking. These outcomes were selected based on prior research, arguing that they should—and actually are—related to risk attitudes (Dohmen and Falk, 2011; Dohmen et al., 2011). The measures that we use to elicit risk attitudes are survey items that ask about willingness to take risks in specific domains on eleven-point Likert scales. In particular, the items refer to the willingness to take risks concerning one's *financial matters*, *career*, and *health*, respectively. Columns 1–3 show results from OLS regressions without further controls, and Columns 4–6 replicate the analyses controlling for a set of socio-demographic characteristics, namely (squared) age, gender, body height, years of education, parental education, log net household income, log wealth, and log debts. Columns 1 and 4 consider the full sample and confirm a positive and significant relationship between risk attitudes and the respective outcomes.

Table 5: Predictive power of domain-specific attitudes towards risk

	Without controls			Including controls		
Sample:	<i>pooled</i> (1)	<i>below</i> (2)	<i>above</i> (3)	<i>pooled</i> (4)	<i>below</i> (5)	<i>above</i> (6)
Dependent variable: <i>Risky financial securities</i>						
Risk attitude	0.0698*** (0.00264)	0.0519*** (0.00359)	0.0863*** (0.00369)	0.0523*** (0.00280)	0.0383*** (0.00383)	0.0665*** (0.00398)
(Partial) $R^2$	0.0827	0.0520	0.114	0.0498	0.0295	0.0737
Observations	9095	4548	4547	7472	3736	3736
$\Delta R^2$	119%, $p < 0.001$			150%, $p < 0.001$		
Dependent variable: <i>Performance pay</i>						
Risk attitude	0.0128*** (0.00174)	0.00808*** (0.00221)	0.0174*** (0.00268)	0.00977*** (0.00199)	0.00491 (0.00252)	0.0150*** (0.00310)
(Partial) $R^2$	0.00870	0.00412	0.0139	0.00487	0.00142	0.0101
Observations	5758	2879	2879	4464	2232	2232
$\Delta R^2$	238%, $p = 0.03$			610%, $p = 0.02$		
Dependent variable: <i>Smoking</i>						
Risk attitude	0.0199*** (0.00154)	0.0175*** (0.00220)	0.0229*** (0.00216)	0.0125*** (0.00175)	0.00923*** (0.00248)	0.0158*** (0.00246)
(Partial) $R^2$	0.0119	0.00888	0.0166	0.00481	0.00258	0.00782
Observations	15162	7581	7581	11652	5826	5826
$\Delta R^2$	87%, $p = 0.04$			203%, $p = 0.06$		

*Note:* The table reports OLS estimates, with binary dependent variables taking the values zero and one. If not stated otherwise, all the data refer to the year 2009. Regressions are based only on respondents who are 18 years or older, and those for performance pay include only respondents up to the age of 66 who work full-time and receive wages or salaries. *Risky financial securities* are, in the SOEP, a residual category of securities without a fixed interest rate, like stocks or options (“other securities”). Since the relevant question was asked on the household level in 2010, the units of observation in the respective regressions are households in that year. *Performance pay* indicates that an employee receives payments from profit-sharing, premiums, or bonuses. Smoking refers to 2010. The variable *risk attitude* in each of the panels refers to the respective domain-specific question asked in the SOEP. The contexts are *financial matters* for holding risky financial securities, *career* for performance pay, and *health* for smoking. The controls used in Columns 4–6 are gender, age, squared age, body height in 2010, years of education, parental education (whether mother and father each have either *Abitur* or *Fachabitur*), log net household income, and log wealth and log debts of the current household in 2007. The last three variables are calculated as  $\ln(\text{euro amount} + 1)$ . For the regressions involving *risky financial securities*, all variables are averaged on the household level, and we base our data only on respondents for whom all information is available individually. The  $p$ -values for the sizes of  $\Delta R^2$  are each based on 10,000 permutations. Heteroskedasticity-robust standard errors in parentheses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Our main interest concerns the pairwise comparisons between Columns 2 and 3 and those between Columns 5 and 6, where we show results for individuals with estimated levels of self-knowledge below and above the median. In all instances, the values of  $R^2$  are higher for individuals with high self-knowledge than for individuals with low self-knowledge, and the values for the full sample are between those for the two sub-samples.<sup>11</sup> This holds both with and without controls included in the regressions (in the regressions with controls, we refer to the partial  $R^2$ ). In all cases, the explanatory power among high self-knowledge respondents is much larger than among the ones with low self-knowledge, ranging from an 87% increase (smoking, without controls) up to an increase of 610% (performance pay, with controls). As the respective  $p$ -values show, the differences in explanatory power are statistically significant. We note that these results hold for a non-cognitive skill—risk attitude—that is different and mostly unrelated to the set of traits that we used to estimate  $\hat{\tau}$  (the Big Five). This suggests that self-knowledge does, in fact, generalize to different aspects of people’s personalities.

### 5.3 Determinants of Preferences

An active literature seeks to uncover the individual determinants of preferences and personality (e.g., Sutter and Kocher, 2007; Croson and Gneezy, 2009; Falk et al., 2018). Understanding how, e.g., age and gender affect preferences is not only interesting in itself. It is also relevant for gaining a better understanding of group-specific outcomes, such as gender differences with respect to sorting into competitive environments, wage gaps, and occupational choice, to give just one example (see, e.g., Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Dohmen and Falk, 2011; Buser, Niederle, and Oosterbeek, 2014). Here, we use differences in domain-specific risk attitudes associated with gender and height to illustrate that when accounting for differences in self-knowledge, exogenous determinants of preferences may actually have higher explanatory power and yield larger effect sizes than typically inferred.

Table 6 reports the differences associated with gender and body height for two different measures of risk attitudes, both based on the 2009 wave of the SOEP and standardized according to the pooled samples used in Column 1.<sup>12</sup> One measure is the so-called general risk question, asking about the willingness to take risks “in general” and measured on an eleven-point Likert scale, while the other is the first principal component of five domain-specific risk questions, referring to *car driving*, *financial matters*, *sports/leisure*, *career*, as well as *health*. Replicating previous findings,<sup>13</sup> women tend to be less willing to take risks than men, and taller people tend to be more willing to take risks than smaller individuals. Our interest here is to compare samples with high vs. low levels of self-knowledge, as

<sup>11</sup>In Appendix C, we discuss how to interpret differences in the estimated coefficients.

<sup>12</sup>Individuals’ height again refers to 2010, due to availability of data.

<sup>13</sup>See in particular Dohmen et al. (2011) but also Croson and Gneezy (2009) and Falk et al. (2018).

Table 6: Differences in risk attitudes

	Unweighted			Weighted	
Sample:	<i>pooled</i> (1)	<i>below</i> (2)	<i>above</i> (3)	<i>below</i> (4)	<i>above</i> (5)
General risk question					
Female	-0.374*** (0.0153)	-0.348*** (0.0218)	-0.400*** (0.0214)	-0.350*** (0.0219)	-0.410*** (0.0221)
$R^2$	0.0348	0.0297	0.0404	0.0305	0.0416
Observations	16654	8327	8327	8327	8327
$\Delta R^2$	36%, $p = 0.05$			36%, $p = 0.05$	
Height (in '10)	0.0219*** (0.000864)	0.0202*** (0.00121)	0.0236*** (0.00124)	0.0196*** (0.00122)	0.0249*** (0.00127)
$R^2$	0.0425	0.0349	0.0502	0.0333	0.0556
Observations	15134	7567	7567	7567	7567
$\Delta R^2$	44%, $p = 0.02$			67%, $p < 0.01$	
<i>Domain-specific risk questions: first principal component</i>					
Female	-0.433*** (0.0164)	-0.401*** (0.0238)	-0.463*** (0.0226)	-0.405*** (0.0238)	-0.477*** (0.0236)
$R^2$	0.0469	0.0390	0.0558	0.0402	0.0578
Observations	14160	7080	7080	7080	7080
$\Delta R^2$	43%, $p = 0.02$			44%, $p = 0.02$	
Height (in '10)	0.0271*** (0.000947)	0.0252*** (0.00133)	0.0284*** (0.00135)	0.0243*** (0.00135)	0.0305*** (0.00138)
$R^2$	0.0648	0.0532	0.0750	0.0498	0.0855
Observations	12858	6429	6429	6429	6429
$\Delta R^2$	41%, $p = 0.01$			72%, $p < 0.001$	

*Note:* The table reports OLS estimates. All regressions only use respondents who are 18 years or older. The dependent variables are standardized among the respondents who enter the corresponding regression in Column 1. Columns 4 and 5 use inverse probability weights that come from probit regressions of group assignment on gender, a second-order age polynomial, and years of education. Except for height, all data refer to the year 2009. The  $p$ -values for the respective sizes of  $\Delta R^2$  are each based on 10,000 permutations. Heteroskedasticity-robust standard errors in parentheses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

shown in Columns 2 and 3. In all four instances, we consistently find that explanatory power, measured in terms of  $R^2$ , is larger among high- $\tau$  individuals than among low- $\tau$  individuals. These differences are substantial, ranging from 36% to 44%. The  $p$ -values for differences in explanatory power imply statistical significance. An inspection of estimated coefficients further shows that the size of coefficients is always (absolutely) larger for the above-median sample than for the below-median sample. Increased effect sizes are at odds with classical measurement error but in line with the predictions of our model. They also mimic the results from our stylized experiment in Section 4, where we saw a steeper slope between reports and true states for high- $\tau$  relative to low- $\tau$  subjects (see Table 3, Columns 4 and 5).

A potential concern regarding the interpretation of the above results is selection. The latter would imply that the observed patterns reflect that the true explanatory power, as



well as true coefficients, are actually larger among respondents with high self-knowledge. In principle, we cannot rule out such an interpretation with non-experimental data. However, recall from Table 4 that the effects of socio-demographic characteristics on  $\hat{\tau}$  were rather small. It is therefore unlikely that selection plays a major role in our findings. Still, we address this issue explicitly in Columns 4 and 5 by restoring representativeness with respect to observable characteristics using inverse probability weighting. We estimate probit models in which we regress group assignment (below or above median) on gender, a second-order age polynomial, and years of education (all as of 2009). We then invert the predicted probabilities and use them as weights, otherwise replicating the regression from Columns 2 and 3. The results change very little and even tend to become slightly stronger.<sup>14</sup> Thus, the findings suggest that individuals with relatively high levels of self-knowledge do, in fact, contribute more to our understanding of the determinants of non-cognitive skills than their low- $\tau$  counterparts.

We conclude this section with a brief discussion of the differences associated with gender and height in the Big Five personality traits. Table 11 in Appendix D.2 is constructed analogously to Table 6 but analyzes the Big Five rather than risk attitudes. It shows that both effects—higher explanatory power and larger effect sizes for high- $\tau$  relative to low- $\tau$  individuals—are also observed for the Big Five Inventory. As an example, take conscientiousness, which is considered one of the most important personality traits for explaining educational and labor market outcomes (Judge et al., 1999; Hogan and Holland, 2003; Almlund et al., 2011) as well as health and mortality (see, e.g., Bogg and Roberts, 2004; Hill et al., 2011). The comparison of Columns 2 and 3 shows that the gender difference is almost three times as large for the high- $\tau$  compared with the low- $\tau$  individuals and that the difference in  $R^2$  amounts to more than 500%.<sup>15</sup> This also suggests the possibility that treatment effects of childhood interventions on personality traits (see Heckman and Kautz, 2012) could be even larger than previously assumed.

## 6 Conclusion

In this paper, we have suggested a theoretical framework of survey response behavior. We assume that respondents try to provide accurate answers but lack perfect self-knowledge. In addition, survey responses may be affected in terms of subjective scale use, inaccurate beliefs about one’s self-knowledge, differences in the endogenous precision of reports, as well as image or social desirability effects. The framework is kept deliberately simple but

---

<sup>14</sup>A corresponding procedure can, of course, also be applied in the context of the differences in predictive power that were analyzed in Table 5. Although given the controls that are used, it seems less needed at that point, we still report the corresponding results in Table 10 in Appendix D.1. The effect sizes decrease a bit, but the results still support the earlier conclusions.

<sup>15</sup>For the Big Five, we can also show that Cronbach’s alpha, a common psychometric measure for scale consistency, is higher among respondents with high (above-median) levels of  $\hat{\tau}$ . Among low- $\hat{\tau}$  respondents in the SOEP, the average across the five facets is 0.50, while it is 0.67 among high- $\hat{\tau}$  respondents.

could be extended to allow for a richer and more realistic analysis of survey response behavior. For example, we assume that the outcome of inspecting one’s individual characteristics is simply an (exogenous) signal about one’s type. It would be interesting to explore cognitive (and emotional) processes involved in this introspection process in more detail, e.g., the role of limited memory and retrieval, how individuals select representative choice contexts to evaluate their characteristics, or how social comparison or life experience affect introspection. The framework also allows for integrating the role and meaning of response times, which could hold strong practical importance. For example, many binary choice experiments in neuroscience and psychology find that accuracy decreases with response time, in the sense that slower decisions are less likely to be correct (Swensson, 1972; Luce, 1986; Ratcliff and McKoon, 2008).<sup>16</sup> An interesting question is how one can integrate response times into our approach to facilitate the identification of precise responses.

We note that while we have interpreted the model in terms of survey response behavior, it can be applied to any elicitation method where subjects make a decision, i.e., in particular to lab and field experiments. For instance, in typical choice experiments to elicit risk or time preferences, the same issues that we discuss in the context of survey responses also arise. In fact, a main difference in experiments is the provision of incentives, which may increase the accuracy of responses (see Section 2.2.1) but do not solve the issues of limited self-knowledge (in the sense of introspective ability), scale use, or social desirability.

A better understanding of the survey response process may also inform the “optimal” design of research. Conditional on survey respondents’ behavior, we can ask the question of how surveys or other elicitation methods should be designed to extract a maximum amount of information. Such a design perspective would consider research as a principal–agent relationship where agents participate in surveys, experiments, or related research contexts that are designed by researchers who optimize research paradigms conditional on agents’ behaviors. Such an approach could be used to investigate how to design survey items and response scales, when and how incentives should be given, or how to design specific modules meant to correct for expected biases.

A key insight of the model is that we can extract individual differences in self-knowledge based on response patterns, in particular by using the ratio of the variance between characteristics and the variance for a given characteristic over time. Building on this finding, we suggest a consistent and unbiased estimator of self-knowledge, discuss its properties, and apply it to experimental data as well as a large panel data set. We show that the estimator reliably identifies individual differences in the informativeness of answers in the laboratory context where we know true states. Splitting the lab sample

---

<sup>16</sup>Fudenberg, Strack, and Strzalecki (2018) and Alós-Ferrer, Fehr, and Netzer (2021) provide theoretical analyses of the relationship between response times and the accuracy of binary decisions.

into individuals giving answers with high vs. low quality, we further show that reports are much closer to true states for the former than for the latter part of the sample. Repeating the same exercise using a representative panel data set and risk attitudes as an example for non-cognitive skills, we show that for subjects with a high level of self-knowledge, the explained variance is significantly higher than for individuals with low levels of self-knowledge. This holds for regressions where risk attitudes are on either the left- or the right-hand side of the regression equation. These applications illustrate the potential of distinguishing between respondents with high vs. low self-knowledge for improving survey evidence. They suggest further econometric implications for the study of measurement error and highlight the potential of integrating self-knowledge into regression frameworks.

## References

- Alan, Sule, Nazli Baydar, Teodora Boneva, Thomas F. Crossley, and Seda Ertac. 2017. “Transmission of risk preferences from mothers to daughters”. *Journal of Economic Behavior and Organization* 134:60–77.
- Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. “Personality Psychology and Economics”. In *Handbook of the Economics of Education*, ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 4:1–181. 2008. Amsterdam, Netherlands: Elsevier B.V.
- Alós-Ferrer, Carlos, Ernst Fehr, and Nick Netzer. 2021. “Time Will Tell: Recovering Preferences When Choices Are Noisy”. *Journal of Political Economy* 129 (6): 1828–1877.
- Beauchamp, Jonathan P., David Cesarini, and Magnus Johannesson. 2017. “The psychometric and empirical properties of measures of risk preferences”. *Journal of Risk and Uncertainty* 54 (3): 203–237.
- Beauchamp, Jonathan, Daniel J. Benjamin, David I. Laibson, and Christopher F. Chabris. 2020. “Measuring and controlling for the compromise effect when estimating risk preference parameters”. *Experimental Economics* 23:1069–1099.
- Bénabou, Roland, Armin Falk, Luca Henkel, and Jean Tirole. 2020. *Eliciting Moral Preferences*. Working paper. Bonn, Germany: briq – Institute on Behavior & Inequality.
- Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. “hroot: Hamburg Registration and Organization Online Tool”. *European Economic Review* 71:117–120.
- Bogg, Tim, and Brent W. Roberts. 2004. “Conscientiousness and Health-Related Behaviors: A Meta-Analysis of the Leading Behavioral Contributors to Mortality”. *Psychological Bulletin* 130 (6): 887–919.
- Bolsinova, Maria, Paul de Boeck, and Jesper Tijmstra. 2017. “Modelling Conditional Dependence Between Response Time and Accuracy”. *Psychometrika* 82 (4): 1126–1148.
- Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. “Measurement Error in Survey Data”. Chap. 59 in *Handbook of Econometrics*, ed. by James J. Heckman and Edward Leamer, 5:3705–3843. Amsterdam, Netherlands: Elsevier Science B.V.
- Bradburn, Norman M., Seymour Sudman, and Brian Wansink. 2004. *Asking Questions: The Definitive Guide to Questionnaire Design*. 448 pp. San Francisco: Jossey-Bass.
- Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. “Gender, Competitiveness, and Career Choices”. *The Quarterly Journal of Economics* 129 (3): 1409–1447.
- Camerer, Colin F., and Robin M. Hogarth. 1999. “The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework”. *Journal of Risk and Uncertainty* 19 (1-3): 7–42.
- Camerer, Colin, and Dan Lovallo. 1999. “Overconfidence and Excess Entry: An Experimental Approach”. *American Economic Review* 89 (1): 306–318.
- Caplin, Andrew, Dániel Csaba, John Leahy, and Oded Nov. 2020. “Rational Inattention, Competitive Supply, and Psychometrics”. *Quarterly Journal of Economics* 135 (3): 1681–1724.
- Caplin, Andrew, and Mark Dean. 2015. “Revealed Preference, Rational Inattention, and Costly Information Acquisition”. *American Economic Review* 105 (7): 2183–2203.

- Chen, Yuanyuan, Shuaizhang Feng, James J. Heckman, and Tim Kautz. 2020. “Sensitivity of self-reported noncognitive skills to survey administration conditions”. *Proceedings of the National Academy of Sciences* 117 (2): 931–935.
- Cobb-Clark, Deborah A., and Stefanie Schurer. 2012. “The stability of big-five personality traits”. *Economics Letters* 115 (1): 11–15.
- Croson, Rachel, and Uri Gneezy. 2009. “Gender Differences in Preferences”. *Journal of Economic Literature* 47 (2): 448–474.
- Cvitanić, Jakša, Dražen Prelec, Blake Riley, and Benjamin Tereick. 2017. *Honesty via Choice-Matching*. Mimeo.
- Dohmen, Thomas, and Armin Falk. 2011. “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender”. *American Economic Review* 101 (2): 556–590.
- Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2012. “The Intergenerational Transmission of Risk and Trust Attitudes”. *Review of Economic Studies* 79 (2): 645–677.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. “Individual risk attitudes: Measurement, determinants, and behavioral consequences”. *Journal of the European Economic Association* 9 (3): 522–550.
- Drerup, Tilman, Benjamin Enke, and Hans Martin von Gaudecker. 2017. “The precision of subjective data and the explanatory power of economic models”. *Journal of Econometrics* 200 (2): 378–389.
- Edwards, Michael C. 2009. “An Introduction to Item Response Theory Using the Need for Cognition Scale”. *Social and Personality Psychology Compass* 3 (4): 507–529.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. “Global Evidence on Economic Preferences”. *Quarterly Journal of Economics* 133 (4): 1645–1692.
- Falk, Armin, David Huffman, and Uwe Sunde. 2006a. *Do I Have What It Takes? Equilibrium Search with Type Uncertainty and Non-Participation*. IZA Discussion Paper 2531. Bonn, Germany: IZA Institute for the Study of Labor.
- . 2006b. *Self-Confidence and Search*. IZA Discussion Paper 2525. Bonn, Germany: IZA Institute for the Study of Labor.
- Falk, Armin, and Florian Zimmermann. 2013. “A Taste for Consistency and Survey Response Behavior”. *CESifo Economic Studies* 59 (1): 181–193.
- . 2017. “Consistency as a Signal of Skills”. *Management Science* 63 (7): 2049–2395.
- Fischbacher, Urs. 2007. “z-Tree: Zurich toolbox for ready-made economic experiments”. *Experimental Economics* 10 (2): 171–178.
- Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki. 2018. “Speed, Accuracy, and the Optimal Timing of Choices”. *American Economic Review* 108 (12): 3651–3684.
- Gerlitz, Jean-Yves, and Jürgen Schupp. 2005. *Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Dokumentation der Instrumententwicklung BFI-S auf Basis des SOEP-Pretests 2005*. Research Notes 4. Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

- Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study”. *Journal of Political Economy* 127 (4): 1826–1863.
- Heckman, James J., and Tim D. Kautz. 2012. *Hard Evidence on Soft Skills*. NBER Working Paper 18121. Cambridge, MA: National Bureau of Economic Research.
- Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior”. *Journal of Labor Economics* 24 (3): 411–482.
- Hefß, Simon. 2017. “Randomization inference with Stata: A guide and software”. *Stata Journal* 17 (3).
- Hill, Patrick L., Nicholas A. Turiano, Michael D. Hurd, Daniel K. Mroczek, and Brent W. Roberts. 2011. “Conscientiousness and Longevity: An Examination of Possible Mediators”. *Health Psychology* 30 (5): 536–541.
- Hogan, Joyce, and Brent Holland. 2003. “Using Theory to Evaluate Personality and Job-Performance Relations: A Socioanalytic Perspective”. *Journal of Applied Psychology* 88 (1): 100–112.
- Hyslop, Dean R., and Guido W. Imbens. 2001. “Bias from classical and other forms of measurement error”. *Journal of Business and Economic Statistics* 19 (4): 475–481.
- Judge, Timothy A., Chad A. Higgins, Carl J. Thoresen, and Murray R. Barrick. 1999. “The Big Five Personality Traits, General Mental Ability, and Career Success Across the Life Span”. *Personnel Psychology* 52 (3): 621–652.
- Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. 2008. “Imputing risk tolerance from survey responses”. *Journal of the American Statistical Association* 103 (483): 1028–1038.
- Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2020. “The Formation of Prosociality: Causal Evidence on the Role of Social Environment”. *Journal of Political Economy* 128 (2): 434–467.
- Kosse, Fabian, and Friedhelm Pfeiffer. 2012. “Impatience among preschool children and their mothers”. *Economics Letters* 115 (3): 493–495.
- Kyllonen, Patrick, and Jiyun Zu. 2016. “Use of Response Time for Measuring Cognitive Ability”. *Journal of Intelligence* 4 (4): 14.
- Luce, R. Duncan. 1986. *Response Times: Their Role in Inferring Elementary Mental Organization*. New York, NY: Oxford University Press.
- Malmendier, Ulrike, and Geoffrey Tate. 2005. “CEO Overconfidence and Corporate Investment”. *Journal of Finance* 60 (6): 2661–2700.
- Matějka, Filip, and Alisdair McKay. 2015. “Rational inattention to discrete choices: A new foundation for the multinomial logit model”. *American Economic Review* 105 (1): 272–98.
- Niederle, Muriel, and Lise Vesterlund. 2007. “Do Women Shy Away From Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics* 122 (3): 1067–1101.
- Prelec, Dražen. 2004. “A Bayesian Truth Serum for Subjective Data”. *Science* 306 (5695): 462–466.

- Ratcliff, Roger, and Gail McKoon. 2008. “The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks”. *Neural Computation* 20 (4): 873–922.
- Schwarz, Norbert. 2007. “Cognitive Aspects of Survey Methodology”. *Applied Cognitive Psychology* 21:277–287.
- Sims, Christopher A. 2003. “Implications of rational inattention”. *Journal of monetary Economics* 50 (3): 665–690.
- . 1998. “Stickiness”. *Carnegie-Rochester Conference Series on Public Policy* 49:317–356.
- Smith, Vernon L. 1976. “Experimental Economics: Induced Value Theory”. *American Economic Review* 66 (2): 274–279.
- Storey, Hannah. 2016. *Animal Pictograms and Sleep Infographic*. <https://hannah-storey.co.uk/post/160132182299/sleep-infographic>.
- Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. 304 pp. San Francisco: Jossey-Bass.
- Sutter, Matthias, and Martin G. Kocher. 2007. “Trust and trustworthiness across different age groups”. *Games and Economic Behavior* 59 (2): 364–382.
- Swensson, Richard G. 1972. “The elusive tradeoff: Speed vs accuracy in visual discrimination tasks”. *Perception & Psychophysics* 12 (1A): 16–32.

## Appendix A Proofs

**Proof of Lemma 1.** Following the result from Equation (2), the optimal report for any given level of precision and signal is given by

$$r = \frac{\bar{\theta} + \tau x}{1 + \tau}.$$

Plugging into Equation (2.2.1) yields that the utility of the agent given the optimal response above equals

$$u_{\theta}(r, \tau) = -\frac{m}{(1 + \tau)^2} [\bar{\theta} - \theta + \tau(x - \theta)]^2 - \frac{\tau}{a}$$

and, consequently, the agent chooses her precision  $\tau$  to maximize

$$\begin{aligned} \mathbb{E}[u_{\theta}(r, \tau)] &= -\frac{m}{(1 + \tau)^2} \left( \mathbb{E}[(\bar{\theta} - \theta)^2] + \tau^2 \mathbb{E}[(x - \theta)^2] \right) - \frac{\tau}{a} \\ &= -\frac{m}{(1 + \tau)^2} \left( \sigma^2 + \tau^2 \frac{\sigma^2}{\tau} \right) - \frac{\tau}{a} = -\frac{m \sigma^2}{1 + \tau} - \frac{\tau}{a}. \end{aligned}$$

Since  $\mathbb{E}[u_\theta(r, \tau)]$  is strictly concave in  $r$ , the first-order condition yields the optimal level of effort for an interior solution.

$$0 \stackrel{!}{=} \frac{\partial \mathbb{E}[u_\theta(r, \tau)]}{\partial \tau} = \frac{m \sigma^2}{(1 + \tau)^2} - \frac{1}{a} \quad \Rightarrow \quad \tau^* = \sqrt{m a} \sigma - 1$$

□

**Proof of the theorem.** We will prove the result in the more general setting with subjective self-knowledge and scale use as introduced in Sections 2.2.2 and 2.2.3, respectively. The case without subjective self-knowledge and scale use stated in the basic version of the model corresponds to the special case where  $\tilde{\tau}_i = \tau_i$  and  $\phi_i = 1$ .

Throughout the proof, we fix  $\tau_i, \tilde{\tau}_i > 0$  and  $\phi_i \in (0, 1]$ . The answer of agent  $i$  when asked for the  $t^{\text{th}}$  time about the  $k^{\text{th}}$  characteristic is given by

$$r_{ikt} = (1 - \phi_i) c + \phi_i \frac{\bar{\theta} + \tilde{\tau}_i x_{ikt}}{1 + \tilde{\tau}_i}.$$

By assumption, there exist independent, standard normally distributed random variables  $\epsilon_{ikt}, \eta_{ik}$  such that

$$\begin{aligned} x_{ikt} &= \theta_{ik} + \frac{\sigma}{\sqrt{\tau_i}} \epsilon_{ikt}, \\ \theta_{ik} &= \bar{\theta} + \sigma \eta_{ik}. \end{aligned}$$

Plugging into the equation for the agent's responses yields that

$$r_{ikt} = (1 - \phi_i) c + \phi_i \left( \bar{\theta} + \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \sigma \left[ \eta_{ik} + \frac{\epsilon_{ikt}}{\sqrt{\tau_i}} \right] \right). \quad (11)$$

Denote agent  $i$ 's average answer for question  $k$  by  $\bar{r}_{ik} = \frac{1}{T} \sum_{t=1}^T r_{ikt}$ , her average answer over all questions by  $\bar{r}_i = \frac{1}{K} \sum_{k=1}^K \bar{r}_{ik}$ , and similarly  $\bar{x}_{ik} = \frac{1}{T} \sum_{t=1}^T x_{ikt}$ ,  $\bar{\epsilon}_{ik} = \frac{1}{T} \sum_{t=1}^T \epsilon_{ikt}$ ,  $\bar{x}_i = \frac{1}{K} \sum_{k=1}^K \bar{x}_{ik}$ ,  $\bar{\epsilon}_i = \frac{1}{K} \sum_{k=1}^K \bar{\epsilon}_{ik}$ , and  $\bar{\eta}_i = \frac{1}{K} \sum_{k=1}^K \bar{\eta}_{ik}$ . We have that

$$\frac{r_{ikt} - \bar{r}_{ik}}{\phi_i} = \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} (x_{ikt} - \bar{x}_{ik}) = \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \frac{\sigma}{\sqrt{\tau_i}} (\epsilon_{ikt} - \bar{\epsilon}_{ik}). \quad (12)$$

Similarly, we get that

$$\begin{aligned} \frac{\bar{r}_{ik} - \bar{r}_i}{\phi_i} &= \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} (\bar{x}_{ik} - \bar{x}_i) = \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \left( \left( \theta_{ik} + \frac{\sigma}{\sqrt{\tau_i}} \bar{\epsilon}_{ik} \right) - \left( \bar{\theta}_i + \frac{\sigma}{\sqrt{\tau_i}} \bar{\epsilon}_i \right) \right) \\ &= \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \left( (\theta_{ik} - \bar{\theta}_i) + \frac{\sigma}{\sqrt{\tau_i}} (\bar{\epsilon}_{ik} - \bar{\epsilon}_i) \right) \\ &= \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \left( \sigma (\eta_{ik} - \bar{\eta}_i) + \frac{\sigma}{\sqrt{\tau_i}} (\bar{\epsilon}_{ik} - \bar{\epsilon}_i) \right). \end{aligned} \quad (13)$$



We first show that

$$A := \frac{(1 + \tilde{\tau}_i)^2}{\tilde{\tau}_i^2 \sigma^2} \tau_i \sum_{k=1}^K \sum_{t=1}^T \left( \frac{r_{ikt} - \bar{r}_{ik}}{\phi_i} \right)^2$$

is  $\chi^2$  distributed with  $K(T - 1)$  degrees of freedom. It follows from Equation (12) that

$$A = \sum_{k=1}^K \sum_{t=1}^T (\epsilon_{ikt} - \bar{\epsilon}_{ik})^2.$$

We have that  $A_k := \sum_{t=1}^T (\epsilon_{ikt} - \bar{\epsilon}_{ik})^2$  is  $\chi^2$  distributed with  $T - 1$  degrees of freedom as it equals the sum of the squared distance of i.i.d. normals from the mean. As  $A_k, A_{k'}$  are independent for  $k' \neq k$  and  $A = \sum_{k=1}^K A_k$ , it follows that  $A$  is  $\chi^2$  distributed with  $\sum_{k=1}^K (T - 1) = K(T - 1)$  degrees of freedom.

We next argue that

$$B := \frac{(1 + \tilde{\tau}_i)^2}{\tilde{\tau}_i^2 \sigma^2} \frac{1}{1 + \frac{1}{T\tau_i}} \sum_{k=1}^K \left( \frac{\bar{r}_{ik} - \bar{r}_i}{\phi_i} \right)^2$$

is  $\chi^2$  distributed with  $K - 1$  degrees of freedom. It follows from Equation (13) that

$$B = \sum_{k=1}^K (\lambda_{ik} - \bar{\lambda}_i)^2$$

where  $\lambda_{ik} = \frac{1}{\sqrt{1 + \frac{1}{T\tau_i}}} (\eta_{ik} + \frac{1}{\sqrt{\tau_i}} \bar{\epsilon}_{ik})$ . As

$$\text{var}(\lambda_{ik}) = \frac{\text{var}(\eta_{ik}) + \frac{1}{\tau_i} \text{var}(\bar{\epsilon}_{ik})}{1 + \frac{1}{T\tau_i}} = \frac{1 + \frac{1}{\tau_i} \text{var}\left(\frac{1}{T} \sum_{t=1}^T \epsilon_{ikt}\right)}{1 + \frac{1}{T\tau_i}} = 1,$$

the random variables  $(\lambda_{ik})_{k \in \{1, \dots, K\}}$  are i.i.d. standard normal random variables. Again, as  $\lambda_{ik}, \lambda_{ik'}$  are independent for  $k \neq k'$ , it follows that  $B$  is  $\chi^2$  distributed with  $K - 1$  degrees of freedom.

Next, recall that for the Normal distribution, the sample variance  $\frac{1}{T-1} \sum_{t=1}^T (\epsilon_{ikt} - \bar{\epsilon}_{ik})^2$  is independent of the sample mean  $\bar{\epsilon}_{ik}$ . As  $\eta$  is independent of  $\epsilon$  it follows that  $\sum_{t=1}^T (\epsilon_{ikt} - \bar{\epsilon}_{ik})^2$  and  $\lambda_{ik} = \frac{1}{\sqrt{1 + \frac{1}{T\tau_i}}} (\eta_{ik} + \frac{1}{\sqrt{\tau_i}} \bar{\epsilon}_{ik})$  are independent. This implies that  $A$  and  $B$  are independent. As  $A$  and  $B$  are independently  $\chi^2$  distributed it follows that

$$F_i := \frac{\frac{1}{K-1} B}{\frac{1}{K(T-1)} A}$$

follows an  $F$ -distribution with parameters  $K - 1$  and  $K(T - 1)$ .<sup>17</sup> Recall that in Equa-

---

<sup>17</sup>See <https://en.wikipedia.org/wiki/F-distribution#Characterization> (accessed on June 17, 2021).

tion (7), we defined  $\hat{\tau}_i$ .

$$\hat{\tau}_i = \frac{\frac{1}{K-1} \sum_{k=1}^K (\bar{r}_{ik} - \bar{r}_i)^2}{\frac{1}{K(T-1)-2} \sum_{k=1}^K \sum_{t=1}^T (r_{ikt} - \bar{r}_{ik})^2} - \frac{1}{T}$$

Plugging in the definition of  $A$  and  $B$  yields that

$$\begin{aligned} \hat{\tau}_i + \frac{1}{T} &= \frac{K(T-1)-2}{K(T-1)} \frac{\frac{1}{K-1} \sum_{k=1}^K \left( \frac{\bar{r}_{ik} - \bar{r}_i}{\phi_i} \right)^2}{\frac{1}{K(T-1)} \sum_{k=1}^K \sum_{t=1}^T \left( \frac{r_{ikt} - \bar{r}_{ik}}{\phi_i} \right)^2} \\ &= \frac{K(T-1)-2}{K(T-1)} \frac{\frac{1}{K-1} B \frac{\bar{\tau}_i^2 \sigma^2}{(1+\bar{\tau}_i)^2} \left( 1 + \frac{1}{T\tau_i} \right)}{\frac{1}{K(T-1)} A \frac{\bar{\tau}_i^2 \sigma^2}{(1+\bar{\tau}_i)^2} \frac{1}{\tau_i}} \\ &= \frac{K(T-1)-2}{K(T-1)} \times \tau_i \left( 1 + \frac{1}{T\tau_i} \right) \times \frac{\frac{1}{K-1} B}{\frac{1}{K(T-1)} A} \\ &= \frac{K(T-1)-2}{K(T-1)} \times \left( \tau_i + \frac{1}{T} \right) \times F_i. \end{aligned}$$

This establishes the first part of the theorem, i.e., Equation (8). Part 2 of the Theorem follows as  $\mathbb{E}[F_i] = \frac{K(T-1)}{K(T-1)-2}$ .<sup>18</sup> Part 3 follows as

$$\text{var}(F_i) = \mathbb{E}[F_i]^2 \frac{2((K-1) + K(T-1) - 2)}{(K-1)(K(T-1) - 4)}.$$

To prove Part 4, observe that Equation (9) is decreasing in  $T$ , and thus an upper bound is given by setting  $T = 2$ .

$$\begin{aligned} \sqrt{\mathbb{E}[(\hat{\tau}_i - \tau_i)^2 | \tau_i]} &\leq \left( \tau_i + \frac{1}{2} \right) \sqrt{\frac{2((K-1) + K - 2)}{(K-1)(K-4)}} = \left( \tau_i + \frac{1}{2} \right) \sqrt{\frac{4K-6}{(K-1)(K-4)}} \\ &\leq \left( \tau_i + \frac{1}{2} \right) \sqrt{\frac{4}{K-4}} = (2\tau_i + 1) \frac{1}{\sqrt{K-4}}. \end{aligned}$$

This establishes the result. Finally, we note that this result immediately extends to the case of endogenous effort introduced in Section 2.2.1, where for agent-specific ability  $a_i$  and incentives  $m_i$ , the precision is endogenously chosen as  $\tau_i = \sqrt{m_i a_i} \sigma - 1$ .  $\square$

<sup>18</sup>See <https://en.wikipedia.org/wiki/F-distribution> (accessed on June 17, 2021).

## Appendix B Robustness of the Estimator

### B.1 Characteristics with Different Averages and Variances

The estimator introduced in Section 3 assumes that the population means and variances of types are identical for all of the  $K$  characteristics that are being used. Empirically, however, this is usually not the case (at least not exactly). For this reason, we next describe a generalization of the estimator derived in Section 3 to the case where the population mean  $\bar{\theta}_k$  and variance  $\sigma_k^2$  of each characteristic  $k$  is potentially different. We make no assumption about the distribution of these population means and variances, but maintain the assumption that the agent's prior belief equals the distribution of characteristics in the population and that characteristics are independent. Throughout, we maintain the assumption of no scale use, i.e.,  $\phi_i = 1$ .

Fix an infinite sequence of levels of perceived and objective self-knowledge of the respondents,  $\tau_1, \tau_2, \dots$  and  $\tilde{\tau}_1, \tilde{\tau}_2, \dots$ , respectively. We denote by

$$C := \frac{1}{I} \sum_{i=1}^I \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \left( 1 + \frac{1}{T\tau_i} \right)$$

and note that  $C$  is a non-negative constant independent of any specific characteristic. Throughout, we assume that each agent's self-knowledge  $\tau_i$  is bounded from below by  $\underline{\tau}$  which implies that  $C$  is bounded by  $C \leq 1 + \frac{1}{T\underline{\tau}}$ . There exist i.i.d. standard normally distributed random variables  $(\epsilon_{ikt})_{ikt}$  and  $(\eta_{ik})_{ik}$  such that

$$\begin{aligned} x_{ikt} &= \theta_{ik} + \frac{\sigma_k}{\sqrt{\tau_i}} \epsilon_{ikt}, \\ \theta_{ik} &= \bar{\theta}_k + \sigma_k \eta_{ik}. \end{aligned}$$

We get that (without scale use) the agent's response when asked for the  $t^{\text{th}}$  time about characteristic  $k$  is then given by

$$r_{ikt} = \frac{\bar{\theta}_k + \tilde{\tau}_i x_{ikt}}{1 + \tilde{\tau}_i} = \bar{\theta}_k + \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} (x_{ikt} - \bar{\theta}_k) = \bar{\theta}_k + \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \sigma_k \left( \eta_{ik} + \frac{1}{\sqrt{\tau_i}} \epsilon_{ikt} \right).$$

We define the average response by agent  $i$  to question about characteristic  $k$  as  $\bar{r}_{ik} = \frac{1}{T} \sum_{t=1}^T r_{ikt}$  and as  $\bar{r}_k = \frac{1}{I} \sum_{i=1}^I \bar{r}_{ik}$  the average response to question  $k$ .

**Lemma 2.** *The average response to question  $k$  is normally distributed with mean  $\bar{\theta}_k$  and variance*

$$\text{var}(\bar{r}_k) = \frac{\sigma_k^2}{I} C.$$

Furthermore,  $\lim_{I \rightarrow \infty} \bar{r}_k = \bar{\theta}_k$  almost surely.

**Proof.** As  $\eta$  and  $\epsilon$  are normally distributed with mean zero it follows that  $\bar{r}_k$  is normally

distributed and has mean  $\bar{\theta}_k$ . We are thus left to compute the variance of  $\bar{r}_k$ . We define  $\bar{\epsilon}_{ik} = \frac{1}{T} \sum_{t=1}^T \epsilon_{ikt}$  as the average signal shock of agent  $i$  for characteristic  $k$ . As  $\eta_{ik}$  and  $\bar{\epsilon}_{ik}$  are independent across agents, we have that

$$\begin{aligned}
\text{var}(\bar{r}_k) &= \frac{1}{I^2} \sum_{i=1}^I \text{var}(\bar{r}_{ik}) = \frac{1}{I^2} \sum_{i=1}^I \text{var} \left( \bar{\theta}_k + \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \sigma_k \left( \eta_{ik} + \frac{1}{\sqrt{\tau_i}} \bar{\epsilon}_{ik} \right) \right) \\
&= \frac{\sigma_k^2}{I^2} \sum_{i=1}^I \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \text{var} \left( \eta_{ik} + \frac{1}{\sqrt{\tau_i}} \bar{\epsilon}_{ik} \right) \\
&= \frac{\sigma_k^2}{I^2} \sum_{i=1}^I \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \left( 1 + \frac{\text{var}(\bar{\epsilon}_{ik})}{\tau_i} \right) = \frac{\sigma_k^2}{I^2} \sum_{i=1}^I \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \left( 1 + \frac{\frac{1}{T^2} \sum_{t=1}^T \text{var}(\epsilon_{ikt})}{\tau_i} \right) \\
&= \frac{\sigma_k^2}{I^2} \sum_{i=1}^I \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \left( 1 + \frac{1}{T\tau_i} \right).
\end{aligned}$$

The almost sure convergence follows from Kolmogorov's strong law of large numbers for independently but not identically distributed random variables.  $\square$

Similarly, we define the variance in responses to question  $k$  as

$$s_k^2 = \frac{1}{I-1} \sum_{i=1}^I (\bar{r}_{ik} - \bar{r}_k)^2.$$

**Lemma 3.** *We have that the expected sample variance converges almost surely*

$$\lim_{I \rightarrow \infty} s_k^2 = \sigma_k^2 C.$$

**Proof.** As  $\lim_{I \rightarrow \infty} \bar{r}_k = \bar{\theta}_k$  a.s., the sample variance a.s. satisfies

$$\begin{aligned}
\lim_{I \rightarrow \infty} s_k^2 &= \lim_{I \rightarrow \infty} \frac{1}{I-1} \sum_{i=1}^I [(\bar{r}_{ik} - \bar{\theta}_k)^2 + (\bar{\theta}_k - \bar{r}_k)^2 + 2(\bar{r}_{ik} - \bar{\theta}_k)(\bar{\theta}_k - \bar{r}_k)] \\
&= \lim_{I \rightarrow \infty} \frac{1}{I-1} \sum_{i=1}^I [(\bar{r}_{ik} - \bar{\theta}_k)^2 + (\bar{\theta}_k - \bar{r}_k)^2] \\
&= \lim_{I \rightarrow \infty} \frac{I}{I-1} \left[ (\bar{\theta}_k - \bar{r}_k)^2 + \frac{1}{I} \sum_{i=1}^I (\bar{r}_{ik} - \bar{\theta}_k)^2 \right].
\end{aligned}$$

As  $I/(I-1)$  converges to 1 and  $(\bar{\theta}_k - \bar{r}_k)^2$  converges to zero almost surely, we get that almost surely

$$\lim_{I \rightarrow \infty} s_k^2 = \lim_{I \rightarrow \infty} \frac{1}{I} \sum_{i=1}^I (\bar{r}_{ik} - \bar{\theta}_k)^2.$$

Note that  $\bar{r}_{ik} - \bar{\theta}_k$  is independently normally distributed with mean zero and variance

$$\sigma_k^2 \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \left( 1 + \frac{1}{T\tau_i} \right).$$

Thus, we get that

$$\mathbb{E}[(\bar{r}_{ik} - \bar{\theta}_k)^2] = \sigma_k^2 \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \left( 1 + \frac{1}{T\tau_i} \right)$$

and

$$\text{var}((\bar{r}_{ik} - \bar{\theta}_k)^2) = 2\sigma_k^4 \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^4 \left( 1 + \frac{1}{T\tau_i} \right)^2 \leq 2\sigma_k^4 \left( 1 + \frac{1}{T\tau_i} \right)^2.$$

As the variance of  $(\bar{r}_{ik} - \bar{\theta}_k)^2$  is bounded, we can apply Kolmogorov's strong law of large numbers and get that

$$\lim_{I \rightarrow \infty} s_k^2 = \lim_{I \rightarrow \infty} \frac{1}{I} \sum_{i=1}^I (\bar{r}_{ik} - \bar{\theta}_k)^2 = \lim_{I \rightarrow \infty} \frac{1}{I} \sum_{i=1}^I \sigma_k^2 \left( \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \right)^2 \left( 1 + \frac{1}{T\tau_i} \right) = \sigma_k^2 C. \quad \square$$

We define the normalized response  $n_{ikt}$  as the difference between agent  $i$ 's response and the average response, divided by the standard deviation of agents' average responses for the given characteristic  $k$ , i.e.

$$n_{ikt} = \frac{r_{ikt} - \bar{r}_k}{s_k}.$$

Together Lemma 2 and 3 imply the following result.

**Lemma 4.** *The normalized responses times  $\sqrt{C}$  almost surely converge in the number of agents to*

$$\lim_{I \rightarrow \infty} \sqrt{C} n_{ikt} = \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i} \left( \eta_{ik} + \frac{1}{\sqrt{\tau_i}} \epsilon_{ikt} \right) \quad (14)$$

We observe that the above asymptotic distribution for  $I \rightarrow \infty$  of the normalized responses multiplied by  $\sqrt{C}$  does not depend on scale use or the means and variances of characteristics. Moreover, the comparison of Equations (14) and (11) shows that the normalized responses are distributed exactly as if the respondents' scale use parameters  $\phi_i$  equaled one, all means  $\bar{\theta}_k$  were zero, and the variances  $\sigma_k^2$  of characteristics all took the value of  $1/C$ . We define the population-based estimator as

$$\hat{\tau}_i^{POP} = \frac{\frac{1}{K-1} \sum_{k=1}^K (\bar{n}_{ik} - \bar{n}_i)^2}{\frac{1}{K(T-1)-2} \sum_{k=1}^K \sum_{t=1}^T (n_{ikt} - \bar{n}_{ik})^2} - \frac{1}{T}. \quad (15)$$

The proof given for the theorem now yields the following result:

**Proposition.** *For every  $K, T$  that satisfy  $K(T-1) > 4$ .*

Table 7: Accuracy of estimates with different means and variances

	(1)	(2)	(3)	(4)	(5)
$I$ (respondents)	100	10,000	100	100	100
$K$ (characteristics)	15	15	50	15	50
$T$ (waves)	3	3	3	10	10
Correlation	0.68	0.68	0.87	0.76	0.91
Rank correlation	0.76	0.77	0.90	0.82	0.93
Median split	79%	80%	88%	83%	90%

1. The estimator  $\hat{\tau}_i^{POP}$  satisfies almost surely

$$\lim_{I \rightarrow \infty} \hat{\tau}_i^{POP} = \left( \tau_i + \frac{1}{T} \right) \frac{K(T-1) - 2}{K(T-1)} F_i - \frac{1}{T} \quad (16)$$

for some random variable  $F_i$  that is  $F$  distributed with  $K-1, K(T-1)$  degrees of freedom for every fixed vector of parameters  $\tau_i, \sigma, \bar{\theta}$ .

2.  $\hat{\tau}_i^{POP}$  is a consistent estimator for  $\tau_i^{POP}$ , i.e.,  $\lim_{I \rightarrow \infty} \mathbb{E}[\hat{\tau}_i^{POP} | \tau_i] = \tau_i$  almost surely.

3. The standard error of the estimator  $\hat{\tau}_i^{POP}$  in large populations is given by

$$\lim_{I \rightarrow \infty} \sqrt{\mathbb{E}[(\hat{\tau}_i^{POP} - \tau_i)^2 | \tau_i]} = \left( \tau_i + \frac{1}{T} \right) \sqrt{\frac{2((K-1) + K(T-1) - 2)}{(K-1)(K(T-1) - 4)}}. \quad (17)$$

4.  $\hat{\tau}_i^{POP}$  converges to  $\tau_i$  at the rate  $1/\sqrt{K}$  in the number of attributes, and for all  $K > 4$  satisfies the following upper bound independent of the number of repeated observations  $T$

$$\lim_{I \rightarrow \infty} \sqrt{\mathbb{E}[(\hat{\tau}_i^{POP} - \tau_i)^2 | \tau_i]} \leq \frac{2\tau_i + 1}{\sqrt{K-4}}.$$

The properties of the population-based estimator are now asymptotic and do not necessarily hold in small samples. However, the only dimension of the sample size that is relevant for convergence is the number  $I$  of respondents. While, in most applications, the number of characteristics and waves ( $K$  and  $T$ , respectively) will probably be limited, the number of respondents is usually fairly large. The asymptotic properties might, therefore, be a realistic approximation of the actual behavior of the population-based estimator in many relevant contexts, as we illustrate with the simulation results below.

The table replicates Table 1, aside from that the means of the characteristics that are assumed. The means  $\bar{\theta}$  are independently drawn from a Normal distribution with a mean of 5 and a standard deviation of 1. The standard deviations of characteristics,  $\theta$ , are drawn from a log-normal distribution with the parameters  $-1/2$  and 1, such that the expected standard deviation still equals one. A comparison of the result shows that the

performance is almost identical to the case with equal means. This even holds for the cases where the simulated number of respondents is just 100, a sample size that most studies exceed.

## B.2 Correlated Characteristics

We choose the Big Five inventory for estimating self-knowledge because, by design, the five measured traits are close to statistical independence. However, the five traits are each measured with a set of three survey items, which among each other are correlated. This does not impede the logic behind our estimator: subjects with high self-knowledge should give similar answers over time to the same questions, and they should give different answers to questions about different traits. What does not hold here is that estimates are necessarily unbiased. In the stylized experiment presented in Section 4, all assumptions of the estimator were fulfilled, and yet unbiasedness was not the important property that we used for the results in Table 3. Instead, we relied on sample splits, i.e., our aim was to sort subjects according to how much information about the true type was entering their reports. Our interest here is the same, and the estimator remains informative. To gain a better understanding of how correlations in characteristics influence our estimates, we replicate the simulation results from Table 1 with the following modifications: we impose that characteristics are correlated in the same way as answers to the 15 Big Five questions in the 2009 wave of the SOEP, and we replicate all the columns that use 15 characteristics.

Table 8: Accuracy of estimates with correlated characteristics

	(1)	(2)	(3)
$I$ (respondents)	100	10,000	100
$K$ (characteristics)	15	15	15
$T$ (waves)	3	3	10
Correlation	0.65	0.64	0.72
Rank correlation	0.74	0.74	0.80
Median split	78%	78%	81%
Bias	-0.19	-0.19	-0.19

The results are reported in Table 8, whose columns are identically constructed as Columns 1, 2, and 4 in Table 1. The main result is that the fraction of respondents who are correctly classified as having below- or above-median self-knowledge decreases only by about two percentage points, i.e., the informativeness of the median-splits remains.

## Appendix C Implications for OLS Estimates

In the analyses presented in the paper, we concentrate on OLS regressions, where the relevant self-report serves either as the dependent or as an independent variable. To

facilitate understanding of our results, we first summarize the effects that we would expect from  $\tau$  in the light of our model. Table 9 provides a schematic overview of the effects that our model of survey responses predicts for regression coefficients estimated with OLS, formulated in terms of attenuation (bias towards zero;  $-$ ) and amplification (bias away from zero;  $+$ ). The two columns of the table differentiate between the cases of the report being used as the dependent variable (left-hand side of the equation) or as an independent variable (right-hand side of the equation). The respective other variable is assumed to be measured without error. In the upper panel, we distinguish between two channels through which a decrease in  $\tau$  affects estimates: first, increased zero-mean noise around the expected answer, and second, bias in answers towards the population mean due to reduced confidence in one's signals. The lower panel presents the total effects for the three cases of  $\tilde{\tau} < \tau$ ,  $\tilde{\tau} = \tau$ , and  $\tilde{\tau} > \tau$  (see Section 2.2.2).

Table 9: Effect of reduction in self-knowledge  $\tau$  on OLS estimates

Report as:	dependent variable	independent variable
Effect through:		
increased noise	none ( $\circ$ )	attenuation ( $-$ )
decreased $\tilde{\tau}$	attenuation ( $-$ )	amplification ( $+$ )
Overall effect with:		
$\tilde{\tau} < \tau$	$--$	$+$
$\tilde{\tau} = \tau$	$-$	$\circ$
$\tilde{\tau} > \tau$	$-/\circ$	$-$

## C.1 Self-reports as the Dependent Variable

For the report as the dependent variable, it is well known that increased noise per se does not introduce any bias, as stated in the respective table cell. However, in our context, reduced confidence leads to attenuation bias, as we have already seen in the experimental results (see Figure 4a). Formally, assume that we want to estimate the following equation:

$$\theta_i = \beta_0 + \beta_1 y_i + \epsilon_i,$$

where  $y_i$  is the respective realization of the independent variable and  $\epsilon_i$  is an i.i.d. error term with an expected value of zero that is independent of  $y_i$  and the signals that subjects receive. Crucially, the value  $\theta_i$  is not observable and instead replaced with the response  $r_i$ . To gain a deeper insight into the forces behind the composite effect, we use the notation involving subjective self-knowledge (see Section 2.2.2). The asymptotic result of



the standard OLS estimator is derived below.

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\widehat{\text{cov}}(r_i, y_i)}{\widehat{\text{var}}(y_i)} \xrightarrow{p} \frac{\text{cov}(r_i, y_i)}{\text{var}(y_i)} = \frac{\mathbb{E}[(r_i - \bar{r})(y_i - \bar{y})]}{\mathbb{E}[(y_i - \bar{y})^2]} = \frac{\mathbb{E}\left[\frac{\tilde{\tau}(x_i - \bar{\theta})}{1 + \tilde{\tau}}(y_i - \bar{y})\right]}{\mathbb{E}[(y_i - \bar{y})^2]} \\
&= \frac{\tilde{\tau}}{1 + \tilde{\tau}} \frac{\mathbb{E}[(x_i - \theta_i + \theta_i - \bar{\theta})(y_i - \bar{y})]}{\mathbb{E}[(y_i - \bar{y})^2]} = \frac{\tilde{\tau}}{1 + \tilde{\tau}} \frac{\mathbb{E}[(x_i - \theta_i + \beta_1(y_i - \bar{y}) + \epsilon_i)(y_i - \bar{y})]}{\mathbb{E}[(y_i - \bar{y})^2]} \\
&= \frac{\tilde{\tau}}{1 + \tilde{\tau}} \frac{\mathbb{E}[\beta_1(y_i - \bar{y})(y_i - \bar{y})]}{\mathbb{E}[(y_i - \bar{y})^2]} = \frac{\tilde{\tau}}{1 + \tilde{\tau}} \beta_1
\end{aligned}$$

$$\hat{\beta}_0 \xrightarrow{p} \bar{\theta} - \beta_1 \bar{y} = \beta_0 + \beta_1 \bar{y} - \frac{\tilde{\tau}}{1 + \tilde{\tau}} \beta_1 \bar{y} = \beta_0 + \left(1 - \frac{\tilde{\tau}}{1 + \tilde{\tau}}\right) \beta_1 \bar{y}$$

Thus, as long as a decrease in  $\tau$  is accompanied by a decrease in  $\tilde{\tau}$ , the overall effect on the absolute value of the slope parameter  $\beta_1$  is strictly negative.

**An Estimator for  $\tau$  Based on Known True States.** Suppose we know that  $\tau$  is constant in the relevant population, or, alternatively, that all answers were given by the same individual. Suppose also that we know the true states, and we use them as the independent variable, i.e.,  $y_i = \theta_i$  for all  $i$ . It follows that  $\beta_0 = 0$ ,  $\beta_1 = 1$ , and  $\bar{y} = \bar{\theta}$ . For predicted answers, it follows that

$$\hat{r}_i \xrightarrow{p} \bar{\theta} + \frac{\tilde{\tau}}{1 + \tilde{\tau}} (\theta_i - \bar{\theta}) .$$

For the model fit, it holds that

$$\begin{aligned}
R^2 &= 1 - \frac{\sum_{i=1}^I [(r_i - \hat{r}_i)^2]}{\sum_{i=1}^I \left[ \left( r_i - \frac{1}{I} \sum_{i=1}^I r_i \right)^2 \right]} \xrightarrow{p} 1 - \frac{\mathbb{E} \left[ \left( r_i - \bar{\theta} - \frac{\tilde{\tau}}{1 + \tilde{\tau}} (\theta_i - \bar{\theta}) \right)^2 \right]}{\mathbb{E}[(r_i - \bar{r})^2]} \\
&= 1 - \frac{\mathbb{E} \left[ \left( \bar{\theta} + \frac{\tilde{\tau}}{1 + \tilde{\tau}} (x_i - \bar{\theta}) - \bar{\theta} - \frac{\tilde{\tau}}{1 + \tilde{\tau}} (\theta_i - \bar{\theta}) \right)^2 \right]}{\mathbb{E} \left[ \left( \bar{\theta} + \frac{\tilde{\tau}}{1 + \tilde{\tau}} (x_i - \bar{\theta}) - \bar{\theta} \right)^2 \right]} \\
&= 1 - \frac{\left( \frac{\tilde{\tau}}{1 + \tilde{\tau}} \right)^2 \mathbb{E}[(x_i - \theta_i)^2]}{\left( \frac{\tilde{\tau}}{1 + \tilde{\tau}} \right)^2 \mathbb{E}[(x_i - \bar{\theta})^2]} = 1 - \frac{\frac{\sigma^2}{\tau}}{\sigma^2 + \frac{\sigma^2}{\tau}} = \frac{\tau}{1 + \tau} .
\end{aligned}$$

Rearranging yields that  $R^2/(1 - R^2)$  is a consistent estimator for  $\tau$ .

## C.2 Self-reports as the Independent Variable

For the report as an independent variable, noise in the sense of classical measurement error is well known to induce attenuation bias. However, reduced subjective self-knowledge works as a counter-force, inducing amplification, i.e., making the slope of the regression

line *steeper*. To see the intuition, consider a regression line fitted through just two data points with coordinates  $(r_1, z_1)$  and  $(r_2, z_2)$ . The point estimate for the regression coefficient is then given by  $(z_2 - z_1) / (r_2 - r_1)$ . Reduced subjective self-knowledge attenuates the absolute difference between  $r_1$  and  $r_2$ , thereby increasing the estimate. Formally, assume that we want to estimate the unknown coefficients of the following equation:

$$z_i = \gamma_0 + \gamma_1 \theta_i + \eta_i,$$

where  $z_i$  is the respective realization of the dependent variable and  $\eta_i$  an i.i.d. error term with an expected value of zero that is independent of both  $\theta_i$  and the signals that subjects receive. Again, the unknown true values  $\theta_i$  are replaced with reports  $r_i$ , and the asymptotic result of the standard OLS estimator is derived below.

$$\begin{aligned} \hat{\gamma}_1 &\xrightarrow{p} \frac{\widehat{\text{cov}}(z_i, r_i)}{\widehat{\text{var}}(r_i)} = \frac{\text{cov}(z_i, r_i)}{\text{var}(r_i)} = \frac{\mathbb{E}[(z_i - \bar{z})(r_i - \bar{r})]}{\mathbb{E}[(r_i - \bar{r})^2]} \\ &= \frac{\mathbb{E}\left[(\gamma_0 + \gamma_1 \theta_i + \eta_i - \gamma_0 - \gamma_1 \bar{\theta}) \left(\frac{\bar{\theta} + \tilde{\tau} x_i}{1 + \tilde{\tau}} - \bar{\theta}\right)\right]}{\mathbb{E}\left[\left(\frac{\bar{\theta} + \tilde{\tau} x_i}{1 + \tilde{\tau}} - \bar{\theta}\right)^2\right]} \\ &= \frac{\mathbb{E}\left[(\gamma_1 (\theta_i - \bar{\theta}) + \eta_i) \left(\frac{\tilde{\tau}}{1 + \tilde{\tau}} (x_i - \theta_i + \theta_i - \bar{\theta})\right)\right]}{\left(\frac{\tilde{\tau}}{1 + \tilde{\tau}}\right)^2 \mathbb{E}\left[(x_i - \bar{\theta})^2\right]} \\ &= \frac{\gamma_1 \frac{\tilde{\tau}}{1 + \tilde{\tau}} \text{var}(\theta)}{\left(\frac{\tilde{\tau}}{1 + \tilde{\tau}}\right)^2 [\text{var}(\theta) + \text{var}(x | \theta)]} = \frac{\gamma_1 \sigma^2}{\frac{\tilde{\tau}}{1 + \tilde{\tau}} (\sigma^2 + \frac{\sigma^2}{\tau})} = \frac{1 + \tilde{\tau}}{\tilde{\tau}} \frac{\tau}{1 + \tau} \gamma_1 \\ \hat{\gamma}_0 &\xrightarrow{p} \bar{z} - \gamma_1 \bar{\theta} = \gamma_0 + \gamma_1 \bar{\theta} - \hat{\gamma}_1 \bar{\theta} = \gamma_0 + \left(1 - \frac{1 + \tilde{\tau}}{\tilde{\tau}} \frac{\tau}{1 + \tau}\right) \gamma_1 \bar{\theta} \end{aligned}$$

The overall effect of a reduction in  $\tau$  for the report as the independent variable is thus ambiguous. As it turns out, for subjects that are correctly specified about their self-knowledge as assumed in our benchmark model, the effects cancel out exactly. If a reduction of  $\tau$  results in an excess of subjective self-knowledge, estimates are attenuated. In the opposite case, the reverse applies and estimates are amplified.

In sum, contrary to economists' typical understanding of the effects of measurement error in the context of OLS, our model suggests that for responses from surveys, error in an independent variable might not always induce "innocent" attenuation bias but perhaps no bias at all or even amplification and that it always induces attenuation bias when reports are used as the dependent variable.

## Appendix D Robustness tests

### D.1 Accounting for Selection

Table 10: Predictive power of domain-specific attitudes towards risk, with inverse probability weighting

	Without controls			Including controls		
Sample:	<i>pooled</i> (1)	<i>below</i> (2)	<i>above</i> (3)	<i>pooled</i> (4)	<i>below</i> (5)	<i>above</i> (6)
Dependent variable: <i>Risky financial securities</i>						
Risk attitude	0.0698*** (0.00264)	0.0570*** (0.00380)	0.0826*** (0.00385)	0.0523*** (0.00280)	0.0425*** (0.00409)	0.0616*** (0.00412)
(Partial) $R^2$	0.0827	0.0597	0.108	0.0498	0.0351	0.0652
Observations	9095	4548	4547	7472	3736	3736
$\Delta R^2$	81%, $p < 0.001$			86%, $p < 0.01$		
Dependent variable: <i>Performance pay</i>						
Risk attitude	0.0128*** (0.00174)	0.0108*** (0.00245)	0.0165*** (0.00262)	0.00977*** (0.00199)	0.00736** (0.00274)	0.0144*** (0.00309)
(Partial) $R^2$	0.00870	0.00670	0.0132	0.00487	0.00295	0.00966
Observations	5758	2879	2879	4464	2232	2232
$\Delta R^2$	97%, $p = 0.20$			227%, $p = 0.15$		
Dependent variable: <i>Smoking</i>						
Risk attitude	0.0199*** (0.00154)	0.0161*** (0.00222)	0.0239*** (0.00218)	0.0125*** (0.00175)	0.00982*** (0.00253)	0.0152*** (0.00248)
(Partial) $R^2$	0.0119	0.00755	0.0182	0.00481	0.00291	0.00723
Observations	15162	7581	7581	11652	5826	5826
$\Delta R^2$	141%, $p < 0.01$			148%, $p = 0.12$		

*Note:* The table reports OLS estimates, with binary dependent variables taking the values zero and one. If not stated otherwise, all the data refer to the year 2009. The regressions use inverse probability weights that come from probit regressions of group assignment on gender, a second-order age polynomial, and years of education. If values are missing, we assume probabilities of 1/2. Regressions are based only on respondents who are 18 years or older, and those for performance pay include only respondents up to the age of 66 who work full-time and receive wages or salaries. *Risky financial securities* are, in the SOEP, a residual category of securities without a fixed interest rate, like stocks or options (“other securities”). Since the relevant question was asked on the household level in 2010, the units of observation in the respective regressions are households in that year. *Performance pay* indicates that an employee receives payments from profit-sharing, premiums, or bonuses. Smoking refers to 2010. The variable *risk attitude* in each of the panels refers to the respective domain-specific question asked in the SOEP. The contexts are *financial matters* for holding risky financial securities, *career* for performance pay, and *health* for smoking. The controls used in Columns 4–6 are gender, age, squared age, body height in 2010, years of education, parental education (whether mother and father each have either *Abitur* or *Fachabitur*), log net household income, and log wealth and log debts of the current household in 2007. The last three variables are calculated as  $\ln(\text{euro amount} + 1)$ . For the regressions involving *risky financial securities*, all variables are averaged on the household level, and we base our data only on respondents for whom all information is available individually. The  $p$ -values for the sizes of  $\Delta R^2$  are each based on 10,000 permutations. Heteroskedasticity-robust standard errors in parentheses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

## D.2 Big Five

Table 11 replicates Table 6, analyzing differences in the Big Five traits instead of differences in risk attitudes. The results are qualitatively similar to those observed for risk attitudes and quantitatively even stronger.

Table 11: Differences in Big Five

(a) Agreeableness, conscientiousness, and extraversion

	Unweighted			Weighted	
Sample:	<i>pooled</i> (1)	<i>below</i> (2)	<i>above</i> (3)	<i>below</i> (4)	<i>above</i> (5)
<i>Domain: Agreeableness</i>					
Female	0.345*** (0.0155)	0.274*** (0.0207)	0.415*** (0.0229)	0.271*** (0.0208)	0.436*** (0.0236)
<i>R</i> <sup>2</sup>	0.0297	0.0212	0.0386	0.0209	0.0419
Observations	16359	8180	8179	8180	8179
$\Delta R^2$	83%, <i>p</i> < 0.001			100%, <i>p</i> < 0.001	
Height (in '10)	-0.0172*** (0.000862)	-0.0128*** (0.00118)	-0.0215*** (0.00126)	-0.0126*** (0.00118)	-0.0229*** (0.00129)
<i>R</i> <sup>2</sup>	0.0262	0.0162	0.0369	0.0158	0.0413
Observations	14846	7423	7423	7423	7423
$\Delta R^2$	127%, <i>p</i> < 0.001			161%, <i>p</i> < 0.001	
<i>Domain: Conscientiousness</i>					
Female	0.143*** (0.0157)	0.0760*** (0.0209)	0.210*** (0.0233)	0.0757*** (0.0208)	0.210*** (0.0239)
<i>R</i> <sup>2</sup>	0.00510	0.00163	0.00986	0.00164	0.00976
Observations	16359	8180	8179	8180	8179
$\Delta R^2$	506%, <i>p</i> < 0.001			494%, <i>p</i> < 0.001	
Height (in '10)	-0.00876*** (0.000883)	-0.00512*** (0.00118)	-0.0121*** (0.00131)	-0.00513*** (0.00117)	-0.0118*** (0.00136)
<i>R</i> <sup>2</sup>	0.00677	0.00261	0.0116	0.00265	0.0111
Observations	14846	7423	7423	7423	7423
$\Delta R^2$	344%, <i>p</i> < 0.001			319%, <i>p</i> < 0.001	
<i>Domain: Extraversion</i>					
Female	0.197*** (0.0156)	0.122*** (0.0194)	0.272*** (0.0244)	0.127*** (0.0196)	0.255*** (0.0247)
<i>R</i> <sup>2</sup>	0.00967	0.00481	0.0150	0.00527	0.0132
Observations	16359	8180	8179	8180	8179
$\Delta R^2$	212%, <i>p</i> < 0.001			151%, <i>p</i> < 0.01	
Height (in '10)	-0.00254** (0.000877)	0.0000471 (0.00109)	-0.00469*** (0.00137)	-0.0000377 (0.00111)	-0.00343* (0.00139)
<i>R</i> <sup>2</sup>	0.000569	0.000000254	0.00159	0.000000163	0.000863
Observations	14846	7423	7423	7423	7423
$\Delta R^2$	625381%, <i>p</i> < 0.01			529774%, <i>p</i> < 0.01	

Note: The table continues on the next page.

## (b) Neuroticism and openness

	Unweighted			Weighted	
Sample:	<i>pooled</i> (1)	<i>below</i> (2)	<i>above</i> (3)	<i>below</i> (4)	<i>above</i> (5)
<i>Domain: Neuroticism</i>					
Female	0.435*** (0.0152)	0.359*** (0.0198)	0.511*** (0.0232)	0.360*** (0.0199)	0.516*** (0.0237)
$R^2$	0.0471	0.0386	0.0559	0.0391	0.0564
Observations	16359	8180	8179	8180	8179
$\Delta R^2$	45%, $p < 0.01$			44%, $p < 0.01$	
Height (in '10)	-0.0204*** (0.000862)	-0.0169*** (0.00114)	-0.0241*** (0.00130)	-0.0171*** (0.00116)	-0.0244*** (0.00133)
$R^2$	0.0366	0.0299	0.0444	0.0306	0.0456
Observations	14846	7423	7423	7423	7423
$\Delta R^2$	49%, $p = 0.02$			49%, $p = 0.02$	
<i>Domain: Openness</i>					
Female	0.129*** (0.0156)	0.126*** (0.0208)	0.132*** (0.0232)	0.120*** (0.0209)	0.122*** (0.0239)
$R^2$	0.00415	0.00447	0.00393	0.00407	0.00331
Observations	16359	8180	8179	8180	8179
$\Delta R^2$	-12%, $p = 0.79$			-19%, $p = 0.68$	
Height (in '10)	0.00107 (0.000877)	0.00126 (0.00116)	0.000595 (0.00132)	0.00107 (0.00117)	0.00205 (0.00137)
$R^2$	0.000100	0.000157	0.0000283	0.000113	0.000335
Observations	14846	7423	7423	7423	7423
$\Delta R^2$	-82%, $p = 0.61$			197%, $p = 0.73$	

*Note:* The table reports OLS estimates. All regressions only use respondents who are 18 years or older. The dependent variables are standardized among the respondents who enter the corresponding regression in Column 1. Columns 4 and 5 use inverse probability weights that come from probit regressions of group assignment on gender, a second-order age polynomial, and years of education. If values are missing, we assume probabilities of  $1/2$ . Except for height, all data refer to the year 2009. The  $p$ -values for the respective sizes of  $\Delta R^2$  are each based on 10,000 permutations. Heteroskedasticity-robust standard errors in parentheses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

# Online Appendix: Experimental Instructions

*The instructions have been translated from German. Horizontal lines are used to separate screens.*

---

## Welcome

Welcome and thank you for participating in today's study!

For your participation, you will receive a flat fee of €5, which is going to be paid out in cash at the end of the study. During the study, you will respond to estimation tasks. Depending on the quality of your answers, you can additionally earn up to €10. On the following screens, everything will be explained in detail.

During the study, communication with other participants is not allowed and the curtain of your cubicle has to remain closed. Your cellphone has to be switched off and no aids are permitted. On the computer, only use the designated functions and use the mouse and keyboard to make inputs. If you should have any questions, please stick your hand out of the cubicle. One of the experimenters is then going to approach you.

Please now click on "Continue" to proceed.

---

## Your Task

Generally, your task in this experiment is to estimate the height of stylized depictions of men. The more precisely you estimate, the more money you can earn. For that, you will, later on, see a series of pictures with men of different heights.

More precisely, the men are going to be depicted as "stick figures." An example is shown below.

[Picture of a male stick figure]

The men are split into eleven categories, depending on their body heights:

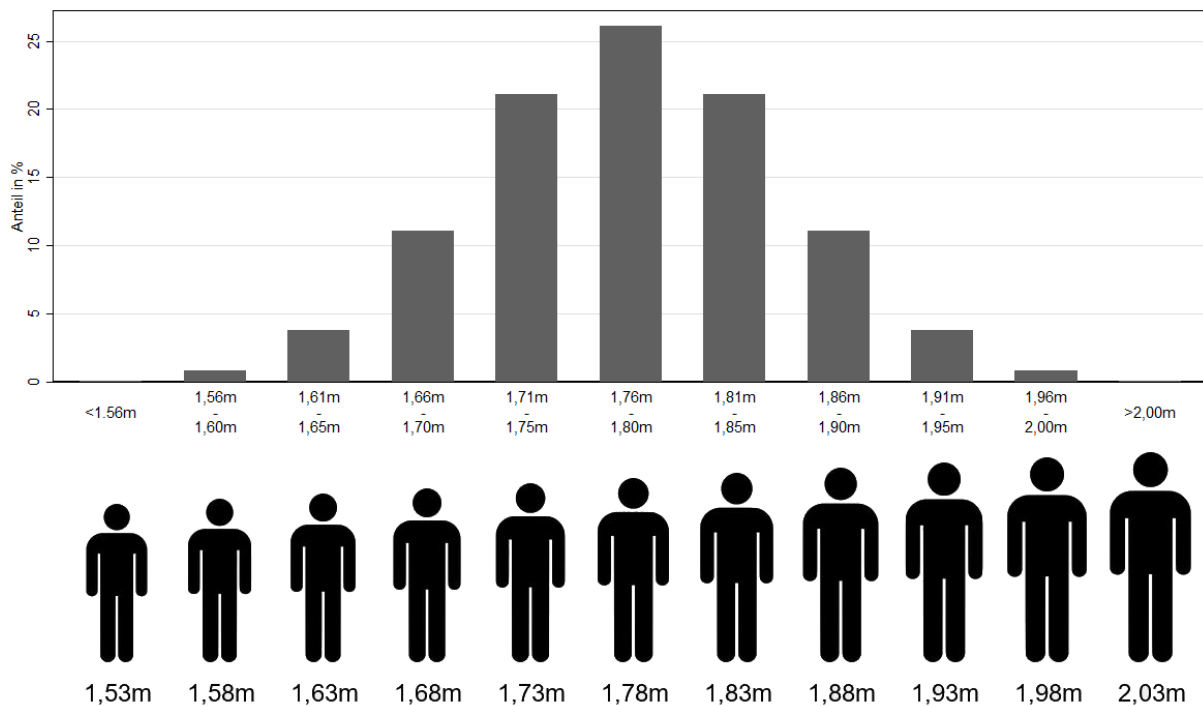
at most 1.55m	1.56m–1.60m
1.61m–1.65m	1.66m–1.70m
1.71m–1.75m	1.76m–1.80m
1.81m–1.85m	1.86m–1.90m
1.91m–1.95m	1.96m–2.00m
at least 2.01m	

---

## Body Heights

As you know, very short and very tall men are found rather infrequently. Most common are men of around 1.78m. Exactly the same holds for the pictures that you are going to see later on. The pictures are informed by the actual height distribution among men in Germany. For that, the data from a large, representative sample of more than 20,000 people in Germany were used. The frequency of observing men of a given height is depicted in the image below.

For your orientation, we have also printed this image for you. It is lying on your desk.



## Body Heights

On the image (on your desk) you see eleven different body heights. For each body height, it is said how often it is observed in the German population. Most common are men of a body height of 1.76m–1.80m, with 26.1%. The second most common are men with body heights of 1.71m–1.75m or 1.81m–1.85m (21.1% each). The third most common is a height of 1.66m–1.70m or 1.86m–1.90m (11.1% each). Considerably less common are very tall and very short men. Heights of 1.61m–1.65m or 1.91m–1.95m occur in only 3.8% of observations, heights of 1.56m–1.60m or 1.96m–2.00m each in only 0.8% of cases. Very uncommon are heights under 1.56m and above 2.00m (0.1% each).

It is important that you understand the relative frequencies of heights since the pictures that will be shown later are drawn from the displayed distribution. Thus, it is considerably more likely that you will see a man with a body height of 1.75m or 1.81m than a man with a body height of 1.58m or 2.03m.

To make the estimation of the body heights easier for you, every picture that will be displayed is accompanied by either a cat or an elephant. The cat has a height of 40cm, and the elephant is 3.50m tall (each at its highest points). In the picture below, you see an average man with a height of 1.78m next to the cat and the elephant, respectively. [two example images here, as described]

---

## Procedure

You will be shown a series of 60 pictures. For this purpose, we will randomly draw 15 different heights from the distribution in the population. Every drawn height will be shown to you four times in total. The accompanying animal and the position on the screen may change.

You will first be shown a countdown in seconds. After the countdown has finished, you will be shown a picture for  $[0.5/7.5]$  seconds. Afterward, the following question will be asked:

How tall was the displayed person?

You can provide your answer on the following scale:

The height of the displayed person was ...

below average						above average				
...–	1.56m–	1.61m–	1.66m–	1.71m–	1.76m–	1.81m–	1.86m–	1.91m–	1.96m–	2.01m–
1.55m	1.60m	1.65m	1.70m	1.75m	1.80m	1.85m	1.90m	1.95m	2.00m	...
○	○	○	○	○	○	○	○	○	○	○

---

## Your Payoff

For each shown picture, there is exactly one correct answer (an interval). For example, if the height of the shown man should be 1.78m, then this would be the answer “1.76m–1.80m.” You always have to select exactly one answer. **At the end of today’s study, one of the shown pictures will randomly be selected for you.** Your answer for this picture then determines the payoff that you receive on top of the €5 flat fee.

If you have chosen exactly the correct option, you will additionally receive €10. The further away you were from the correct answer (how much further to the left or right you should have clicked), the more is deducted from the €10. For this, the deviation (steps to the left or right) is squared and multiplied by 10 cents. The maximal deviation is ten steps (e.g., if you have answered “2.01m–...” but “...–1.55m” would have been correct). In this case, the entire €10 would be deducted.



**You receive more money, the fewer steps are between your selected answer and the correct answer.** The table gives you an overview of the possible deductions and the resulting additional payments. A printed version of this table is also available at your desk.

Deviation (steps)	0	1	2	3	4	5	6	7	8	9	10
Deduction (€)	0.00	0.10	0.40	0.90	1.60	2.50	3.60	4.90	6.40	8.10	10.00
Additional payment (€)	10.00	9.90	9.60	9.10	8.40	7.50	6.40	5.10	3.60	1.90	0.00

## Control Questions

Please respond to a few questions regarding your comprehension. Feel free to use the printout at your desk as an aid.

- In each case, which of the two is more likely: the picture depicts a man with a height of ...
  - 1.76m–1.80m [correct]  
2.01m–...
  - 1.81m–1.85m  
1.76m–1.80m [correct]
  - 1.76m–1.80m [correct]  
1.71m–1.75m
  - 1.66m–1.70m  
1.81m–1.85m [correct]
- How much money would be deducted from the additional €10?
  - Correct would be “1.76m–1.80m.” You responded “2.01m–...” [€2.50]
  - Correct would be “2.01m–...” You responded “...–1.55m.” [€10.00]
  - Correct would be “1.76m–1.80m.” You responded “1.81–1.85m.” [€0.10]
  - Correct would be “1.86m–1.90m”. You responded “1.76m–1.80m.” [€0.40]
- Suppose you have missed the picture of the man, but you nonetheless must give an estimate. What is the best answer? [1.76m–1.80m]

Thank you for your responses! Please wait.

## Trial Run

Before you see the 60 pictures and estimate the heights, there will first be a trial run. You will see ten pictures and subsequently have to estimate the height of the respective man you saw. Unlike later, you are afterward informed about the correct answer.

This trial run is unrelated to the final payout and is meant to introduce you to the task.

The pictures will be displayed for [0.5/7.5] seconds, exactly as in later rounds.

When you are ready, click on “Begin”.

---

Practice task [n]/10

---

[Countdown]

---

[Picture]

---

How tall was the shown person?

The height of the displayed person was ...

below average						above average				
...–	1.56m–	1.61m–	1.66m–	1.71m–	1.76m–	1.81m–	1.86m–	1.91m–	1.96m–	2.01m–
1.55m	1.60m	1.65m	1.70m	1.75m	1.80m	1.85m	1.90m	1.95m	2.00m	...
○	○	○	○	○	○	○	○	○	○	○

---

*Nine more practice rounds.*

---

Correct answer: [e.g., 1.71m–1.75m]

Your answer: [e.g., 1.81m–1.85m]

---

Thank you for your responses! Please wait.

---

## Beginning of the Main Part

Thank you for completing the trial rounds.

You can now begin with the main part of the study. At the end of the study, one of your following responses will be chosen and determine how much additional money you earn.

---

Task [n]/60

---

*60 rounds like the practice rounds but without feedback.*

---

## Further Questions

Thank you for completing the main part.

Please now also respond to a few more additional questions.

How difficult did you feel was the task? [very easy – very difficult; seven-point scale]

How sure were you about your responses? [very unsure – very sure; seven-point scale]

---

## Further Questions

Big Five questionnaire (BFI-S; Gerlitz and Schupp, 2005)

Scale-use module

Bayesian updating question

---

## Personal Details

Your gender: female male diverse

Your age (in years):

Your body height (in cm):

---

Do you have any final comments?

---

Thank you for your participation in this study!

You will receive a flat fee of €5.

In addition, answer no. [n] was chosen to determine your additional payoff. Due to the deviation of your answer from the correct answer you will additionally receive [X] euros and [Y] cents.

We will soon begin with the payouts. Please wait at your seat and keep the curtain of your cubicle closed until your cabin number is called. Then, please enter the adjoining

room and remember to take the card on which your cabin number is printed with you and return it.

---