Limited Self-knowledge and Survey Response Behavior

Armin Falk* Thomas Neuber† Philipp Strack‡

November 22, 2020

Preliminary draft

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 estimates. 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 gaps 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.

Keywords: survey research, rational inattention, lab experiment, non-cognitive skills, preferences

JEL Codes: C83, D83, C91, D91, J24

*briq – Institute on Behavior & Inequality and University of Bonn; armin.falk@briq-institute.org

†University of Bonn; thomas.neuber@uni-bonn.de

‡Yale University; philipp.strack@yale.edu. 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 an increasing 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 or less informative respondents based only on *patterns* of their response behavior. Hence, this paper makes two main contributions. It offers a framework on how to model and understand survey response behavior in general, and it derives a method how 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. The model takes seriously 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 variation in "self-knowledge" reflects the fact that individuals differ in their capacity to retrieve or memorize relevant information about themselves, engage more or less in reflecting who they are, or it may simply reflect a lack of life experience in the domain of interest. Limited self-knowledge is modeled as an imperfect signal that the respondent receives about her true type. The informativeness of the signal varies between respondents. 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, a Bayesian agent's 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 to the variance between characteristics, 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 consistent over time if a person knows herself either very well or not at all. This result cautions against the use of simple consistency 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 is one of our key results. It 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 conditions for the choice of signal precision and discuss implications for how survey quality responds 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 expected 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, respectively. 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 as if they are prone to subjective scale use.

In the second part of the paper, we take the theoretical result—that signal precision about one's type can be inferred from the ratio of between- and within-variance—to the data. 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 (i.e., the variance of responses between items) and the within-variance (i.e., 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. Using simulations, we illustrate the performance of the estimator for realistic sample sizes. We study various combinations of sample size, the number of survey items, and waves (periods). The estimator performs quite 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 0.75.

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 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, on average, underestimated. Second, subjects who are estimated to be more reliable (using the variances of their reports) actually provide more accurate reports. Based on the estimator of self-knowledge, 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 apply our empirical estimator to predict subjects' treatment status, i.e., we are able to predict whether subjects were assigned to the treatment condition with either high or low signal precision, respectively.

Finally, we apply our estimator to a large representative panel data set (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 chose 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 higher for the above-median sample than for the below-median sample. Moreover, the difference in R^2 between the two sub-samples amounts to 45%. 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 381% higher in the above-median sample compared with the below-median sample.

Our paper is related to different strands of 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; Matějka and McKay, 2015; Caplin and Dean, 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 who are well informed 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 effects 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—i.e., deviations in answers that 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 to use 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).[1] Despite not referring to survey measures, a related contribution comes from Beauchamp et al. (2019), who analyze how accounting for the "compromise effect"—whereby subjects tend to give answers close to 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, shows its consistency, 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.

[1] From a psychological perspective, processes underlying 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 (Section 2.1). Second, we derive how patterns of the answering behavior reveal the informational content of responses, providing the intuition for how we later estimate self-knowledge (Section 2.2). 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 (Section 2.3).

2.1 Framework

Introspection and Self-knowledge. The context in which we are interested 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.[2] The agent's true type is denoted by θ, and we assume that it is normally distributed in the population with mean $\bar{\theta}$ and variance σ^2.

When asked about her type θ, 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 mean equal to the agent's type θ and variance σ^2/τ. The parameter $\tau > 0$ hence indicates the precision of the signal relative to the variance in the population. The higher the value of τ, the more precise the signal that an individual receives about herself. We refer to τ as *self-knowledge*.

Response Behavior. After reflecting on her true type θ, the respondent reports her answer. We assume that she seeks to provide a response r that is as precise as possible, i.e., interests of the researcher and respondent are perfectly aligned.[3] 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.,

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

Hence, she reports her best guess of her type $r = \mathsf{E}[\theta \mid x]$. 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

$$r = \frac{\bar{\theta} + \tau\, x}{1 + \tau} \, . \tag{2}$$

[2]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.

[3]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 believed true type for reputational or "social desirability" reasons. For a discussion, see Section 2.3.4.

Intuitively, the higher her self-knowledge τ, the more precise the respondent's signal, and the more weight that 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.

2.2 Response Patterns

We now explore the implications of limited self-knowledge for response patterns. We are particularly interested in the variances in reports, both conditionally and unconditionally, on an agent's type. These variances will allow us to identify differences in self-knowledge.

Expected Report. It follows from Equation 2 that the expected report conditional on the true type θ equals

$$\mathsf{E}[r \mid \theta] = \frac{\bar{\theta} + \tau\,\theta}{1 + \tau}\,. \tag{3}$$

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

Between-variance. Consider now the variance of the expected report conditional on an agent's type. 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 variance between different characteristics), we refer to it as the *between-variance*. It is given by

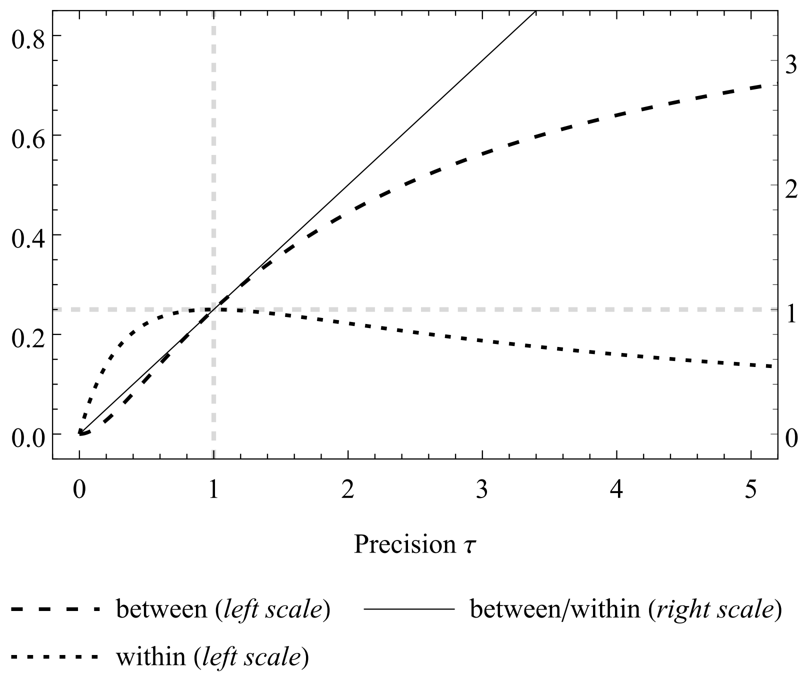
$$\begin{aligned} \sigma^2_{\text{between}} &\equiv \mathsf{var}[\mathsf{E}[r \mid \theta]] = \mathsf{var}\left[\frac{\bar{\theta} + \tau\,\theta}{1 + \tau}\right] \\ &= \left(\frac{\tau}{1+\tau}\right)^2 \mathsf{var}[\theta] = \left(\frac{\tau}{1+\tau}\right)^2 \sigma^2\,. \end{aligned} \tag{4}$$

The between-variance is strictly increasing in self-knowledge τ. 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.

Within-variance. Now suppose that the researcher observes the respondent responding multiple times to questions about the *same* characteristic. We call the resulting variation in answers the *within-variance* of the agent's reports, i.e., the variance conditional on her type. It is given by

$$\begin{aligned}\sigma^2_{\text{within}} \equiv \text{var}[r \mid \theta] &= \text{var}\left[\frac{\bar{\theta} + \tau\, x}{1+\tau}\,\middle|\,\theta\right] \\ &= \left(\frac{\tau}{1+\tau}\right)^2 \text{var}[x \mid \theta] = \frac{\tau}{(1+\tau)^2}\,\sigma^2\,. \end{aligned} \tag{5}$$

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



Note: Variances $\sigma^2_{\text{between}}$ and σ^2_{within} as functions of τ (values on the left axis). The dotted line shows the ration of the two variances, which is equal to τ (values on the right axis).

Figure 1: Theoretical variances

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 τ. As τ 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 τ 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, σ^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.

The discussion reveals that both the between- and within-variance contain information about the respondent's level of self-knowledge τ. While a large between-variance is always "good news," indicating high levels of τ, a low within-variance can reflect either high or low levels of τ, 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:

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

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

Our paper builds on this fundamental insight. We show that the relationship between the variance 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.

2.3 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 in 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 τ can be inferred as the ratio of the between- and within-variance is robust.

2.3.1 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 τ. 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 precision of the signal x is no longer fixed at a given level of σ^2/τ. Instead, τ is chosen by the agent at a cost τ/a for some ability $a \in \mathrm{R}_+$. The utility function (corresponding to Equation 1) equals

$$u_\theta(r,\tau) = -m\,(r-\theta)^2 - \frac{\tau}{a}\,. \tag{7}$$

Here, $m \in \mathrm{R}_+$ measures the motivation of the respondent to provide an accurate answer, and it can be thought of as either extrinsic or intrinsic motivation.[4] Assume that $m\,a > \sigma^{-2}$, 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* $\tau^* = \sqrt{m\,a}\,\sigma - 1$.

The chosen signal precision τ^* is increasing in both incentives m and ability a, i.e., higher incentives or ability generate more precise signals.

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.

2.3.2 Subjective Self-knowledge

The basic framework introduced in Section 2.1 assumes that the respondent knows the relative precision τ 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, respectively. In either case, this will result in a wedge between the optimal and the actual response, again complicating inference about respondents' true types.

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

[4] 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.

under-confident if $\tilde{\tau} < \tau$, and she is over-confident if $\tilde{\tau} > \tau$, respectively. We assume that the agent is naive and when determining her survey response, she applies relative weight 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

$$\sigma^2_{\text{between}} = \left(\frac{\tilde{\tau}}{1+\tilde{\tau}}\right)^2 \sigma^2 \, .$$

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 τ, 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.

$$\sigma^2_{\text{within}} = \left(\frac{\tilde{\tau}}{1+\tilde{\tau}}\right)^2 \frac{\sigma^2}{\tau} \, .$$

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 τ through the variance of the signal.[5]

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

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

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

2.3.3 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 category. However, if response scales are subjectively interpreted, responses may differ. The mapping from an intended response to some scale may depend on individual-

[5]Observe that only for $\tilde{\tau} \to \infty$, the model predicts classical measurement error.

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 estimate of τ that we proposed in 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 r, therefore, two agents may come up with different actual responses. We assume that the agent’s response is scaled away from some point $c \in \mathsf{R}$, e.g., the center of the scale, by a factor $\phi \in \mathsf{R}$. The report and its expected value (corresponding to Equation 3) are then given by

$$r = (1-\phi)\,c + \phi\left(\frac{\bar{\theta} + \tau\, x}{1+\tau}\right), \qquad \mathsf{E}[r \mid \theta] = (1-\phi)\,c + \phi\left(\frac{\bar{\theta} + \tau\, \theta}{1+\tau}\right).$$

Therefore, depending on ϕ, actual responses may be pushed towards the center or the boundaries of the scale, rendering the interpretation of responses more difficult. This holds in particular if ϕ 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

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

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

$$\begin{aligned} \sigma^2_{\text{within}} \equiv \mathsf{var}[r \mid \theta] &= \mathsf{var}\left[(1-\phi)\,c + \phi\,\frac{\bar{\theta} + \tau\,x}{1+\tau} \,\middle|\, \theta\right] \\ &= \phi^2\left(\frac{\tau}{1+\tau}\right)^2 \mathsf{var}[x \mid \theta] = \phi^2\,\frac{\tau}{(1+\tau)^2}\,\sigma^2\,. \end{aligned}.$$

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

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

2.3.4 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 respondents feel more comfortable reporting that they are an honest person rather than a dishonest person. Such concerns can be integrated into our framework by adding a desirable answer $d \in$ 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

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

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 θ and the desirable answer

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

The respondent thus acts as if subject to subjective scale use, as introduced in Section 2.3.3. The main difference between subjective scale use and desirability arises in the context of multiple agents and characteristics: while the scale use parameters (ϕ, c) are naturally agent-specific, the desirability parameters (ψ, 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 of 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 θ_{ik} the value of the k^{th} characteristic for agent i and assume that characteristics are independently normally distributed in the population with a mean $\bar{\theta}$ and variance σ^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 a mean θ_i and variance σ_i^2/τ_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 τ_i. In Section 2, we have shown that τ equals the 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 between-variance and the average variance of answers for a given characteristic as an approximation of the within-variance. Denote agent i's average answer for 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

$$\hat{\tau}_i = \frac{\frac{1}{K-1}\sum_{k=1}^{K}\left(\bar{r}_{ik} - \bar{r}_i\right)^2}{\frac{1}{K(T-1)-2}\sum_{k=1}^{K}\sum_{t=1}^{T}\left(r_{ikt} - \bar{r}_{ik}\right)^2} - \frac{1}{T}\,. \tag{8}$$

The enumerator in the first part 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[6] and a correction term of $1/T$ is subtracted from the ratio. These two adjustments are necessary to ensure that the estimator is unbiased.

Theorem 1 establishes that $\hat{\tau}_i$ is a consistent, unbiased estimator of self-knowledge τ_i and describes its properties.

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

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

$$\hat{\tau}_i = \left(\tau_i + \frac{1}{T}\right)\frac{K(T-1)-2}{K(T-1)}F_i - \frac{1}{T} \tag{9}$$

   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 τ_i, i.e., $\mathsf{E}[\hat{\tau}_i \,|\, \tau_i] = \tau_i$.*

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

$$\sqrt{\mathsf{E}[(\hat{\tau}_i - \tau_i)^2 \,|\, \tau_i]} = \left(\tau_i + \frac{1}{T}\right)\sqrt{\frac{2((K-1)+K(T-1)-2)}{(K-1)(K(T-1)-4)}}\,. \tag{10}$$

4. *$\hat{\tau}_i$ is a consistent estimator and converges to τ_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 the number of repeated observations T:*

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

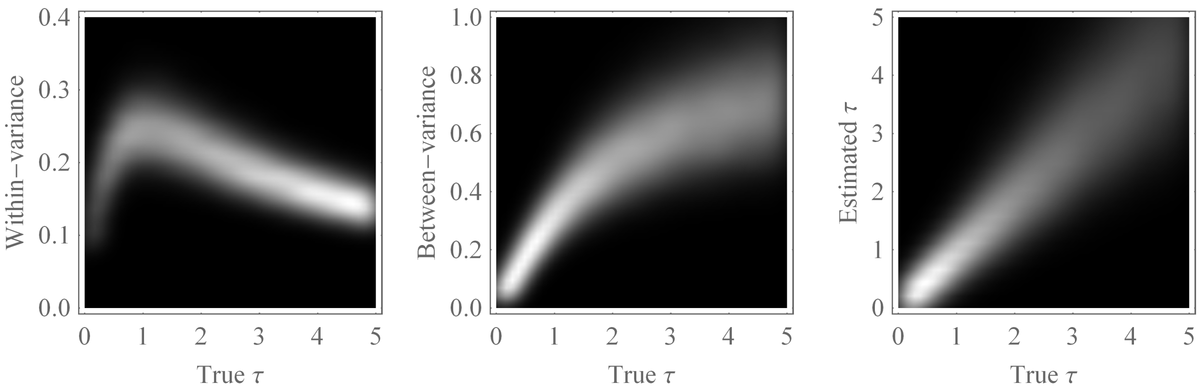
The proof of Theorem 1 is provided in Appendix A. Part 4 of Theorem 1 shows that additional questions are more valuable than additional waves for retrieving precise

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

estimates. 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 1. *As we show in the proof of Theorem 1 in Appendix A, the properties of the estimator derived extend unchanged to the model with endogenous e ort, subjective self-knowledge, and subjective scale use. We state the properties here without these extensions for the ease of exposition.*

Numerical Illustration of the Estimator. Next, we illustrate our model and the behavior of the estimator using numerical simulations. For all illustrations, agents' levels of self-knowledge τ_i are drawn from the uniform distribution between 0.1 and 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 σ^2 equals 1.



(a) Within-variance (b) Between-variance (c) Estimated τ

Note: Kernel-density estimates, where lighter shading corresponds to a higher estimated density.[7] Each panel is based on the same 100,000 simulations, each with I =1,000 hypothetical individuals, for whom reports about $K = 50$ characteristics are observed $T = 3$ times.

Figure 2: Simulations

Figure 2 displays the joint distribution of the true level of self-knowledge τ_i, the sample within-variance, the sample between-variance, and estimated self-knowledge $\hat{\tau}_i$. 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 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, and

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	0.673	0.667	0.871	0.762	0.911
Rank corr.	0.751	0.757	0.899	0.822	0.929
Median split	79.0%	79.3%	88.1%	82.9%	90.4%

report the average value of three measures for the quality of our estimates: Pearson's correlation and Spearman's rank correlation between estimated and true self-knowledge and the proportion of simulated agents correctly assigned to having a level of self-knowledge above or below its median value based on estimated self-knowledge. 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 0.673 and 0.751 shown in Column 1 for $I = 100$, $K = 15$, $T = 3$ suggest that the estimator is already informative about self-knowledge for modest sample sizes. This is confirmed by 79.0% 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. This has virtually no effect on the quality of predictions, 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 (in line with Part 4 of Theorem 1). Relative to these increases, the increased power stemming from a higher number of answers per characteristic in Column 4 (ten instead of three) is rather small but still noticeable (in line with Part 4 of Theorem 1, 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 90.4%. Summarizing, 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 τ) while at the same time *knowing* the true state θ. Accordingly, we can study whether our estimator is successful

[7] Gaussian kernel; bandwidth selection according to Silverman's rule.

in identifying subjects whose reports are relatively more informative than those of others. In addition, we *exogenously* vary the quality of the signal subjects receive about true types. In particular, we run two treatments with either high or low signal quality and test whether our estimator of $\hat{\tau}$ is capable of predicting subjects' treatment status, i.e., whether a subject received high- or low-quality signals. These tests are difficult with non-experimental data unless true states are known to the researcher and the precision of signals can be exogenously varied.

4.1 Design of Experiment

To create a choice environment with known types θ and an exogenous variation in knowledge τ, 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 (θ_i), we therefore 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.

Table 2: Choice categories

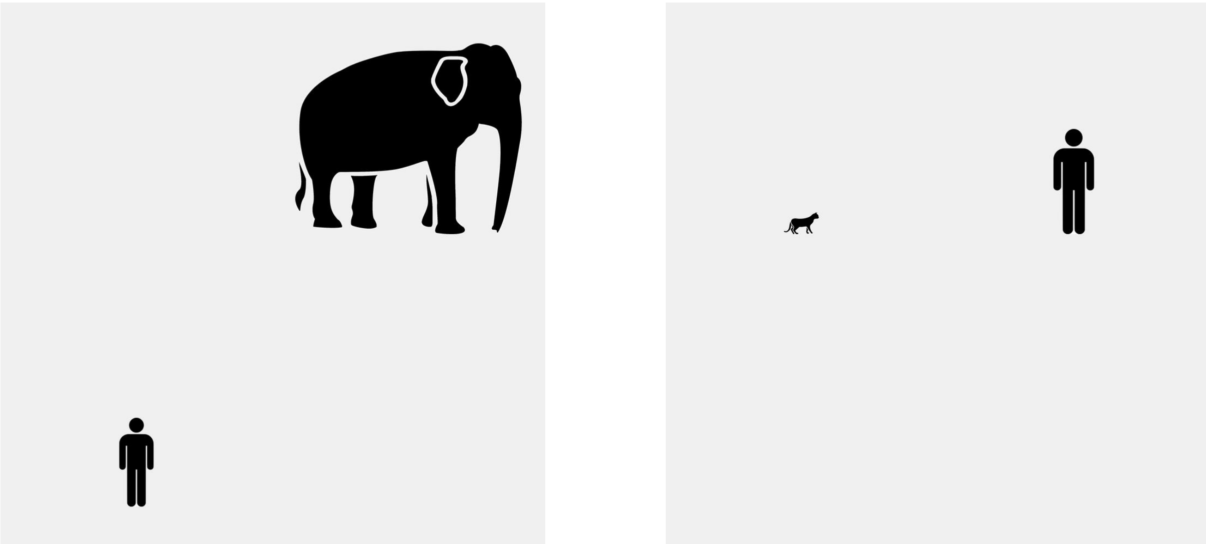
0	1	2	3	4	5	6	7	8	9	10
<1.56	1.56–1.60	1.61–1.65	1.66–1.70	1.71–1.75	1.76–1.80	1.81–1.85	1.86–1.90	1.91–1.95	1.96–2.00	>2.00
0.1%	0.8%	3.8%	11.1%	21.1%	26.1%	21.1%	11.1%	3.8%	0.8%	0.1%

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

For example, Category 3 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 Instructions in the Appendix).

Subjects were informed that a total of 15 distinct sizes were independently drawn from the eleven categories and shown four times. Specifically, subjects were told that they would see four blocks each comprising these 15 distinct sizes. Within each block, the order and location of the male figures were random. 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.

To facilitate the estimation task and vary the presentation style of the screens, next to a 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 3.40 meters and the cat 0.40 meters, respectively. 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.



(a) Example 1 (b) Example 2

Note: Two examples of signals that were shown to subjects. The panel on the left shows a male figure with a height of 1.63m along with the elephant, which is 3.40m tall. The male figure on the right corresponds to 1.93m and the cat has a height of 40cm. Animal pictograms by Hannah Storey.

Figure 3: Example screens

Payoff Function. Subjects had an incentive to estimate the shown size of the male figure as precisely as possible. The payoff function, π, implements a quadratic loss function and is exactly the same as in the model (see Equation 1) with

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

where θ indicates the true type (size of the male figure) and r a subject’s report. For payment, 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 in fact saw 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 €$10-(1-8)^2\times$€$0.10=$€5.10. Note that we chose an endowment of €10 to rule out losses even if the difference between true and estimated type was maximal.

Signal Precision and Treatments. To exogenously vary the precision τ of the signal, we ran two treatments that only differed in terms of how long subjects saw each of the 60 screens. In 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. Within a lab session, treatments were randomly assigned. 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 the true type and biased towards the average of the true types, which in the experiment is a value of 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 only uses the assumption that knowledge τ is finite for any subject. Hypotheses 2–4 exploit individual-specific information about τ, either in terms of treatment differences (Long vs. Short) or using the estimator introduced in Section 3.

Hypothesis 2. *Estimates $\hat{\tau}$ should be larger in the Long treatment than in the Short treatment.*

An implication of Hypothesis 2 is that the estimates for τ 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 the treatment to which a given subject was assigned.

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

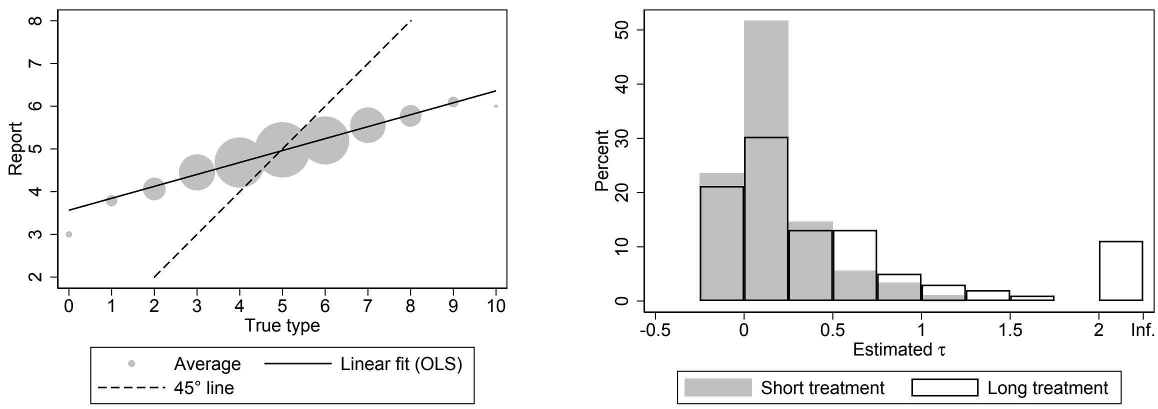
Hypothesis 3. *The lower a given subject's τ, the stronger the reports' bias towards the average value of the characteristic, i.e., five.*

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 τ, which may reflect ability or treatment status.

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

Hypothesis 4 is our main hypothesis. It states that using reports of subjects who have a high empirical estimate of $\hat{\tau}$ yields higher explanatory power of estimates 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 1, i.e., answers are biased towards the sample mean (see Column 1 of Table 3 below for details).



(a) Reports vs. true types

(b) Estimates of τ by treatment

Figure 4: Results from the experiment

To test the further hypotheses, we apply the estimator from Equation 8 to our experimental data. Recall that each given size that a subject saw was shown four times.

Therefore, we treat the respective four answers given by a subject as referring to the same characteristic. Hence, we observe $K = 15$ characteristics and $T = 4$ waves.[8] Figure 4b shows the distribution of $\hat{\tau}$, separately for the Short and the Long treatment (gray and transparent, respectively).

In support of Hypothesis 2, estimates of τ 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 τ yields a significant positive coe□cient value (average treatment e□ect $= 0.37$; $p < 0.001$, two-sided).

For the tests of Hypotheses 3 and 4, we turn to Table 3. Column 1 corresponds to

Table 3: Relationship between true types and reports

	Dependent variable: Report				
Subjects	*all*	*by treatment*		*by* $\hat{\tau}$	
		Short	*Long*	*low*	*high*
	(1)	(2)	(3)	(4)	(5)
True type	0.279***	0.190***	0.366***	0.162***	0.404***
	(0.0167)	(0.0169)	(0.0260)	(0.0138)	(0.0250)
Constant	3.565***	3.973***	3.177***	4.100***	3.012***
	(0.0799)	(0.0859)	(0.122)	(0.0741)	(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
ΔR^2		139%, $p < 0.001$		394%, $p < 0.001$	

Note: 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 τ are available. Tests for the significance of differences in R^2 are each based on 10,000 simulations randomly assigning relevant subjects to counterfactual groups of corresponding sizes. Standard errors clustered at the subject level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the fitted line shown in Figure 4a and regresses reports on true types for the full sample. Columns 2 and 3 replicate Column 1 separately for the two treatments, Short and Long. In comparison with the pooled sample, slopes are less steep for the Short and steeper for the Long treatment, respectively. The three possible pairwise di□erences in slopes (full sample, Long and Short treatment) are all statistically significant ($p < 0.001$). This is in line with a successful treatment manipulation of τ 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 τ, the estimation coe□cient between report and true type is larger than for below-median subjects (Column 4) and the whole sample (Column 1). Again, the three possible pairwise

[8] We use the information that, e.g., signals 3, 18, 33, and 48 referred to the same underlying type, but we do not use any information about what that type was.

di[?]erences are statistically significant ($p < 0.001$).[9]

To test Hypothesis 4, which states that the predictive power of estimating true types should increase in a population's level of τ, we compare the respective values of R^2 for the full sample (Column 1) and all sub-samples (Columns 2–5), respectively. The data confirm our hypothesis: relative to the Short treatment, the value of R^2 in the Long treatment is more than doubled (compare Columns 2 and 3; $p < 0.001$).[10] For the two sub-samples based on the estimator $\hat{\tau}$, the e[?]ects are even larger (compare 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 ($p < 0.001$). In addition to supporting our hypothesis, these comparisons show that the estimator $\hat{\tau}$ is more informative than knowledge about a subject's 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 self-knowledge.

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 di[?]erent true states, and it is therefore informative about τ. Moreover, the individual slope parameter reveals how much weight is assigned to signals, and thus it is informative about the level of subjective self-knowledge $\tilde{\tau}$. Several observations can be made. First, the individual regression levels of R^2 and the respective slope parameters are strongly positively related, with a rank correlation of 0.83 ($p < 0.001$, two-sided, $N = 188$).[11] This positive correlation supports the central assumption of the model that agents place more weight on their signal (measured in terms of higher regression slopes), the higher their level of self-knowledge (measured in terms of higher values of R^2). Second, using the individual values of R^2 allows us to provide a further test of the validity of our estimator from Section 3, which does *not* use information about true types. In fact, we find that the values of R^2 are strongly correlated with the individual values of $\hat{\tau}$. The rank correlation is 0.83 ($p < 0.001$, two-sided, $N = 188$). In light of our model, this relationship can be analyzed even more thoroughly. The R^2 can be transformed into an "objective" measure of τ according to the

[9] Note that the differences between sub-samples in Columns 4 and 5 are more pronounced than those between Columns 2 and 3: the coefficient in the low-τ sub-sample (Column 4) is smaller than the one for the Short treatment in Column 2 ($p = 0.060$) and the high-τ coefficient in Column 5 is larger than the Long treatment coefficient in Column 3 ($p = 0.045$).

[10] To calculate p-values, we replicated the respective analyses 10,000 times but with respondents being allocated to the two sub-groups at random. Each of the counterfactual analyses resulted in an absolute difference between the values of R^2 for the two artificially-created groups. The respective p-value is the share of these absolute differences that were larger than the one observed for the actual sample-split according to the estimates $\hat{\tau}$. The procedure was applied for Table 3, as well as for Tables 5, 6, and 13.

[11] 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.

formula $\tau = R^2/(1 - R^2)$. The Pearson correlation between this objective measure τ and our estimate $\hat{\tau}$ is 0.98 ($p < 0.001$, two-sided, $N = 188$). This relationship is not mechanic since the identification of the two measures rests on entirely different properties of the data: the objective measure uses true states, while our 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 patterns of behavior. An alternative to using this estimator is to simply ask respondents directly how accurate or reliable they think that their responses are. The use of such survey items is widespread and common practice. At the end of the experiment, we asked two such items and can compare their discriminatory power relative to our estimate $\hat{\tau}$. In particular, we asked subjects "how difficult" they thought the estimation task had been and "how reliable" they thought their answers were. Both questions were asked on a seven-point Likert scale. Reassuringly, responses to these two items are strongly negatively correlated ($r = -0.59$; $p < 0.001$, two-sided, $N = 199$). 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 ($p = 0.45$, $N = 188$). However, the rank correlation between self-reported precision and the individual-level values of R^2 is only 0.08 ($p = 0.29$, $N = 188$), i.e., very small and, in particular, much smaller than the respective correlation between R^2 and $\hat{\tau}$ ($p = 0.83$). These results suggest that—in contrast to our estimator—, the survey items of self-reported precision contain little 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 are based on self-reports. In particular, we estimate τ using answers to the Big Five personality inventory from multiple waves and split the sample by the median level 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 = 21,135). The estimator introduced in Section 3 assumes that expected values of types are identical across characteristics and that characteristics are independent. Empirically, however, average answers may differ for different characteristics. To estimate self-knowledge, we therefore suggest subtracting population means as follows:[12]

1. For each question, we calculate the average response over all waves. Then, for each individual response, we subtract the respective average to retrieve demeaned responses.

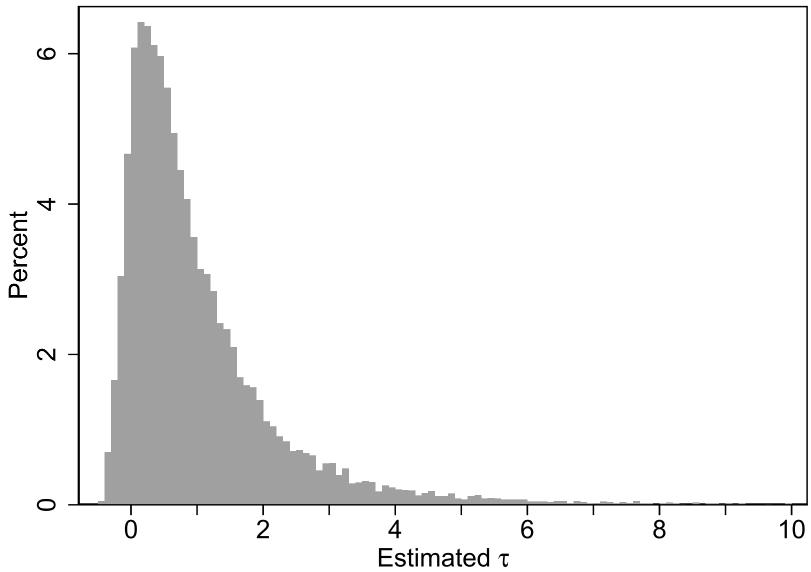
2. We apply the estimator from Equation 8 to the demeaned responses.

Figure 5 shows the distribution of $\hat{\tau}$ for the SOEP population. Apparently, there is substantial heterogeneity in these estimates, suggesting substantial variation in latent self-knowledge. The median value is 0.67, and for about 63% of respondents the estimate $\hat{\tau}$ is smaller than one.

The main focus of this section is to show how the 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 discuss potential determinants of τ, 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 not significantly correlated with gender. With respect to age, Column 2 reveals a hump-shaped relationship with self-knowledge. Descriptively, the latter increases over the life cycle, peaking at an age of about 40 before it mildly declines. However, effect sizes and values of R^2 are fairly small. Given that self-knowledge might reflect differences in cognitive skills, we also consider an effect of education (see Column 3). The correlation is significant and indicates that one more

[12] In Appendix B.1, we show that this slightly modified estimator is consistent and its properties are virtually identical to those of the one introduced in Equation 8. In Appendix B.2, we also show that the estimator is informative if the characteristics used to estimate τ are correlated.



Note: Distribution of $\hat{\tau}$ in the German SOEP. Estimates that are larger than ten (60 out of 21,135) are omitted.

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

year of education is on average associated with an increase of about 0.07 in the level of self-knowledge. In Column 4, we regress self-knowledge simultaneously on all determinants. Education seems to dominate, as becomes apparent when comparing the values of R^2 between columns. However, even the combined R^2 of 0.038 is fairly low, suggesting that the estimates of self-knowledge contain much information above and beyond socio-demographic characteristics.

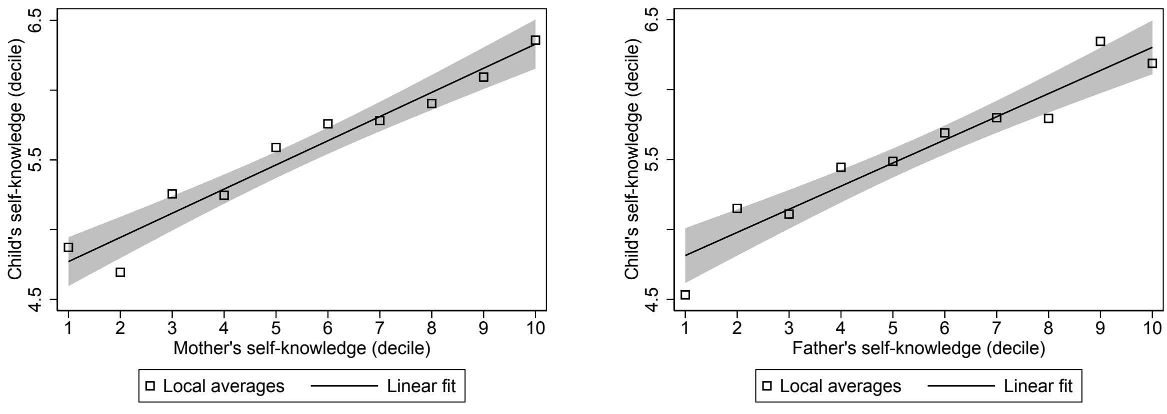
If self-knowledge is a trait, it might be intergenerationally transmitted, similar to, e.g., risk aversion, trust, patience, and social preferences (Dohmen et al., 2012; Kosse and Pfei er, 2012; Alan et al., 2017; Kosse et al., 2020). Such transmission could come from various sources, e.g., imitation, exposure to similar social environments, or genetic dispositions. Using SOEP household panel data, we can match 3,569 respondents to their mothers and 2,960 respondents to their fathers, respectively. Figure 6 shows the relationship between the estimated levels of self-knowledge for parents and their children. 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). Parents' and children's estimated levels of self-knowledge are positively correlated. The rank correlation coe cients of children's estimates of $\hat{\tau}$ with those of their parents are 0.17 ($p < 0.001$) for mothers and 0.16 for fathers ($p < 0.001$).[13]

[13]Differences in slopes or intercepts for fathers and mothers are not statistically significant.

Table 4: Correlations with τ

	Dependent variable: $\hat{\tau}$			
	(1)	(2)	(3)	(4)
Female	-0.00372			0.0249
	(0.0149)			(0.0162)
Age (in '11)		0.0104***		0.00524
		(0.00235)		(0.00291)
Age^2 (in '11)		-0.000135***		-0.0000750**
		(0.0000223)		(0.0000272)
Edu. years (in '11)			0.0722***	0.0698***
			(0.00313)	(0.00317)
Constant	0.975***	0.838***	0.0774*	0.0464
	(0.0109)	(0.0577)	(0.0382)	(0.0808)
Observations	20923	20923	16151	16151
R^2	0.000	0.005	0.036	0.038

Note: Table reports OLS estimates. Individuals for whom $\hat{\tau}$ lies above the 99^{th} percentile are excluded. Age as well as years of education refer to the year 2011. Heteroskedasticity-robust standard errors in parentheses. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$.



(a) From mother (b) From father

Note: Graphs plot average percentiles to which children's $\hat{\tau}$ belong for the deciles of $\hat{\tau}$ for mothers (left) and father (right). Deciles refer to the full sample, i.e., thresholds are the same for parents and for their children. Shaded areas around the regressions lines represent 95% confidence intervals.

Figure 6: Intergenerational transmission of self-knowledge

5.2 Predicting Outcomes

To illustrate how we can increase the explanatory power of non-cognitive skills by accounting for individual estimates of $\hat{\tau}$, we study the effect of risk attitudes on various economic outcomes. Similar to the analysis in Table 3, in Table 5 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. To give coefficients a natural interpretation, self-reported risk attitudes are standardized within the given (sub-)sample.[14] 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 in the respective columns. Table 5 presents the empirical results for three different economic outcomes, related to risk attitudes: receiving performance related pay, holding stocks, 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 items that ask about willingness to take risks in specific domains on eleven-point Likert scales. In particular, the items refer to willingness to take risks concerning one's *career*, *financial matters*, 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, including age, gender, body height, parental education, log wealth, log debts, and log gross household income. Columns 1 and 4 consider the full sample and confirm a positive and significant relationship between risk attitudes and the respective outcomes. Our main interest is the comparison of Columns 2 and 3 and Columns 4 and 5, where we show results for individuals with estimated levels of self-knowledge below or above the median, respectively. In all instances, the values of R^2 are higher for individuals with high self-knowledge than for individuals with low self-knowledge or the full sample, respectively. This holds both with and without controls included in the regressions (in the regressions with controls, we refer to the partial R^2).[15] In most cases, the explanatory power among high self-knowledge respondents is multiple times larger than among the ones with low self-knowledge, ranging from a 74% increase (smoking, without controls) up to an increase of 404% (performance pay, with controls). These differences are also statistically significant, as shown by the respective p-values displayed in Table 5 (for the construction of p-values, see Footnote 10). We note that these results hold for a non-cognitive skill (risk attitudes) that is different and mostly unrelated to the set of traits that we used to

[14] The qualitatively identical results with non-standardized risk attitudes are reported in Table 10 in Appendix C.1

[15] In Appendix B.3, we discuss differences in the estimated coefficients for the respective sub-samples.

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

	Without controls			Including controls		
Sample:	*pooled* (1)	*below* (2)	*above* (3)	*pooled* (4)	*below* (5)	*above* (6)
	Dependent variable: *Investing in stocks*					
Risk attitude	0.113***	0.0900***	0.132***	0.0919***	0.0716***	0.111***
	(0.00397)	(0.00573)	(0.00538)	(0.00426)	(0.00606)	(0.00595)
(Partial) R^2	0.0566	0.0361	0.0774	0.0396	0.0238	0.0574
Observations	15202	7601	7601	12081	6041	6040
ΔR^2		115%,	$p < 0.001$		142%,	$p < 0.001$
	Dependent variable: *Performance pay*					
Risk attitude	0.0384***	0.0225**	0.0513***	0.0299***	0.0179*	0.0408***
	(0.00516)	(0.00705)	(0.00748)	(0.00568)	(0.00775)	(0.00829)
(Partial) R^2	0.00918	0.00326	0.0158	0.00559	0.00202	0.0102
Observations	5612	2806	2806	4509	2255	2254
ΔR^2		386%,	$p < 0.01$		404%,	$p = 0.05$
	Dependent variable: *Smoking*					
Risk attitude	0.0452***	0.0398***	0.0526***	0.0298***	0.0197***	0.0403***
	(0.00351)	(0.00492)	(0.00503)	(0.00395)	(0.00551)	(0.00567)
(Partial) R^2	0.0119	0.00924	0.0161	0.00517	0.00231	0.00924
Observations	15146	7573	7573	12122	6061	6061
ΔR^2		74%,	$p = 0.06$		300%,	$p = 0.01$

Note: Table reports OLS estimates. The variable *risk attitude* in each of the panels refers to the respective domain-specific question asked in the SOEP, standardized within the given (sub-)sample. The contexts are *financial matters* for investing in stocks, *career* for performance pay, and *health* for smoking. The controls used in Columns 4–6 are gender, age, body height, parental education, log wealth, log debts, and log gross household income. All data are from 2009, except for information on body height and smoking, which due to a lack of availability, are from 2010. The model for performance pay includes only respondents up to the age of 66 who are working full-time and who are not self-employed. Tests for the significance of differences in R^2 are each based on 10,000 simulations randomly assigning relevant subjects to counter-factual groups of corresponding sizes. Heteroskedasticity-robust standard errors in parentheses. $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.

estimate $\hat{\tau}$ (the Big Five). This suggests that self-knowledge does, in fact, carry over to different facets of one's personality.

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 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 gender and height effects on domain-specific risk attitudes to illustrate that accounting for differences in self-knowledge, exogenous determinants of preferences may actually have higher explanatory power and yield larger effect sizes than typically assumed. Again, the measures of risk attitudes are standardized within sub-samples.[16]

Table 6 reports the effects of gender and body height on six measures of risk attitudes, all measured on an eleven-point Likert scale in the 2009 wave of the SOEP. The measures comprise the so-called general risk question, asking about willingness to take risks "in general," as well as five domain-specific risk questions, referring to risk taking in the domains of *car driving*, *financial matters*, *sports/leisure*, *career*, as well as *health*.

Replicating previous findings,[17] women tend to be less willing to take risks than men (see Column 1), and taller people tend to be more willing to take risks than smaller individuals (Column 4). Our interest here is to compare samples with high vs. low levels of self-knowledge, i.e., the comparison of Columns 2 and 3 as well as 5 and 6. For all twelve possible comparisons (six risk domains, two determinants), we consistently find that explanatory power as measured in terms of R^2 is larger among high-τ individuals than among low-τ individuals. These differences are often quite substantial and range from 19% (effect of height on career) up to 108% (gender effect on health). They are also mostly statistically significant, as indicated by the respective p-values (for the construction of p-values, see Footnote 10). These 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-$\hat{\tau}$ counterparts.

A closer inspection of estimated coefficients in Columns 2, 3, 5, and 6 further shows that the size of coefficients is always larger for the above-median sample than for the below-median sample. These differences are often significant. 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-τ relative to low-τ subjects (see Table 3, Columns 4 and 5).

A potential concern regarding the above interpretation 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. While, in principle, we cannot rule out such an interpretation with non-experimental data, we believe that it is unlikely in the present case for several reasons. First, recall from Table 4 that the effects of socio-demographic characteristics on $\hat{\tau}$ were rather small. Second, if we account for selection using inverse probability weighting, the results remain very similar. This is shown in Table 12 in Appendix C.2, where we report regression results that

[16] For qualitatively identical results with non-standardized risk attitudes, see Table 11 in Appendix C.1

[17] See in particular Dohmen et al. (2011) but also Croson and Gneezy (2009) and Falk et al. (2018).

replicate the above results based on weights that restore representatives. This suggests that the effects observed for subjects with high self-knowledge bring us closer to accurate measurement, suggesting that both the “true” explanatory power and the size of the coefficients are at least as high for the *entire* sample as those reported for the above-median respondents.

We conclude this section with a brief discussion on the effects of gender and height on the Big Five personality traits. Table 13 in the Appendix 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-τ relative to low-τ 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 effect is more than three times as large for the high-τ compared with the low-τ individuals and that the difference in R^2 amounts to almost 1300%. All effects shown in Table 13 remain qualitatively similar when we account for inverse probability weighting, as shown in Table 14.

Table 6: Di erences in risk attitudes

	Gender			Height		
Sample:	*pooled* (1)	*below* (2)	*above* (3)	*pooled* (4)	*below* (5)	*above* (6)
			Risk domain: *general*			
Coefficient	-0.374***	-0.344***	-0.404***	0.0220***	0.0202***	0.0235***
	(0.0153)	(0.0217)	(0.0215)	(0.000866)	(0.00121)	(0.00125)
R^2	0.0348	0.0295	0.0406	0.0425	0.0346	0.0502
Observations	16633	8317	8316	15103	7552	7551
ΔR^2		38%, $p = 0.04$			45%, $p = 0.02$	
			Risk domain: *car driving*			
Coefficient	-0.403***	-0.375***	-0.433***	0.0248***	0.0232***	0.0260***
	(0.0156)	(0.0222)	(0.0220)	(0.000893)	(0.00125)	(0.00128)
R^2	0.0407	0.0352	0.0468	0.0534	0.0456	0.0602
Observations	15742	7871	7871	14314	7157	7157
ΔR^2		33%, $p = 0.07$			32%, $p = 0.05$	
			Risk domain: *financial matters*			
Coefficient	-0.376***	-0.332***	-0.420***	0.0190***	0.0163***	0.0212***
	(0.0154)	(0.0219)	(0.0218)	(0.000896)	(0.00125)	(0.00129)
R^2	0.0352	0.0275	0.0440	0.0316	0.0225	0.0407
Observations	16523	8262	8261	15011	7506	7505
ΔR^2		60%, $p < 0.01$			81%, $p < 0.01$	
			Risk domain: *sports/leisure*			
Coefficient	-0.323***	-0.299***	-0.349***	0.0231***	0.0206***	0.0250***
	(0.0155)	(0.0220)	(0.0219)	(0.000875)	(0.00124)	(0.00125)
R^2	0.0261	0.0224	0.0303	0.0470	0.0361	0.0563
Observations	16344	8172	8172	14848	7424	7424
ΔR^2		36%, $p = 0.11$			56%, $p < 0.01$	
			Risk domain: *career*			
Coefficient	-0.275***	-0.259***	-0.291***	0.0215***	0.0206***	0.0219***
	(0.0164)	(0.0233)	(0.0232)	(0.000912)	(0.00127)	(0.00131)
R^2	0.0189	0.0167	0.0212	0.0410	0.0363	0.0434
Observations	14618	7309	7309	13251	6626	6625
ΔR^2		27%, $p = 0.32$			19%, $p = 0.31$	
			Risk domain: *health*			
Coefficient	-0.310***	-0.254***	-0.366***	0.0173***	0.0145***	0.0195***
	(0.0154)	(0.0218)	(0.0217)	(0.000868)	(0.00123)	(0.00123)
R^2	0.0240	0.0161	0.0335	0.0263	0.0179	0.0345
Observations	16625	8313	8312	15096	7548	7548
ΔR^2		108%, $p < 0.001$			93%, $p < 0.01$	

Note: Table reports OLS estimates. Tests for the significance of differences in R^2 are each based on 10,000 simulations randomly assigning relevant subjects to counterfactual groups of corresponding sizes. Heteroskedasticity-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

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 preferences, 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 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, such as 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 et al., 1986; Ratcliff and McKoon, 2008).[18] 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, in particular 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 is the provision of incentives in experiments, which may increase the accuracy of responses (see Section 2.3.1) but do not solve issues of lack of self-knowledge, 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 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-

[18]Fudenberg, Strack, and Strzalecki (2018) provide a theoretical analysis of the relationship between response times and the accuracy of binary decisions in the drift-diffusion model.

knowledge based on response patterns, in particular 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 panel data set. We show that the estimator reliably identifies individual differences in self-knowledge in the laboratory context where we know true states. Splitting the lab sample in high vs. low self-knowledge individuals, 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 how to integrate 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.

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. 2019. "Measuring and controlling for the compromise e□ect when estimating risk preference parameters". *Experimental Economics*: forthcoming.

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 E□ects 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 Institite 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): 3651–84.

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., 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.

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, et al. 1986. *Response times: Their role in inferring elementary mental organization*. 8. Oxford University Press on Demand.

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". In *Carnegie-rochester conference series on public policy*, 49:317–356. Elsevier.

Smith, Vernon L. 1976. "Experimental Economics: Induced Value Theory". *American Economic Review* 66 (2): 274–279.

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 (1): 16–32.

Appendix A Proofs

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

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

Plugging into Equation 7 yields that the utility of the agent given the optimal response above equals

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

and, consequently the agent choose her precision τ to maximize,

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

The first-order condition yields the optimal level of effort for an interior solution.

$$0 \overset{!}{=} \frac{\partial\, \mathsf{E}[u(r,\tau)]}{\partial\, \tau} = \frac{m\,\sigma^2}{(1+\tau)^2} - \frac{1}{a} \quad \Rightarrow \quad \tau^* = \sqrt{m\,a}\,\sigma - 1\,.$$

As $\tau \mapsto \mathsf{E}[u(r,\tau)]$ is strictly convex it follows that τ^* is the optimal effort. □

Proof of Theorem 1. We will prove the result in the more general setting with subjective self-confidence and scale use as introduced in Sections 2.3.2 and 2.3.3. The case without subjective self-confidence and scale use stated in the main text corresponds to special case where $\tilde{\tau}_i = \tau_i$ and $\phi_i = 1$.

The answer of agent i when asked for the t^{th} time about the k^{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 agents responses yields that

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

We thus have that

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

Similarly, we get that

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

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

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

We have that $A_k = \sum_{t=1}^{T}\left(\epsilon_{ikt} - \bar{\epsilon}_{ik}\right)^2$ is χ^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_j are independent for $j \neq k$ and $A = \sum_{k=1}^{K} A_k$ it follows that A is χ^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 χ^2 distributed with $K-1$ degrees of freedom. It follows from (13) that

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

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 $\lambda_{ik}, \lambda_{ij}$ are independent for $k \neq j$, it follows that B is χ^2 distributed with $K-1$ degrees of freedom.

Next, recall that for the Normal distribution, the sample variance $\frac{1}{T-1}\sum_{t=1}^{T}\left(\epsilon_{ikt}-\bar{\epsilon}_{ik}\right)^2$ is independent of the sample mean $\bar{\epsilon}_{ik}$. As η is independent of ϵ it follows that $\sum_{t=1}^{T}\left(\epsilon_{ikt}-\bar{\epsilon}_{ik}\right)^2$ 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 χ^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 $K(T-1)$.[19] Recall that in Equation 8 we defined $\hat{\tau}_i$

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

Plugging in the definition of A and B yields that

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

This establishes (9). Part 2 of the Theorem follows as $\mathsf{E}\left[F_i\right] = \frac{K(T-1)}{K(T-1)-2}$.[20] Part 3 follows as

$$\mathrm{var}(F_i) = \mathsf{E}\left[F_i\right]^2 \frac{2((K-1)+K(T-1)-2)}{(K-1)(K(T-1)-4)} \,.$$

To prove Part 4 observe that (10) is decreasing in T and thus an upper bound is given by setting $T=2$

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

[19] See https://en.wikipedia.org/wiki/F-distribution#Characterization.

[20] See https://en.wikipedia.org/wiki/F-distribution.

This establishes the result. Finally, we note that this result immediately extends to the case of endogenous e□ort introduced in Section 2.3.1, where for agent-specific ability a_i and incentives m_i, the precision is endogenously chosen as $\tau_i = \sqrt{m_i\, a_i}\, \sigma - 1$. □

Appendix B Empirical Details

B.1 Di□erent Average Characteristics

The estimator introduced in Section 3 assumes that the population means 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 derive in Section 3 to the case where the population mean $\bar{\theta}_k$ of each characteristic is k potentially di□erent. We make no assumption about the distribution of these population means, but maintain the assumption that the agent's prior belief equals the distribution of characteristics in the population. Throughout, we maintain the assumption of no scale use $\phi_i = 1$. The response of agent i when asked for the t^{th} time about characteristic k is then given by

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

We note that the mean answer $\bar{r}_{ik}$ of each agent i when asked about characteristic k unconditional on his type is $\bar{\theta}_k$. Thus, by the strong law-of-large-numbers, the mean answer $\bar{r}_k$ in the population converges to $\bar{\theta}_k$ when the population gets large. By assumption, there exists independent, standard normally distributed random variables $\epsilon_{ikt}, \eta_{ik}$ such that

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

We first note that $r_{ikt} - \bar{r}_{ik}$ is the same as in the case with equal means (see Equation 12) as the mean of the characteristic cancels out if one subtracts an agent's mean response

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

Furthermore, we have that

$$\begin{aligned} \bar{r}_{ik} &= \bar{\theta}_k + \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i}(\bar{x}_{ik} - \bar{\theta}_k) = \bar{\theta}_k + \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i}(\bar{\theta}_k + \sigma\eta_{ik} + \frac{\sigma}{\sqrt{\tau_i}}\bar{\epsilon}_{ik} - \bar{\theta}_k) \\ &= \bar{\theta}_k + \frac{\tilde{\tau}_i}{1 + \tilde{\tau}_i}(\sigma\eta_{ik} + \frac{\sigma}{\sqrt{\tau_i}}\bar{\epsilon}_{ik})\,. \end{aligned}$$

As $\lim_{N\to\infty}\bar{r}_k = \bar{\theta}_k$ we have that

$$\lim_{N\to\infty}(\bar{r}_{ik}-\bar{r}_k) = \frac{\tilde{\tau}_i}{1+\tilde{\tau}_i}(\sigma\eta_{ik} + \frac{\sigma}{\sqrt{\tau_i}}\bar{\epsilon}_{ik}).$$

This yields that

$$\lim_{N\to\infty}\left[(\bar{r}_{ik}-\bar{r}_k) - \frac{1}{K}\sum_{k'=1}^{K}(\bar{r}_{ik'}-\bar{r}_{k'})\right] = \frac{\tilde{\tau}_i}{1+\tilde{\tau}_i}(\sigma(\eta_{ik}-\bar{\eta}_i) + \frac{\sigma}{\sqrt{\tau_i}}(\bar{\epsilon}_{ik}-\bar{\epsilon}_i)), \tag{14}$$

and thus is the same as in (13) when the means of different characteristics are equal. Consequently, $r_{ikt}-\bar{r}_{ik}$ has the same distribution as in the case with equal means and the expression in (14) converges to a random variable that has exactly the same distribution as $\bar{r}_{ik}-\bar{r}_i$ in the case where the means are identically. We define the population-based estimator

$$\hat{\tau}_i^{POP} = \frac{\frac{1}{K-1}\sum_{k=1}^{K}\left(\bar{r}_{ik}-\bar{r}_k-\frac{1}{K}\sum_{k'=1}^{K}(\bar{r}_{ik'}-\bar{r}_{k'})\right)^2}{\frac{1}{K(T-1)-2}\sum_{k=1}^{K}\sum_{t=1}^{T}(r_{ikt}-\bar{r}_{ik})^2} - \frac{1}{T} \tag{15}$$

Applying, Theorem 1 then yields the following result:

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

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

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

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 τ_i^{POP}, i.e., $\lim_{N\to\infty}\mathsf{E}[\hat{\tau}_i \mid \tau_i] = \tau_i$.*

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

$$\lim_{N\to\infty}\sqrt{\mathsf{E}[(\hat{\tau}_i^{POP}-\tau_i)^2 \mid \tau_i]} = \left(\tau_i+\frac{1}{T}\right)\sqrt{\frac{2((K-1)+K(T-1)-2)}{(K-1)(K(T-1)-4)}}. \tag{17}$$

4. *$\hat{\tau}_i^{POP}$ converges to τ_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*

$$\lim_{N\to\infty}\sqrt{\mathsf{E}[(\hat{\tau}_i^{POP}-\tau_i)^2 \mid \tau_i]} \le \frac{2\tau_i+1}{\sqrt{K-4}}.$$

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 N 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 to the actual behavior of the population-based estimator in many relevant contexts, as we illustrate with the simulation results below.

Table 7: Accuracy of estimates with different means

	(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	0.669	0.667	0.870	0.760	0.910
Rank corr.	0.746	0.758	0.899	0.820	0.928
Median split	78.7%	79.3%	88.0%	82.6%	90.3%
Bias	-0.029	0.005	-0.028	-0.034	-0.032

The table replicates Table 1, aside from that the means of the characteristics, $\bar{\theta}$, are independently drawn from a Normal distribution with mean 5 and standard deviation 1. A comparison of the result shows that the performance is almost identical to the case with equal means. This even holds true for the cases where the simulated number of respondents is just 100, a number 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 pertaining to 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. The same is our interest here, 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.

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 in assigning respondents to

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	0.632	0.626	0.717
Rank corr.	0.721	0.729	0.796
Median split	77.4%	77.9%	81.3%
Bias	-0.193	-0.190	-0.187

having a level of self-knowledge that lies above or below the population average, the fraction of correctly classified respondents decreases by less than two percentage points, i.e., the informativeness of the median-splits remains.

B.3 OLS Coefficients

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 τ in the light of our model. Table 9 provides a schematic overview of the effects that the 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

Table 9: Effect of reduction in self-knowledge τ on OLS estimates

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

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 τ 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 2.3.2).

B.3.1 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 ϵ_i is and i.i.d. error terms with an expected value of zero that is independent of y_i and the signals that subjects receive. Crucially, the value θ_i is not observable and instead replaced with the response r_i. To gain a deeper insight into the forces behind the composite e□ect, we use the notation involving subjective self-knowledge (see Section 2.3.2. The asymptotic result of the standard OLS estimator is derived below.

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

$$\hat{\beta}_0 = \bar{\theta} - \hat{\beta}_1\,\bar{y} = \beta_0 + \beta_1\,\bar{y} - \frac{\tilde{\tau}}{1+\tilde{\tau}}\,\beta_1\,\bar{y} = \beta_0 + \left(1 - \frac{\tilde{\tau}}{1+\tilde{\tau}}\right)\beta_1\,\bar{y}$$

Thus, as long as a decrease in τ is accompanied by a decrease in $\tilde{\tau}$, the overall e□ect on the absolute value of the slope parameter β_1 is strictly negative.

B.3.2 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 coe□-cient is then given by $(z_2 - z_1)\,/\,(r_2 - r_1)$. Reduced subjective self-knowledge attenuates the absolute di□erence between r_1 and r_2, thereby increasing the estimate. Formally, assume that we want to estimate the unknown coe□cients 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 η_i an i.i.d. error terms with an expected value of zero that is independent of both θ_i and the signals that subjects receive. Again, the unknown true values θ_i are replace with reports r_i, and the asymptotic result of the standard OLS estimator is derived below.

$$\hat{\gamma}_1 = \frac{\mathsf{cov}(z_i, r_i)}{\mathsf{var}(r_i)} = \frac{\mathsf{E}[(z_i - \bar{z})(r_i - \bar{r})]}{\mathsf{E}\left[(r_i - \bar{r})^2\right]} = \frac{\mathsf{E}\left[\left(\gamma_0 + \gamma_1 \theta_i + \epsilon_i - \gamma_0 - \gamma_1 \bar{\theta}\right)\left(\frac{\bar{\theta}+\tilde{\tau}\, x_i}{1+\tilde{\tau}} - \bar{\theta}\right)\right]}{\mathsf{E}\left[\left(\frac{\bar{\theta}+\tilde{\tau}\, x_i}{1+\tilde{\tau}} - \bar{\theta}\right)^2\right]}$$

$$= \frac{\mathsf{E}\left[\left(\gamma_1\left(\theta_i - \bar{\theta}\right) + \epsilon_i\right)\left(\frac{\tilde{\tau}}{1+\tilde{\tau}}\left(x_i - \theta_i + \theta_i - \bar{\theta}\right)\right)\right]}{\left(\frac{\tilde{\tau}}{1+\tilde{\tau}}\right)^2 \mathsf{E}\left[\left(x_i - \bar{\theta}\right)^2\right]} = \frac{\gamma_1 \frac{\tilde{\tau}}{1+\tilde{\tau}} \mathsf{var}(\theta)}{\left(\frac{\tilde{\tau}}{1+\tilde{\tau}}\right)^2 [\mathsf{var}(\theta) + \mathsf{var}(x \mid \theta)]}$$

$$= \frac{\gamma_1 \sigma^2}{\frac{\tilde{\tau}}{1+\tilde{\tau}}\left(\sigma^2 + \frac{\sigma^2}{\tau}\right)} = \frac{1+\tilde{\tau}}{\tilde{\tau}} \frac{\tau}{1+\tau} \gamma_1$$

$$\hat{\gamma}_0 = \bar{z} - \hat{\gamma}_1 \bar{\theta} = \gamma_0 + \gamma_1 \bar{\theta} - \hat{\gamma}_1 \bar{\theta} = \gamma_0 + \left(1 - \frac{1+\tilde{\tau}}{\tilde{\tau}} \frac{\tau}{1+\tau}\right) \gamma_1 \bar{\theta}$$

The overall e ect of a reduction in τ 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 e ects cancel out exactly. If a reduction of τ 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 e ects from 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 C Robustness tests

C.1 Results for non-standardized self-reports

Table 10: Predictive power of non-standardized domain-specific attitudes towards risk

	Without controls			Including controls		
Sample:	*pooled*	*below*	*above*	*pooled*	*below*	*above*
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: *Investing in stocks*						
Risk attitude	0.0519***	0.0398***	0.0627***	0.0431***	0.0324***	0.0533***
	(0.00183)	(0.00254)	(0.00256)	(0.00200)	(0.00274)	(0.00286)
(Partial) R^2	0.0566	0.0361	0.0774	0.0396	0.0238	0.0574
Observations	15202	7601	7601	12081	6041	6040
Dependent variable: *Performance pay*						
Risk attitude	0.0133***	0.00715**	0.0190***	0.0105***	0.00579*	0.0152***
	(0.00178)	(0.00224)	(0.00277)	(0.00199)	(0.00250)	(0.00310)
(Partial) R^2	0.00918	0.00326	0.0158	0.00559	0.00202	0.0102
Observations	5612	2806	2806	4509	2255	2254
Dependent variable: *Smoking*						
Risk attitude	0.0198***	0.0178***	0.0226***	0.0129***	0.00872***	0.0172***
	(0.00154)	(0.00220)	(0.00216)	(0.00171)	(0.00243)	(0.00241)
(Partial) R^2	0.0119	0.00924	0.0161	0.00517	0.00231	0.00924
Observations	15146	7573	7573	12122	6061	6061

Note: Table reports OLS estimates. 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 investing in stocks, *career* for performance pay, and *health* for smoking. The controls used in Columns 4–6 are gender, age, body height, parental education, log wealth, log debts, and log gross household income. All data are from 2009, except for information on except for information on body height and smoking, which due to a lack of availability, are from 2010. The model for performance pay includes only respondents up to the age of 66 who are working full-time and that are not self-employed. Tests for the significance of differences in R^2 are each based on 10,000 simulations randomly assigning relevant subjects to counterfactual groups of corresponding sizes. Heteroskedasticity-robust standard errors in parentheses. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$.

Table 11: Differences in non-standardized risk attitudes

	Gender			Height		
Sample:	*pooled* (1)	*below* (2)	*above* (3)	*pooled* (4)	*below* (5)	*above* (6)
			Risk domain: *general*			
Coefficient	-0.817***	-0.756***	-0.878***	0.0479***	0.0441***	0.0510***
	(0.0334)	(0.0476)	(0.0469)	(0.00189)	(0.00264)	(0.00270)
R^2	0.0348	0.0295	0.0406	0.0425	0.0346	0.0502
Observations	16633	8317	8316	15103	7552	7551
			Risk domain: *car driving*			
Coefficient	-1.029***	-0.962***	-1.097***	0.0632***	0.0597***	0.0657***
	(0.0399)	(0.0568)	(0.0559)	(0.00228)	(0.00322)	(0.00323)
R^2	0.0407	0.0352	0.0468	0.0534	0.0456	0.0602
Observations	15742	7871	7871	14314	7157	7157
			Risk domain: *financial matters*			
Coefficient	-0.798***	-0.705***	-0.891***	0.0401***	0.0343***	0.0449***
	(0.0328)	(0.0465)	(0.0461)	(0.00189)	(0.00264)	(0.00272)
R^2	0.0352	0.0275	0.0440	0.0316	0.0225	0.0407
Observations	16523	8262	8261	15011	7506	7505
			Risk domain: *sports/leisure*			
Coefficient	-0.834***	-0.771***	-0.896***	0.0594***	0.0530***	0.0638***
	(0.0401)	(0.0567)	(0.0563)	(0.00225)	(0.00319)	(0.00319)
R^2	0.0261	0.0224	0.0303	0.0470	0.0361	0.0563
Observations	16344	8172	8172	14848	7424	7424
			Risk domain: *career*			
Coefficient	-0.736***	-0.695***	-0.772***	0.0576***	0.0552***	0.0580***
	(0.0439)	(0.0625)	(0.0614)	(0.00244)	(0.00341)	(0.00348)
R^2	0.0189	0.0167	0.0212	0.0410	0.0363	0.0434
Observations	14618	7309	7309	13251	6626	6625
			Risk domain: *health*			
Coefficient	-0.750***	-0.613***	-0.887***	0.0417***	0.0349***	0.0471***
	(0.0373)	(0.0527)	(0.0525)	(0.00210)	(0.00296)	(0.00297)
R^2	0.0240	0.0161	0.0335	0.0263	0.0179	0.0345
Observations	16625	8313	8312	15096	7548	7548

Note: Table reports OLS estimates. Tests for the significance of differences in R^2 are each based on 10,000 simulations randomly assigning relevant subjects to counter-factual groups of corresponding sizes. Heteroskedasticity-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

C.2 Accounting for Selection

The below table replicates Table 6 but, in columns that use split samples (2, 3, 5, and 6), inverse probability weighting is used to account for selection on observable characteristics. The underlying probabilities are prediction from Probit models regressing above vs. below values of $\hat{\tau}$ on gender, age, squared age, and years of education (all as of 2011). For observations where any necessary information are missing, the probability is set to 0.5.

Table 12: Differences in risk attitudes, with inverse probability weighting

	Gender			Height		
Sample:	*pooled* (1)	*below* (2)	*above* (3)	*pooled* (4)	*below* (5)	*above* (6)
			Risk domain: *general*			
Coefficient	-0.374***	-0.345***	-0.409***	0.0220***	0.0195***	0.0243***
	(0.0153)	(0.0219)	(0.0222)	(0.000866)	(0.00122)	(0.00128)
R^2	0.0348	0.0298	0.0411	0.0425	0.0326	0.0532
Observations	16633	8317	8316	15103	7552	7551
			Risk domain: *car driving*			
Coefficient	-0.403***	-0.366***	-0.446***	0.0248***	0.0224***	0.0269***
	(0.0156)	(0.0226)	(0.0226)	(0.000893)	(0.00129)	(0.00131)
R^2	0.0407	0.0335	0.0491	0.0534	0.0423	0.0645
Observations	15742	7871	7871	14314	7157	7157
			Risk domain: *financial matters*			
Coefficient	-0.376***	-0.342***	-0.415***	0.0190***	0.0167***	0.0213***
	(0.0154)	(0.0226)	(0.0218)	(0.000896)	(0.00130)	(0.00129)
R^2	0.0352	0.0288	0.0435	0.0316	0.0233	0.0420
Observations	16523	8262	8261	15011	7506	7505
			Risk domain: *sports/leisure*			
Coefficient	-0.323***	-0.294***	-0.356***	0.0231***	0.0198***	0.0259***
	(0.0155)	(0.0223)	(0.0225)	(0.000875)	(0.00126)	(0.00127)
R^2	0.0261	0.0216	0.0313	0.0470	0.0334	0.0607
Observations	16344	8172	8172	14848	7424	7424
			Risk domain: *career*			
Coefficient	-0.275***	-0.259***	-0.301***	0.0215***	0.0197***	0.0235***
	(0.0164)	(0.0236)	(0.0239)	(0.000912)	(0.00130)	(0.00134)
R^2	0.0189	0.0167	0.0223	0.0410	0.0331	0.0498
Observations	14618	7309	7309	13251	6626	6625
			Risk domain: *health*			
Coefficient	-0.310***	-0.256***	-0.375***	0.0173***	0.0140***	0.0203***
	(0.0154)	(0.0221)	(0.0222)	(0.000868)	(0.00125)	(0.00125)
R^2	0.0240	0.0164	0.0347	0.0263	0.0167	0.0376
Observations	16625	8313	8312	15096	7548	7548

Note: Table reports OLS estimates. Heteroskedasticity-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

C.3 Effects on Big Five

Table 13 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 13: Differences in Big Five

	Gender			Height		
Sample:	*pooled* (1)	*below* (2)	*above* (3)	*pooled* (4)	*below* (5)	*above* (6)
			Domain: *Agreeableness*			
Coefficient	0.344***	0.290***	0.394***	-0.0172***	-0.0132***	-0.0207***
	(0.0155)	(0.0220)	(0.0218)	(0.000863)	(0.00127)	(0.00117)
R^2	0.0296	0.0210	0.0388	0.0261	0.0149	0.0388
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Conscientiousness*			
Coefficient	0.142***	0.0585**	0.219***	-0.00873***	-0.00417**	-0.0124***
	(0.0157)	(0.0222)	(0.0221)	(0.000885)	(0.00128)	(0.00123)
R^2	0.00504	0.000854	0.0119	0.00673	0.00149	0.0140
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Extraversion*			
Coefficient	0.197***	0.156***	0.231***	-0.00245**	-0.000182	-0.00405***
	(0.0156)	(0.0221)	(0.0220)	(0.000878)	(0.00127)	(0.00122)
R^2	0.00964	0.00610	0.0133	0.000532	0.00000284	0.00149
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Neuroticism*			
Coefficient	0.434***	0.364***	0.496***	-0.0204***	-0.0169***	-0.0233***
	(0.0153)	(0.0218)	(0.0214)	(0.000863)	(0.00126)	(0.00119)
R^2	0.0470	0.0330	0.0614	0.0365	0.0245	0.0493
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Openness*			
Coefficient	0.128***	0.124***	0.134***	0.00118	0.00253*	-0.000510
	(0.0156)	(0.0221)	(0.0220)	(0.000878)	(0.00124)	(0.00124)
R^2	0.00411	0.00381	0.00448	0.000124	0.000548	0.0000236
Observations	16338	8169	8169	14815	7408	7407

Note: Table reports OLS estimates. Tests for the significance of differences in R^2 are each based on 10,000 simulations randomly assigning relevant subjects to counterfactual groups of corresponding sizes. Heteroskedasticity-robust standard errors in parentheses. $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$.

Table 14 replicates Table 13 but, in columns that use split samples (2, 3, 5, and 6), inverse probability weighting is used to account for selection on observable characteristics. The underlying probabilities are prediction from Probit models regressing above vs. below values of $\hat{\tau}$ on gender, age, squared age, and years of education (all as of 2011). For observations where any necessary information are missing, the probability is set to 0.5.

Table 14: Differences in Big Five, with inverse probability weighting

	Gender			Height		
Sample:	*pooled* (1)	*below* (2)	*above* (3)	*pooled* (4)	*below* (5)	*above* (6)
			Domain: *Agreeableness*			
Coefficient	0.344***	0.286***	0.414***	-0.0172***	-0.0129***	-0.0219***
	(0.0155)	(0.0221)	(0.0224)	(0.000863)	(0.00128)	(0.00121)
R^2	0.0296	0.0206	0.0419	0.0261	0.0143	0.0433
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Conscientiousness*			
Coefficient	0.142***	0.0576**	0.222***	-0.00873***	-0.00422***	-0.0123***
	(0.0157)	(0.0222)	(0.0226)	(0.000885)	(0.00128)	(0.00127)
R^2	0.00504	0.000841	0.0122	0.00673	0.00154	0.0138
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Extraversion*			
Coefficient	0.197***	0.159***	0.215***	-0.00245**	-0.0000615	-0.00315*
	(0.0156)	(0.0224)	(0.0223)	(0.000878)	(0.00129)	(0.00124)
R^2	0.00964	0.00630	0.0116	0.000532	0.000000324	0.000909
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Neuroticism*			
Coefficient	0.434***	0.366***	0.502***	-0.0204***	-0.0171***	-0.0235***
	(0.0153)	(0.0220)	(0.0219)	(0.000863)	(0.00129)	(0.00121)
R^2	0.0470	0.0336	0.0622	0.0365	0.0249	0.0500
Observations	16338	8169	8169	14815	7408	7407
			Domain: *Openness*			
Coefficient	0.128***	0.113***	0.126***	0.00118	0.00247*	0.000654
	(0.0156)	(0.0223)	(0.0226)	(0.000878)	(0.00126)	(0.00127)
R^2	0.00411	0.00321	0.00390	0.000124	0.000526	0.0000388
Observations	16338	8169	8169	14815	7408	7407

Note: Table reports OLS estimates. Heteroskedasticity-robust standard errors in parentheses. $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$.