American Economic Review 2026, 116(1): 88–118
https://doi.org/10.1257/aer.20211217

Risk Preferences and Field Behavior: The Relevance of Higher-Order Risk Preferences[†]

By Sebastian O. Schneider and Matthias Sutter*

Using new methods, we measure the intensities of higher-order risk preferences (prudence and temperance) in an incentivized experiment with 658 adolescents. Aligned with theory, we find that higher-order risk preferences are strongly related to field behavior, including prevention, health, addictive behavior, and financial decision-making. Most importantly, we show that ignoring prudence and temperance can yield misleading conclusions about the relation of risk preferences to field behavior, and that survey measures of risk tolerance often relate to field behavior because they capture higher-order risk preferences. (*JEL* C83, D81, D91, J13)

Risk is an inherent part of life: Decisions about occupation, education, finances, or health behavior, to name just a few, regularly involve at least some degree of risk. Consequently, measuring risk preferences is important for a better understanding of behavior in various domains of life. While a plethora of research has empirically documented a connection between commonly used measures of risk aversion and health outcomes, educational and financial success, or occupational choices (e.g., Anderson and Mellor 2008; Dohmen et al. 2011; Becker et al. 2012), there is also plenty of work that fails to predict field behavior with subjects' risk aversion (e.g., Sutter et al. 2013; Galizzi, Machado, and Miniaci 2016; Charness et al. 2020; Samek et al. 2021), or reports rather weak relations between experimental measures and standard survey questions on risk tolerance (e.g., Frey et al. 2017, or Crosetto 2025).

*Schneider: Max Planck Institute for Behavioral Economics (email: sschneider@coll.mpg.de); Sutter: Max Planck Institute for Behavioral Economics; University of Cologne and University of Innsbruck (email: matthias.sutter@coll.mpg.de). Pietro Ortoleva was the coeditor for this article. For helpful comments we thank Zvonimir Bašić, Laura Breitkopf, Gary Charness, Stefano DellaVigna, Markus Dertwinkel-Kalt, Sebastian Ebert, Helga Fehr-Duda, Dalila Figueiredo, Matteo Galizzi, Glenn Harrison, James Heckman, Matthias Heinz, Alex Imas, Charles Noussair, Manuela Puente-Beccar, Bettina Rockenbach, Theresa Schaetze, Hannah Schildberg-Hörisch, Stefan Schmidt, Sonja Settele, Jana Tissen, Sebastian Tonke, Stefan Trautmann, Peter Wakker, Daniel Wiesen, Johannes Wohlfart, the editor Pietro Ortoleva, and three referees, plus participants at IMEBESS, M-BEPS, TIBER, ESA, NCBEE, BEH-net, German Economic Association, CESifo, iHEA, EEA-ESEM, CEHD Chicago's and Jinan University's Frontiers of Measurement and Formation of Skills, Boston University, George Mason University, Goethe-University Frankfurt, University of Mainz, University of Amsterdam, UC Santa Barbara, and University of Chicago. We also thank Christa Kuon-Rehm, Karl Müller, Christian Neubauer, Eva Schneider, and Klaus Schneiderhan for their support and coordination of the study in their respective schools and all our assistants. This paper replaces our former working paper "Higher Order Risk Preferences: New Experimental Measures, Determinants and Field Behavior." This study was approved by the IRB of the University of Innsbruck (Certificate of good standing 27/2018) and it was preregistered with the Open Science Foundation (osf.io/n7v2y). Financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC 2126/1-390838866 is gratefully acknowledged.

†Go to https://doi.org/10.1257/aer.20211217 to visit the article page for additional materials and author disclosure statement(s).

The somewhat mixed results on the relationship between risk aversion and field behavior may be related to measurement error, causing imperfect correlations (see Gillen, Snowberg, and Yariv 2019, for a recent illustration). However, another potential explanation beyond measurement error is that risk comes in different forms. In fact, starting with the case of precautionary saving (Leland 1968), a growing body of theoretical contributions suggests that for explaining a broad range of field behavior, *higher-order* risk preferences (in particular *prudence* and *temperance*) might be relevant, as the *second-order* risk preference *risk aversion* cannot unambiguously explain behavior in these contexts (Kimball 1990, 1993; Gollier and Pratt 1996; Eeckhoud and Gollier 2005). Focusing on the "wrong aspect" of risk preferences and limiting the consideration to risk aversion only (which applies to all "null-results" mentioned above) might therefore explain, at least partially, the often weak and mixed empirical relation between risk preferences and field behavior. Particularly so, because the higher-order risk preferences prudence and temperance not only play a crucial role in explaining intertemporal behavior under uncertainty, they may also capture attitudes towards important aspects of the distribution of a risk beyond its mean and variance: left-skewness and kurtosis aversion, respectively (Modica and Scarsini 2005; Denuit and Eeckhoudt 2010; Ebert 2013). Imagine you should allocate a mean-zero risk, and you can decide to allocate it to a state of higher wealth or to a state of lower wealth. Following Eeckhoudt and Schlesinger (2006), preferring to take risk in the good state of the world classifies as prudent, which in turn implies decreasing risk aversion. Similarly, temperance can be interpreted as the preference to allocate two independent mean-zero risks across two states of the world, as opposed to accepting both of them in the same state of the world (Eeckhoudt and Schlesinger 2006). This means that temperate choices avoid accumulating risks, but rather spread them out over different states of the world.

In this paper, we present and use novel ways to measure risk aversion and the higher-order risk preferences prudence and temperance, relate them to field behavior, and contribute to clarifying the often puzzling picture emerging from relating risk aversion to field behavior. We build on the empirically validated method developed by Schneider, Ibanez, and Riener (2025) for eliciting utility-based measures of higher-order risk preferences. Just as risk aversion equals a negative second derivative under expected utility theory, prudence and temperance can be defined as a positive third and negative fourth derivative of the utility function. Here, we develop a sophisticated elicitation procedure for this method as well as an easy-to-understand implementation that is additionally robust to well-known anomalies in decision-making. The method accounts for defective measurement caused by, for example, decision and approximation errors in the elicitation procedure, and is the first to yield nonparametric utility-based intensity measures in the spirit of the Arrow-Pratt measure for risk aversion. In addition, we develop a simple, raw measurement to approximate utility-based prudence intensity. As utility-based intensity measures are the measures applied in related theoretical work (discussed below) and as common parametric approaches imply a dependence structure between the orders of risk preferences that is not matched by empirical data (Ebert and Wiesen 2014), our nonparametric methods provide the ideal basis to relate intensity measures of higher-order risk preferences to field behavior.

We examine the distribution and the determinants of higher-order risk preferences (and of risk aversion) in a sample of 658 children and adolescents, aged 10 to 21 years. First, we find, in the aggregate, children and adolescents are risk averse, prudent, and temperate, which is in line with findings from adults (Trautmann and van de Kuilen 2018). Moreover, we extend the standard finding of gender differences in risk aversion (e.g., Croson and Gneezy 2009) to apply to higher orders of risk as well: Females are not only more risk averse, but also more prudent and more temperate than males (as reported for adults by Ebert and Wiesen 2014). Age has no effect on the intensity of any (higher-order) risk preference.

Second, we relate our individual experimental measures of risk aversion, prudence and temperance to self-reported behavior in the field. We focus on (to the best of our knowledge) all domains of everyday life for which theoretical models have linked the intensities of higher-order risk preferences to behavior. This captures health-related behavior (e.g., Courbage and Rey 2006), eco-friendly behavior (e.g., Bramoullé and Treich 2009), prevention effort to lower the probability of an undesired event (Eeckhoudt and Gollier 2005; Menegatti 2009), and household finance (e.g., Kimball 1990). Except for the area last mentioned (Noussair, Trautmann, and van de Kuilen 2014), there are no empirical studies on the relationship between higher-order risk preferences and field behavior in these domains (Trautmann and van de Kuilen 2018). We find that prudence and, to a lesser extent, temperance are very important for predicting field behavior, as suggested by theory. To start, we can replicate in our sample of children and adolescents the seminal result of Noussair, Trautmann, and van de Kuilen (2014) that financial decision-making is related to prudence and temperance. Drawing on a much larger set of field behavior, we uncover many novel insights beyond the domain of financial decision-making. Most importantly, prudence is strongly related to health-related behavior, but risk aversion is not. This holds particularly true for addictive behavior like the obsessive use of smartphones, but also extends to including smoking and drinking, or being overweight. Prudence also matters for other domains of field behavior, such as prevention behavior or the preference for a competitive income. Temperance is related to investment, prevention behavior and cautious planning. Our experimental measure of risk aversion, however, is only related to prevention and eco-friendly behavior, and to a preference for competitive income, while it is unrelated to all other domains. The latter could be mistaken as meaning that risk preferences play no role for other, important fields of behavior, such as health-related behavior and financial decision-making. Yet, they do, as our findings for higher-order risk preferences reveal. This suggests that higher-order risk preferences should be taken into account when studying the relation of risk preferences to field behavior to avoid misleading conclusions. Reassuringly, we can rule out a series of alternative explanations for our findings, including biased measures, measurement error and noise, probability weighting, reference dependence, and random decision-making.

Third, we examine how frequently used survey measures (e.g., Dohmen et al. 2011) relate to our set of field behavior. Usually, such survey measures are interpreted as capturing risk aversion, and they are more often successful predictors for field behavior compared to experimental measures of risk aversion, as our extensive literature review shows. This is also true in our sample of children and adolescents. However, we are the first to show that survey measures are related to

field behavior because they are correlated with higher-order risk preferences up to the same degree as risk aversion is. This insight, together with the previous results, can help understand why survey measures of risk aversion are often related to field behavior, while experimental measures of risk aversion are not.

Our paper contributes to the literature in several ways. The first contribution is to provide a unified experimental framework to measure higher-order risk preferences (and risk aversion) for a large sample of children and adolescents, using novel methods to elicit nonparametric utility-based intensity measures. To date, the standard approach to obtain intensity measures of higher-order risk preferences has been to count the number of decisions in the binary, "behavioral" classification tasks due to Eeckhoudt and Schlesinger (2006) that indicate prudence or temperance. For the lack of a theoretical relation to any proper intensity measure or an empirical justification (for a discussion, see, e.g., Noussair, Trautmann, and van de Kuilen 2014), the interpretation of this "count measure" as an intensity measure is, however, questionable. Moreover, as a discrete measure with limited support and variation, its suitability to empirically relate to field behavior is restricted. Our measures overcome all these imperfections, while our developed implementation of the method, i.e., our elicitation tasks are arguably easier to understand for participants than the compound-lottery tasks based on Eeckhoudt and Schlesinger (2006), which makes the measurement even faster. Even though this helps avoiding decision errors to begin with, our sophisticated method additionally accounts for errors in measurement, which is particularly important for prediction purposes (Gillen, Snowberg, and Yariv 2019). Consequently, our sophisticated measures demonstrate superior predictive power and greater robustness to noise and decision errors when compared to our raw measures. Moreover they entail a measure for temperance. Our novel, raw measure of (utility-based) prudence, in turn, is plain and intuitive. It is also easily implemented and operationalized (and still has a clear theoretical interpretation as intensity measure and a wider support than the "count measure"). Next, while risk aversion has been studied extensively, including this age group (see Sutter, Zoller, and Glätzle-Rützler 2019, for a review), for prudence the evidence in this group is limited to Heinrich and Shachat (2020). They analyze the number of prudent decisions as proxy for prudence intensity from a sample of 362 Chinese children and adolescents. Unlike us, Heinrich and Shachat (2020) do not compare theoretically predicted relations of prudence to field behavior and disregard temperance altogether, which has, so far, never been studied with children or adolescents.

Our second contribution is that we are the first to connect higher-order risk preferences to a much wider range of field behavior than before, and the first to study their relation to adolescents' behavior. So far, empirical work is confined to the seminal study of Noussair, Trautmann, and van de Kuilen (2014) about how higher-order risk preferences of Dutch adults relate to their financial decisions. We also consider the financial behavior of our adolescents, confirming the findings of Noussair, Trautmann, and van de Kuilen (2014) also for our younger age group. Yet, our focus lies on providing a comprehensive approach, including, in particular, the theoretically related domains of field behavior that have not been linked to higher-order risk preferences empirically, some of which might be considered most formative in this age group. For example, smoking or drinking during the teenage years has high predictive power for also showing such behavior in adulthood (DeWit et al. 2000;

Paul et al. 2008). Thus, a key domain of our study is health-related behavior, where we focus on excessive smartphone usage, which has been linked to mental health issues, like depression, and poor well-being of adolescents (Przybylski and Weinstein 2017; Orben and Przybylski 2019). We successfully relate prudence to health-related behavior and financial decision-making. We also link (higher-order) risk preferences empirically to adolescents' eco-friendly behavior and their preventive behavior to avoid undesired events.

Our third, and arguably most important, contribution relates back to the often weak or inconclusive empirical results on the relation of risk-aversion measures to field behavior, and is unprecedented in the literature. On the one hand, we illustrate with experimental measures how weak or even null results regarding the relationship of risk preferences to field behavior may be due to omitting prudence and temperance—even above and beyond issues of measurement error. For example, we show how focusing on risk aversion and neglecting higher-order risk preferences results in wrongly concluding that health-related behavior is unrelated with risk—whereas a measure of prudence would clearly reveal such a relationship. On the other hand, the commonly used one-item survey measure on risk tolerance (e.g., Falk et al. 2018) often successfully relates to field behavior, even in cases where experimental measures of risk aversion do not. We can replicate this pattern, and show that in our data this is the case because survey measures capture prudence and temperance and account for them when relating to field behavior, whereas in the literature they are usually assumed to be measures of risk aversion (only). These insights provide a novel view on previous findings and demonstrate the importance of accounting for higher-order risk preferences in empirical analyses of theoretically related behavior.

I. Experimental Design

A. *General Setup*

Subject Pool.—We ran the experiment in four German schools in fall 2018 during regular school hours. In all schools, classes were selected randomly such that at least one class per grade from grades six, eight, ten, and twelve participated in the study. In total, 658 children and adolescents took part voluntarily and with their parents' informed consent.[1] 99 percent of our participants were between 10 and 18.5 years old (see Supplemental Appendix Table D-1 for details).

General Experimental Setup.—The study was run on tablet computers using oTree (Chen, Schonger, and Wickens 2016). Instructions were the same in every session.[2] As we describe in the following subsection, we, first, elicited risk and time preferences, then measured cognitive abilities and, finally, administered an extensive survey on field behavior; see Schneider and Sutter (2026) for the data collected.

In the experiment, adolescents could earn up to 280 "taler." We explained the conversion rate from taler to euro carefully and varied it depending on age, such that the maximal earnings corresponded to 120 percent of the weekly amount of

[1] Less than 5 percent of parents in participating classes did not provide informed consent.
[2] Instructions are available in Supplemental Appendix F.

pocket money recommended by the German Federal Ministry of Family Affairs for the different age cohorts (see the website Familien-wegweiser.de 2018). This was done in order to hold the relative value of a taler constant across the cohorts. In total, adolescents made 18 decisions in our risk-preference and four decisions in our time-preference tasks. In addition, we included three tasks to validate our measures of prudence. Among all decision tasks, one was randomly selected for payment by the computer. If one of the risk tasks was selected, and the participant chose an option with a random element, chance was resolved by the computer. The payoffs in these tasks ranged from 0 to 140 taler. If one of the time-preference tasks was selected for payment and the later payment was chosen, it was handed over to the student in an anonymized way at the specified date. The payoffs in these tasks ranged from 100 to 140 taler. Payments (in cash) also included a show-up fee of 70 taler and up to 70 taler for the cognitive-ability tasks.

B. *Risk and Time Tasks*, *Risk Survey Items*, *and Cognitive Ability Tasks*

Risk.—Elicitation of Certainty Equivalents: With the exception of the validation measure, all our experimental risk preference measures (introduced below) build on certainty equivalents. A certainty equivalent is the sure amount of money that makes a subject indifferent between a lottery and a sure amount of money. To find these amounts, we use the so-called bisection approach, in which participants have to choose between a sure payoff and a lottery in an iterative manner.[3] To keep it as simple as possible, we use only two-outcome lotteries, with the two outcomes being equally likely. Doing so not only facilitates understanding, but at the same time keeps concerns about the possibility of probability weighting within limits (as we show in Supplemental Appendix B.3). The choice in the first decision task was between 70 taler for sure and an equal chance of getting zero or 140 taler (see Figure A-1 in Supplemental Appendix A for a screenshot of our easy-to-understand choice task). If a subject chose the sure payoff, the amount of the sure payoff would be decreased in the next iteration, whereas if they chose the lottery, the sure payoff would be increased. From three such iterations, we approximate indifference amounts for a specific lottery, i.e., the certainty equivalents, by taking the mid-point of the interval in which it must lie. See Supplemental Appendix A.3 for a detailed, exemplary computation of a certainty equivalent for a lottery depending on participants' decisions.

To learn about risk preferences on a wider range of outcomes (and to establish utility curves, as we describe in the next section), we elicited the certainty equivalents of five more lotteries with amounts in the interval from 0 to 140 taler. Hence, subjects made 18 choices in total (for six certainty equivalents) in this part of the elicitation. Departing from the first lottery, denoted $L(0,140)$ for short, we create the next two lotteries: Denoting the elicited certainty equivalent to $L(0,140)$ with

[3]The bisection approach is not theoretically incentive compatible, as it may lead subjects not to reveal their preferences truthfully in order to increase their expected payoff (e.g., Harrison 1986). Ex post, by analyzing aggregate and individual choice behavior, we can investigate whether this has happened systematically, but we find no evidence for adolescents gambling the method (see Supplemental Appendix A.4). In light of our subject pool, following Bardsley et al. (2010), we had purposefully chosen this *behaviorally* incentive-compatible method, to avoid the complexity of *theoretically* incentive-compatible alternatives such as choice lists.

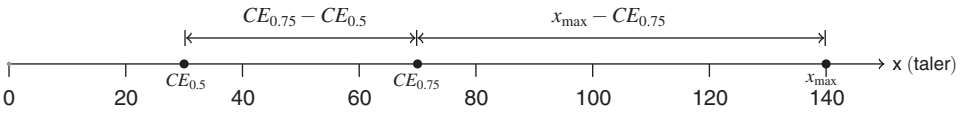

FIGURE 1. DYNAMICALLY OPTIMIZED ELICITATION OF CERTAINTY EQUIVALENTS

$CE_{0.5}$ (as the expected value is just half the maximal payoff), we next elicit the certainty equivalents $CE_{0.25}$ and $CE_{0.75}$ to the lotteries $L(0, CE_{0.5})$ and $L(CE_{0.5}, 140)$, respectively.

Dynamically Optimized Certainty Equivalent Method.— Now one of our innovations comes into play that will become important, as shown in the next section: In addition to these three elicited certainty equivalents, and depending on the differences between certainty equivalents, we elicit either $CE_{0.125}$ or $CE_{0.375}$, and either $CE_{0.625}$ or $CE_{0.875}$.

Figure 1 provides an illustration: In this scenario, the distance between $CE_{0.75}$ and $x_{max} = 140$ is larger than the distance between $CE_{0.5}$ and $CE_{0.75}$. Therefore, we elicit $CE_{0.875}$ (and not $CE_{0.625}$), to minimize the differences in elicited certainty equivalents. Put differently, for every participant we make sure that they decide over a wide range of monetary amounts. We thus avoid situations where, for large regions of monetary outcomes, no fine-grid decision is available. A last certainty equivalent minimizes the by then largest difference in certainty equivalents.[4]

Behavioral Prudence Task (Validation Task): In order to cross-check our new measures of (higher-order) risk preferences introduced below, we *additionally* elicited prudence using a version of the widely used "behavioral" classification task going back to Eeckhoudt and Schlesinger (2006). To account for our sample, we adapted the tasks from Heinrich and Shachat (2020), where compound lotteries are reduced following Maier and Rüger (2011), and let subjects make decisions in three additional tasks between two equally likely four-outcome options of the type (30, 70, 100, 100) versus (50, 50, 80, 120), see Figure F-1 in the Supplemental Appendix for an illustration of such a task. In the second task, the amounts 10, 90, 100, 100 (Option A) versus 50, 50, 60, 140 (Option B) were used, and 60, 100, 100, 100 (Option A) versus 80, 80, 80, 120 (Option B) were the corresponding amounts in the third task.

Risk Survey Measures: Finally, we also administered a survey question on risk attitudes. This was done because the literature often uses a single-item survey question about one's willingness to take risks, and then relates the responses to field behavior. Specifically, we use the standard question on willingness to take risk in

[4] If, for example, up to this last step, the certainty equivalents $CE_{0.125}$, $CE_{0.25}$, $CE_{0.5}$, $CE_{0.625}$, and $CE_{0.75}$ have been elicited, this could be one of the following certainty equivalents: $CE_{0.0625}$, $CE_{0.1875}$, $CE_{0.375}$, $CE_{0.5625}$, $CE_{0.6875}$, $CE_{0.875}$. See http://horp-otree-apps.sebastianoschneider.com/ for an illustration of the implementation (Schneider 2025a).

general, first included in the German Socio-Economic Panel (SOEP; Wagner, Frick, and Schupp 2007; Dohmen et al. 2011; Falk et al. 2018). Additionally, we use an adapted subset of the DOSPERT questionnaire (Weber, Blais, and Betz 2002) that was built to assess risk in the domains of health/safety as well as recreational, ethical, social, and financial decisions.[5] The survey questions used to address general risk-taking behavior are printed in Supplemental Appendix E.8. Adding survey questions to our study design allows us to relate our experimental, utility-based measures, in particular those of prudence and temperance, to those survey measures and examine if it is appropriate to interpret the latter as a measure of risk aversion, or whether they also capture higher-order risk preferences.

Time.—To measure time preferences, we elicited with a bisection approach the so-called future equivalent: the amount of money that makes a participant indifferent between receiving a given payment right away and receiving the elicited amount with a three-week delay. As our measure of time preference (or impatience), we use the ratio of the future equivalent to the earlier payoff. See Supplemental Appendix A.1 for details and Chabris et al. (2008) on the importance of accounting for time discounting when studying risk-taking behavior.

Cognitive Abilities.—Concerning our measure for cognitive abilities, we focus on fluid intelligence, and use a time-weighted average of scores in a standard matrix test and in a symbol-digit-correspondence task (Dohmen et al. 2010). See Supplemental Appendix A.2 for details.

C. *Selection of Field Behavior and Survey Design to Elicit Field Behavior*

At the end of experimental sessions, we administered an extensive questionnaire on field behavior. The selection of domains was motivated by the availability of theoretical models that related higher-order risk preferences to everyday field behavior. Then, for every identified domain of behavior, we carefully developed survey items to capture the given behavior adjusted to adolescents' reality of life, drawing on our rich set of best practices for this age group (Sutter, Zoller and Glätzle-Rützler 2019). Here we present the intuition for each domain and typical questions for elicitation. Supplemental Appendix E contains the full (verbatim) questionnaire. There, we also provide some theoretical background, including a general framework (in Supplemental Appendix E.1) to illustrate the potential role of prudence for the various domains of field behavior.

Health-Related Behavior.—Prudence has been linked theoretically to the health domain, for example by studying multivariate risk taking (e.g., Eeckhoudt, Rey, and Schlesinger 2007), modeling health-related prevention effort (e.g., Courbage and Rey 2006), the demand for medical care (e.g., Dardanoni and Wagstaff 1990),

[5] We use a subset of these questions to account for our underage sample: some questions (e.g., having an affair with a married person or cheating on one's tax return) would induce low variation and seemed inappropriate to ask adolescents as young as age 10. Moreover, we adapted some questions (e.g., using a helmet when riding a bike instead of riding a motorbike).

or medical treatment decisions (e.g., Bleichrodt, Crainich, and Eeckhoudt 2003). In the life reality of adolescents, health-related behavior is arguably best captured by considering behavior that risks their health status (or avoids such behavior).

To test the importance of higher-order risk preferences for health-related behavior, we include several questions in our questionnaire to capture such behavior. Notably, we include a novel, self-constructed scale consisting of six questions to capture smartphone and social-media addiction, as this kind of addictive behavior has gained tremendous attention over the last decade (e.g., Hormes, Kearns, and Timko 2014; Andreassen 2015; He, Turel, and Bechara 2017). Yet, to our knowledge, it has been ignored in the risk-taking literature so far.

Our scale on abusive smartphone usage comprises questions such as "When I feel bad or when I face a difficult task, I distract myself with my smartphone," "I feel uncomfortable (e.g., nervous or fretful or disquiet or a bit sad) when I cannot use my smartphone for a considerable time, because of an empty battery, no signal, or because my smartphone was taken away," or "I often check my phone while eating with my family to see if there are any news." In addition to abusive smartphone usage, our addictive behavior index comprises smoking and drinking behavior. Behavior that generally is a risk to health additionally includes items on, e.g., the BMI or physical inactivity. See Section E.3 in Supplemental Appendix E for all items used, as well as theoretical considerations based on our illustrative model (in Supplemental Appendix E.1) on how prudence might relate to health-related behavior.

Financial Decision-Making.—Prudence was already linked to financial decision-making, in particular precautionary saving by Leland (1968) via a positive third derivative of the utility function (see the illustrative model in Supplemental Appendix E.1). Temperance is theoretically related with less risky investment as a reaction to greater background risk (Kimball 1992; Gollier and Pratt 1996). The impact of prudence on insurance demand is negative when there is uncertainty about the value of the insured item (e.g., Fei and Schlesinger 2008). Using the "behavioral" elicitation method by Eeckhoudt and Schlesinger (2006), Noussair, Trautmann, and van de Kuilen (2014) have been the first to explore the relation between higher-order risk preferences and financial decision-making with experimental measures.

With respect to financial decision-making, we extend their seminal paper by studying adolescents. Naturally, adolescents take fewer financial decisions than adults. However, in particular in Germany, their financial education has long started. They all receive regular pocket money, cash gifts, or earn money in a side job, for which reason they are legally allowed to make financial decisions, including insurance choices, themselves (even at a smaller scale than adults). To study decision-making adjusted to their situations, and to investigate whether our sample follows the same general patterns as found for adults in Noussair, Trautmann, and van de Kuilen (2014), we include questions on saving, risky investments and insurance coverage.

To collect field behavior on saving, we ask students for example what fraction of a gifted 50 euro bill they would save, or how they handle their pocket money, where possible answers range from "I spend everything quickly" to "I save everything."

The equivalent to adults' preference for risky investment is assessed with questions like "Have you ever used money that was originally intended for something else at a subsequent date (e.g., for holidays or a present), for a bet or invested it in stocks"? To address insurance demand, we ask for the possession of a bike or phone insurance, and whether students bought it themselves. See Section E.4 in Supplemental Appendix E for the full list of items targeting financial behavior, as well as sketches of the theoretical models linking higher-order risk preferences to financial decision-making.

General Prevention and Eco-Friendly Behavior.—Prevention behavior (as an effort to reduce the probability of an adverse event, Ehrlich and Becker 1972) has been theoretically connected to prudence. In one-period models, the preventative effort and the potential loss are contemporaneous. Contrary to intuition, in this setting, and with equally likely loss and no-loss states, prudence has a negative impact on the optimal level of prevention, since the prudent agent prefers to accumulate wealth to face future risks instead of investing in preventative effort (Eeckhoudt and Gollier 2005). In two-period models, the preventative effort precedes the potential loss. In that setting, the relation between prudence and the optimal level of prevention is positive (Menegatti 2009). Regarding eco-friendly behavior, the model by Bramoullé and Treich (2009) addresses global warming and climate change outside the prevention context and predicts a negative effect of prudence.

In our survey, we ask for actions preventing an unwanted event that might arguably happen on the same day (one-period model) and in the more distant future (two-period model). Moreover, we add questions on eco-friendly behavior.

We ask, for example, for agreement to the statement, "Since I think of packing something to eat and drink during longer journeys by bus, train or car, I am not hungry or thirsty in such situations," to capture effort provision in order to prevent hunger, an event that is likely to happen on the same day. Contrarily, agreement to the statement, "Since I do not know yet what I would like to become later, I try to get good grades to keep all possibilities open to me," indicates effort provision in order to prevent a missed chance in the future. We address eco-friendly behavior by asking, for example, whether students separate their waste, use reusable coffee cups, bottles, or shopping bags, take their bike when possible (instead of a bus or car), or turn down the heating if leaving a room. See Section E.5 in Supplemental Appendix E for the questions as well as theoretical considerations on the relationships of short- and long-term prevention effort, eco-friendly behavior, and higher-order risk preferences.

Planning Behavior.—In a seminal paper, Kimball (1990) developed the "theory of the optimal response of decision variables to risk (which includes precautionary saving as a subcase)," and *defined* prudence as the "sensitivity of the optimal choice of a decision variable to risk" (Kimball 1990, p. 54), thus the degree to which plans are adapted to risk.

We test if we find support for prudence in this more general sense as a measure of cautious planning. The decision situations considered in our questions are, e.g., preparation for a class test, or being on time for a meeting.

We ask participants how much more time they invest in the given situations, if risk increases by, for example, a more uncertain scope of the class test, or by red lights on the way to the meeting. See Section E.6 in Supplemental Appendix E for the exact wording of the items used and some notes on how to think of this kind of behavior within our illustrative model.

Preference for Competitive Income.—Selecting into a job as civil servant has been related to risk aversion and prudence by Fuchs-Schündeln and Schündeln (2005): They attribute a lower heterogeneity in precautionary savings to self-selection of risk averse and prudent individuals into low income risk occupations.

We hence include in our survey questions about participants' preferences for a competitive income. We ask questions on the preference for risky/riskless occupations in the future (e.g., "Would you like to become a civil servant?"). We moreover add a tangible proxy for the choice of job type, and ask whether participants would have preferred a fixed or a tournament payment instead of the piece rate payment for the cognitive ability tasks. Uncertainty about future income causes prudent individuals to engage in precautionary behavior by accumulating more money in the present. For this type of questions, prudent individuals might prefer a payment scheme that yields a higher payoff in expectation. See Section E.7 in Supplemental Appendix E for the exact wording of all items used.

Building Indices of Field Behavior.—For most of the questionnaire, we obtained data for all 658 students.[6] We build indices for the different domains of behavior. Importantly, all indices contain all information that we have gathered for an individual in the respective domain. Since both meaning and variation in a given question might differ between the age groups, and because some questions were not asked to the younger cohorts (e.g., on drinking alcohol), we first aggregate results per age cohort. For every domain of field behavior and every age cohort, we first run a principal component analysis (PCA) using all the variables resulting from the questions that we had asked in the given age cohort for a given domain of field behavior. If all the variation in all those variables should be reduced to just one variable or dimension, the first component of the PCA would be the best pick. A component is essentially a linear combination with component-specific weights for all the variables, and it is given by the vector of weights, with one weight for each variable. We thus build, for every domain of field behavior, indices per age cohort using weights from the first component of the PCA.[7] We then center and standardize each of these age-cohort indices so that for all age cohorts, the resulting index is on the same scale. For final aggregation, i.e., to obtain just one index of field behavior for every domain, for each individual, we simply pick the relevant age-cohort index in accordance with our

[6] Some questions, however, were not asked to the youngest students (e.g., drinking or dating behavior), in accordance with our agreements with participating schools and to get meaningful results (see the respective notes to a question in Supplemental Appendix E).

[7] For example, for planning behavior, our outcome for the twelfth grade is created as linear combination $.031 \times \text{Q1} + 0.24 \times \text{Q2} + 0.45 \times \text{Q3}$, where the questions ask about preparation for a class test (Q1), travel planning (Q2), and handing in a term paper (Q3). For the tenth graders, e.g., the weight on the class test (Q1) decreases to 0.22, so Q1 seems to have less importance for them, perhaps as results do not enter their school-leaving certificate.

pre-registered analysis plan.[8] Therefore, all of our dependent variables are centered, standardized, and continuous, and we thus use OLS regression.

II. Methods to Measure (Higher-Order) Risk Preferences

Now we present our methods to obtain raw and sophisticated measures of (higher-order) risk preferences based on the experimentally elicited certainty equivalents as described in the previous section. Here, we first describe our raw measures. We develop a raw, simple measure of prudence in order to show that our results are essentially the same, even if we use measures that result from (few) certainty equivalents by means of addition, subtraction, division and multiplication only. We then present the sophisticated method building on Schneider, Ibanez, and Riener (2025), to which we contribute by introducing what we call the dynamically optimized certainty equivalent method and with our easy-to-understand implementation.

Our methods build on utility-based definitions of higher-order risk preferences to be able to obtain utility-based intensity measures. This is important from a theoretical point, as all the theoretical work relating higher-order risk preferences to field behavior involves these measures, and there are no alternatives to elicit them cleanly. Neither is there a clear relation between these measures and the prudence and temperance premia resulting from the method by Ebert and Wiesen (2014), nor to the "count measure" building on Eeckhoudt and Schlesinger (2006). This measure counts the number of prudent decisions in "behavioral," binary classification tasks and uses this number as an approximation of prudence intensity.

Another argument for our methods and the resulting utility-based measures is that they allow for a very quick and easy-to-understand elicitation task. Using this implementation, all the information to construct intensity measures of risk aversion, prudence and temperance can be elicited in about two minutes, compared to the one-and-a-half-hour format of the method by Ebert and Wiesen (2014). Moreover, using our easy-to-understand task also avoids the high and increasing complexity inherent in the so-called risk apportionment method (Eeckhoudt and Schlesinger 2006), where with every order of risk preference an additional compound lottery is introduced. For example, to measure temperance, each option consists of the combination of three lotteries. As ambiguity aversion seems to be almost identical to aversion to compound lotteries, choice behavior in these compound tasks might actually reflect a mixture of higher-order risk preferences and ambiguity aversion (e.g., Chapman et al. 2023). While reduction of compound lotteries is possible, the results from Heinrich and Shachat (2020) call for caution in using this approach already for prudence (at least with adolescents): For one of the three tasks they employ to measure prudence, behavior cannot be distinguished from random choices even when pooling all age cohorts. Our methods, in turn, are very intuitive and cognitively accessible, even to a sample of adolescents or

[8] The two indices capturing insurance demand and the preference for a competitive income are not aggregated using PCA. In case of no missing values, they consist of only two binary questions (plus two follow-up questions) and four binary questions, respectively. Hence, their support consists of 9 and 16 elements, respectively. Here, using weights from a PCA per age cohort would shift the support for every age cohort marginally, thus introducing noise in the measure when aggregating the indices rather than precision. Therefore, we compute z-scores for every item and add them.

the general population. Thus, they allow for measuring prudence and temperance with important samples, including in repeated household panels: Participants only have to choose repeatedly between a safe outcome and a two-outcome lottery with equiprobable outcomes.

Lastly, our sophisticated method allows to address errors in measurement that obviously complicate relating experimental measures with anything else (see Gillen, Snowberg, and Yariv 2019, for an illustration). Actually, already our raw measure can be expected to be a better measure compared to alternatives (due to a larger support, and hence, variation, for being a continuous measure, together with a very low level of complexity). Yet, the predictive quality of the sophisticated measures is unmatched by the raw measures. The combination of using both sets, however, is valuable beyond showing the robustness of our results to using plain and transparent measures. It is also very helpful to explore (and rule out) potential issues of noise-induced (correlated) measurement error, as we explain in Supplemental Appendix B.2.

A. *Raw Experimental Measures of Risk Aversion and Prudence*

We first show how we compute simple, nonprocessed intensity measures for risk aversion and prudence. For a simple, "behavioral" (i.e., model-free) intensity measure of risk aversion, one certainty equivalent suffices. We define the normalized risk premium for a two-outcome lottery with equally likely outcomes as

$$r^{CE} = \big((x_{high} - x_{low})/2 + x_{low} - CE\big)/(x_{high} - x_{low}),$$

where CE denotes the certainty equivalent, and x_{high} and x_{low} denote the high and low outcomes of the corresponding lottery, respectively. It is simply the standardized difference between the expected value and the certainty equivalent, where positive values imply risk aversion, and negative values imply risk-seeking behavior. For the first certainty equivalent that we elicit, we can compute this measure and compare it across all individuals, as $x_{high} = 140$, and $x_{low} = 0$ for all participants in the first lottery.

The risk premium, and hence, our simple measure r^{CE}, is proportional to the Arrow-Pratt measure $r = -u''/u'$ under EUT, a result well-known due to Pratt (1964). For our data, this raw, behavioral measure of risk aversion r^{CE} will be highly correlated with r as derived from our sophisticated method (introduced below), as we will see in the results Section III.

For the (utility-based) intensity measure of prudence $p = u'''/u'$, a similar link to choice behavior can be established: From Leland (1968, eq. 19), it follows that

$$p = (-r)^2 - r'. \tag{1}$$

Hence, those who exhibit decreasing risk aversion $(-r' > 0)$, that is, those who ask for a lower risk premium for the same risk as their wealth level increases, are prudent (sufficient, but not necessary condition). In addition, from (1), we see that *risk aversion is not at all a necessity for prudence*, emphasizing the importance of a nonparametric approach to "allow" for prudent risk seekers.

We can use equation (1) to build a simple measure of prudence that is based on choice patterns in only a few choices, which we still can easily interpret: We approximate the derivative of r with the change in r^{CE} that is associated with an increase in the lowest outcome of the corresponding lotteries, i.e., an increase in the level of wealth to which risk is added. As the first three certainty equivalents that we elicit correspond to the same levels of utility for all participants ($u = 0.25, 0.5,$ and 0.75, see Section IB), and to reduce the influence of decision errors, we average two measures of decreasing risk aversion and the corresponding three measures of risk aversion r^{CE}, and use as our simple measure of prudence:[9]

$$(2) \qquad p = r^2 - r' \approx p^{CE} = \left(\left(r^{CE_{0.25}} + r^{CE_{0.5}} + r^{CE_{0.75}}\right)/3\right)^2 + \left(\left(r^{CE_{0.5}} - r^{CE_{0.75}}\right)/CE_{0.5} + \left(r^{CE_{0.25}} - r^{CE_{0.75}}\right)/CE_{0.5}\right)/2.$$

Despite a slightly more complex expression compared to r^{CE}, we will see in the results Section III that also our simple prudence measure p^{CE} is highly correlated with its sophisticated counterpart p.

B. *Sophisticated Method for Measuring (Higher-Order) Risk Preferences*

The intuition of the sophisticated method is as follows: First, we determine certainty equivalents (as described above). These values yield x-coordinates of the points of the utility function that we seek to obtain, and their y-values result from straight-forward calculations under EUT as explained below. Then, we transform these points on the utility function (utility points for short) into individual utility functions, using a (local) nonparametric regression approach for this (generalized) interpolation task. Finally, utility-based intensity measures can be derived from predicted utility functions. We now explain each step in detail.

Part 1: From Certainty Equivalents to Utility Points.—From the lottery choices of participants, i.e., the elicited six certainty equivalents, we obtain eight utility points as follows. We, first, normalize the utility function, such that for the highest possible outcome across all tasks $x_{max} = 140$ taler, we set $u(x_{max}) = 1$, and for $x_{min} = 0$ taler, the lowest possible outcome in our procedure, we have $u(x_{min}) = 0$. Then, the expected utility of a lottery with these two equally likely outcomes is $0.5u(x_{max}) + 0.5u(x_{min}) = 0.5$. As, per definition, subjects should be indifferent between receiving the lottery and the certainty equivalent $CE_{0.5}$ for this lottery, the utility to them must be the same. Thus, $u(CE_{0.5}) = 0.5$. We thus have fixed the third utility point $(CE_{0.5}, 0.5)$. Iterating this procedure, we obtain $0.5u(x_{min}) + 0.5u(CE_{0.5}) = 0.25 = u(CE_{0.25})$. Similarly, we obtain $(CE_{0.75}, u(CE_{0.75}))$ with the certainty equivalent $CE_{0.75}$ corresponding to the lottery $L(CE_{0.5}, x_{max})$.

[9] The first three certainty equivalents are $CE_{0.5}$, $CE_{0.25}$, and $CE_{0.75}$, with vectors $(x_{min}, x_{max}) = (0, 140)$, $(0, CE_{0.5})$, and $(CE_{0.5}, 140)$ containing the outcomes of the corresponding lotteries. Thus, for $r^{CE_{0.5}}$ and for $r^{CE_{0.25}}$ the wealth level to which risk is added is 0, and for $r^{CE_{0.75}}$ it is $CE_{0.5}$.

Improving Predictions with the Dynamically Optimized Certainty Equivalent Method.—Now our contribution to the method, an extension of the certainty equivalent method, comes into play: As we wish to predict utility functions with our experimental measurements (i.e., the certainty equivalents), it is central to have information about individuals' utility on a wide range of monetary outcomes. In statistics, it is well-known that an increase in the variation of measurement points (the certainty equivalents in our case) increases prediction quality. Hence, to avoid concentration of measurement points, and avoid situations where for large regions of outcomes no measurement (utility point) is available, we use the *dynamically optimized certainty equivalent method* (described in Section IIB). Consequently, we obtain either utility points $(CE_{0.125}, 0.125)$ or $(CE_{0.375}, 0.375)$ and either $(CE_{0.625}, 0.625)$ or $(CE_{0.875}, 0.875)$, depending on which of the options would increase the variation of measurement points. In a similar fashion, the last utility point is obtained. That way, we ensure the derivation of meaningful preference measures both from a conceptual and a statistical point of view.

So far, we have explained how we obtain six utility points from our experimentally elicited certainty equivalents (in addition to the two fixed points), and presented our contribution regarding the elicitation procedure. We now move on to the statistical part.

Part 2: Nonparametric Prediction of Utility Functions from Utility Points.—We now describe how we obtain nonparametric utility functions from the elicited utility points by means of a (generalized) interpolation approach, in a way that accounts for errors or noise in the utility points. This is the statistical core of the method, developed in Schneider, Ibanez and Riener (2025) and implemented in the R-package utilityFunctionTools (Schneider 2025b), which we use in this paper. The method relies on a nonparametric regression approach—penalized (P-)spline regression (Eilers and Marx 1996)—that has been adapted to the case of individual utility functions such that (utility-based) intensity measures of higher-order risk preferences can be computed. Here, we give a brief, nontechnical overview of the method to highlight the key advantages for our purpose. In Supplemental Appendix A.5 we provide a more extensive summary, including the technical details.

As the name reveals, penalized (P-)spline regression is a spline regression approach that involves a penalty. The idea of spline regression essentially is to make polynomial regression more flexible without severely over- or underfitting the data (a well-known complication with polynomial regression, see Runge 1901). This is achieved by cutting the interval, on which the regression polynomials (such as $x, x^2, x^3, \dots$) are defined, into potentially too many subintervals. The polynomial functions to regress on are now piecewise locally defined, that is, only on few subintervals they are different from zero. This generates the necessary flexibility to fit utility functions as it allows to have different functional forms in different subintervals. For example in the first subinterval, we can have a cubic function, and in the last, we can have a rather linear one. Here, we use polynomials of degree six in order to allow for a nonlinear fourth derivative of the utility function, which characterizes temperance.

The penalty is the crucial ingredient with which we implement error correction. At the same time, it addresses over- and underfitting and allows fitting a curve of

degree six even with less than seven points (we use eight). Simply put, our penalty measures how "unsmooth," "wiggly" or "rough" a utility curve will be. Technically speaking, the penalty captures the variation in regression coefficients of the polynomial functions defined on close-by subintervals (see Supplemental Appendix A.5 for details). It is simply added with a to be determined weight λ to the standard linear regression objective function. The higher λ, the more we weight the penalty compared to the usual objective of minimizing squared residuals. Hence, the more we weight (i.e., penalize) roughness, the more we relax the objective to fit the data perfectly. At the same time, an increase in λ results in an increasingly smooth utility function, resulting from a lower variation in neighboring regression coefficients, i.e., a lower penalty. Figure 2 illustrates predicted utility functions, depending on different values for λ.

How does λ relate to noise and error correction? Obviously, with perfectly measured utility points, we want the fit to the points to be as high as possible ($\lambda = 0$). If we, similar to instrumental variable approaches to address errors in variables (e.g., Durbin 1954; Gillen, Snowberg, and Yariv 2019), interpret inconsistent variation that cannot be predicted with other measurements as noise or error, we may use $\lambda > 0$ to correct for this: A higher λ attaches, roughly speaking, a lower weight to a single utility point, and this applies in particular to points that deviate from the shape of the utility curve that the remaining points suggest. The idea of error correction hence is a "global" one, where all information is considered (as opposed to no error correction as in, e.g., linear interpolation, where the focus is "local" and only two points are considered at a time). Regarding this "global" focus, our approach is similar to a parametric fitting approach. However, it is much more flexible, and allows also for cases that would be ruled out by the common parametric utility functions (such as prudent individuals who behave in a risk-seeking manner; something that our approch can capture).

The penalty weight λ then is determined using cross-validation following the line of thought just described. This procedure determines the shape that the majority of points suggests by picking the λ that maximizes the predictive quality of the predicted utility function in the following sense. For every individual utility function to be fitted, we iteratively leave two (elicited) points out from fitting the model such that they can be used for prediction.[10] We do this for every possible combination of leaving out two points, then we average the squared prediction error for the points left out, and obtain a mean squared prediction error for a given individual and given penalty weight λ. As we vary λ, we obtain a mapping from penalty weights to predictive quality measured by mean squared prediction error. For every individual, we then choose the penalty weight that maximizes the predictive quality of the resulting utility function in this sense. Thus, a decision error resulting in a shifted point will lead for a low λ to low predictive quality of the fitted model for the other points; as λ increases, this shifted point becomes less important, and so the predictive quality increases.

More generally, for a "noisy" respondent, whose points deviate a lot from their own "trend," the predictive quality of a fitted function for the points left out will be poor for a low penalty weight. It will increase with a higher penalty weight (i.e.,

[10] We have chosen two points, corresponding to one-third of elicited utility points, because choice reversals for one-third of choices are commonly assumed and observed (e.g., Abdellaoui, Bleichrodt, and Paraschiv 2007).

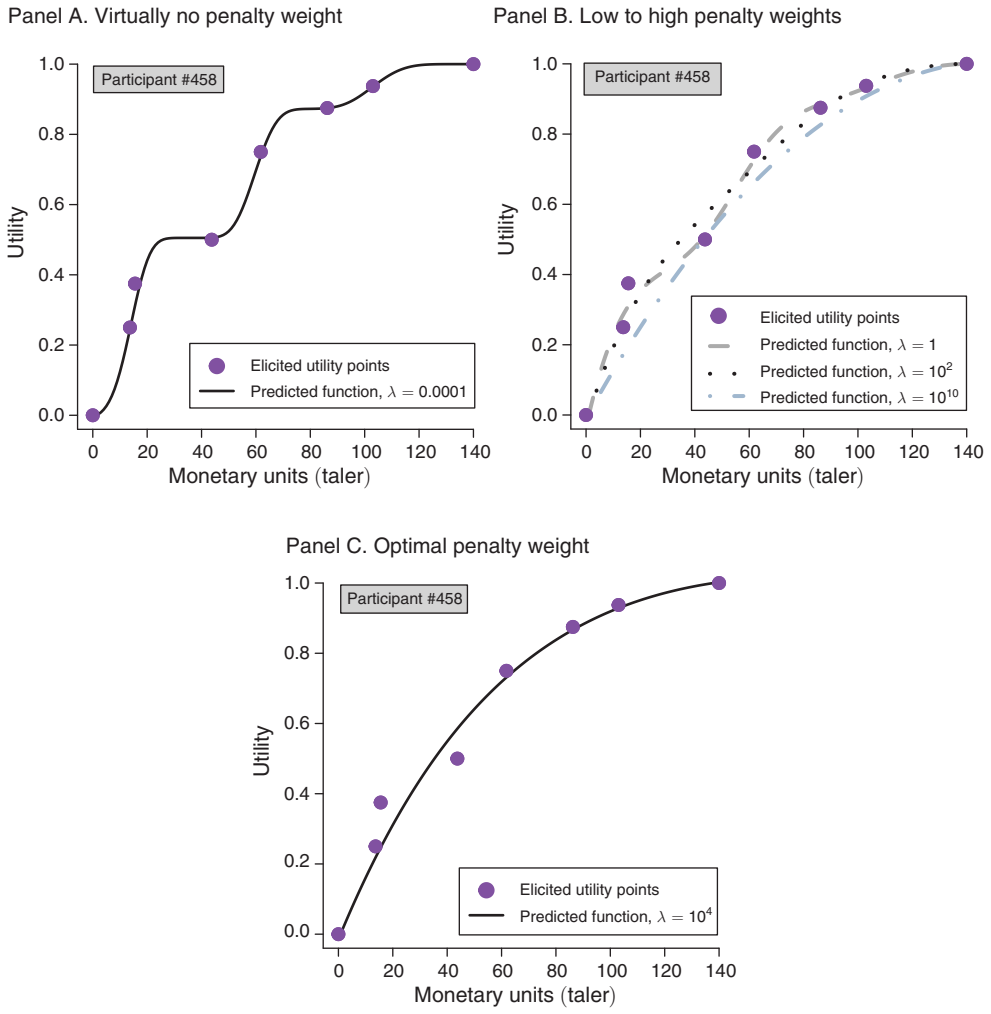

FIGURE 2. PREDICTED UTILITY FUNCTIONS FOR DIFFERENT PENALTY WEIGHTS λ

Notes: This figure shows elicited utility points and predicted utility functions using our method for different values of the penalty weight λ for a representative individual (with respect to their prudence intensity). In panel C, two points that are classified as noisy by the cross-validation procedure do not inform the shape of the predicted utility curve.

higher degree of smoothness). Roughly speaking, the less coherent (the noisier) the picture that any one of all the possible subsets of utility points delivers, the higher the necessary smoothness. And the higher the necessary smoothness, the lower the curvature (see Figure 2, panel B) and the higher-order derivatives, too. Since our measures of (higher-order) risk preferences (defined below) are functions of the curvature and higher-order derivatives, this rules out that we mistakenly interpret a noisy participant as particularly risk averse or risk seeking, (im)prudent, or (in)temperate.[11]

[11] In practice, the theoretical possibility of noise-induced downward bias in our sophisticated measures that this feature is creating has only limited relevance, as we show in Supplemental Appendix B.2. Even when adding arguably exaggerated degrees of noise to our data, correlation coefficients with the true, undistorted measures remain

Measures of (Higher-Order) Risk Preferences and Their Derivation.—Derivatives of the predicted utility functions can be calculated analytically with a closed form solution following our spline-regression approach. This is a crucial advantage for obtaining utility-based measures of risk preferences.

For individual risk attitude, we use the Arrow-Pratt measure (Pratt 1964), $r = -u''/u'$, where positive (negative) values indicate risk aversion (risk seeking) and risk neutrality corresponds to $r = 0$. We measure an individual's prudence level with the measure popularized by Crainich and Eeckhoudt (2008), $p = u'''/u'$, where positive (negative) values indicate (im)prudence.[12] Positive values also correspond to downside risk aversion (Menezes, Geiss, and Tressler 1980). Moreover, u'''/u' is also a measure of left-skewness aversion (Modica and Scarsini 2005). As an individual measure of temperance, we use the measure by Denuit and Eeckhoudt (2010), $t = -u^{(iv)}/u'$, where again positive (negative) values indicate (in)temperance. Denuit and Eeckhoudt (2010) show that t is also a measure of kurtosis aversion.

To obtain these measures, we first predict for each derivative separately its value at 1,000 evenly spaced points covering the interval from x_{min} to x_{max}. Then we build the mean of the predictions for each derivative, which yields individual measures of u', u'', u''' and $u^{(iv)}$. Finally, we compute the above measures. Note that the division by u' is only a normalization of the utility function so as not to change the measure if utility is multiplied by a constant; it has, however, no further meaning (Pratt 1964). Performing this normalization only after averaging over the different wealth levels has two advantages: First, the dependence of the measures on wealth is "averaged" out. This is desired, as we are connecting these measures to field behavior in general instead of at a specific wealth level. Second, we avoid adding noise to the measures (resulting, for example, from division by values that are virtually zero).

III. Results I: Higher-Order Risk Preferences and Their Determinants

In our pre-analysis plan, we have specified directional hypotheses for the relation between risk aversion, prudence and temperance, and the influence factors age, cognitive ability, and gender. For impatience and all other possible influence factors investigated (see the notes to Table 1), we correct p-values for multiple hypothesis testing (Romano and Wolf 2005a, b, 2016) as implemented by Clarke, Romano, and Wolf (2020). For ease of exposition, all tests reported in this paper are two-sided. Before we investigate the (higher-order) risk preferences, their determinants and their relations, we empirically address the quality of our measures.

high (0.99 for risk aversion, 0.98 for prudence, and 0.96 for temperance), which is still the case when we vary the simulation approach. In the absence of noise, our measures are unbiased (see Schneider, Ibanez, and Riener 2025). Nevertheless, in the spirit of a sandwich estimator, this noise-induced downward bias (even if primarily of theoretical relevance) allows to approximate "the true" relationship with field behavior in the presence of noise from below. Our raw measures, in turn, will yield the approximation from above, due to a theoretically possible, noise-induced upward bias (see Supplemental Appendix B.2). As results using both sets of measures coincide, as we will see, noise can only have a negligible impact on our measures, and no impact on our results.

[12] Note that our measure is different from the well-known measure introduced by Kimball (1990) in order to be able to compare risk-averse and risk-seeking subjects: For a prudent individual, the Kimball measure $-u'''/u''$ might be positive or negative depending on their risk attitudes.

Table 1—Influence Factors of Risk Aversion, Prudence, and Temperance

	Risk aversion (r) (1)		Prudence (p) (2)		Temperance (t) (3)	
Age (in years)	−0.014	(0.015)	−0.007	(0.019)	0.002	(0.015)
Cognitive ability	−0.132	(0.043)	−0.062	(0.058)	−0.118	(0.042)
Female (= 1)	0.259	(0.091)	0.205	(0.105)	0.166	(0.085)
Impatience	−0.889	(0.275)	−0.584	(0.234)	−0.703	(0.283)
Other factors	10		10		10	
R^2	0.08		0.06		0.06	
Observations	658		658		658	

Notes: OLS regressions of utility-based measures of (higher-order) risk preferences. Positive coefficients imply increasing risk aversion, prudence, and temperance, which are expressed in standard deviations. Cognitive ability scores are standardized, such that above average scores are positive. Other possible influence factors controlled for are relative math grade, relative German grade (where positive variables imply above average performance relative to the grade), the amount of pocket money per week, relative BMI, the number of siblings, religion, migration background, parents' education, as well as their occupation; see Supplemental Appendix Tables D-3, D-4, and D-5 for detailed regressions results, and Table D-6 for regression results of simple measures of risk aversion (r^{CE}) and prudence (p^{CE}). Robust standard errors clustered at the session level in parentheses.

A. *Identification, Validation, and Sanity Checks*

Identification of (higher-order) risk preferences with our methods is evident, as our measures are directly derived from the theoretical definitions. Since we use two-outcome lotteries only, identification of, e.g., prudence comes *with the change of decisions as the amounts of the lotteries change*; it cannot be derived from an isolated decision in these tasks.

We nevertheless start with an empirical validation of our methods to show that the theoretically apparent identification of (higher-order) risk preferences is unaffected by the operationalization of our methods. We find a significantly positive relation ($p < 0.001$) between the number of prudent choices in the "behavioral" classification tasks à la Eeckhoudt and Schlesinger (2006) (see "Behavioral Prudence Task" in Section IB) and our two utility-based intensity measures of prudence (p and p^{CE}); see Table D-2 in the Supplemental Appendix.[13]

Next, we investigate the relation between our raw and our sophisticated measures. The Pearson correlation between the nonprocessed measure r^{CE} and the risk attitude r (that is derived from a predicted utility function according to Section II) is large and significant, and we obtain $\rho\left(r, r^{CE}\right) = 0.87$ (p-value < 0.001).

[13] While it is probably indisputable that a relationship should exist between our (utility-based) intensity measure and the "count measure" based on the behavioral classification tasks due to Eeckhoudt and Schlesinger (2006), it is less clear what to expect regarding the nature of their relation and its size, for the unclear interpretation of the latter. Moreover, Heinrich and Shachat (2020) report that for one of the three behavioral classification tasks that they employ (and that we use as validation tasks here), behavior cannot be distinguished from random choices even when pooling all age cohorts; for the youngest cohort, even for all three behavioral classification tasks choices are not different from random choices in their study. Hence, complexity and resulting measurement error is an issue, too. As a consequence, the correlation coefficients have no clear interpretation, either. These issues notwithstanding, when looking at our oldest age cohort, we find that the median intensity measure p of those who never chose the prudent option in the behavioral classification task is negative, where it is positive for those who chose the prudent option three times. The overall median of this age cohort is positive and right in the middle of these figures.

Moreover, and despite a slightly more complex operationalization compared to r^{CE}, the Pearson correlation between our raw prudence measure p^{CE} and the sophisticated measure p (that is derived from a predicted utility function) is large and significant: $\rho\left(p, p^{CE}\right) = 0.71$ (p-value < 0.001).

B. *Risk Aversion and Its Determinants*

We estimate a significantly positive mean Arrow-Pratt coefficient of risk aversion, expressed in standard deviations, of $r = 0.46$ (median: 0.35), with 0 indicating risk neutrality (p-value < 0.001, Wilcoxon signed-rank test). For 70 percent of our sample, we estimate a positive Arrow-Pratt coefficient, implying risk aversion (see Figure D-1 in the Supplemental Appendix for the full distribution). A regression including a measure for cognitive abilities and demographic background variables is shown in column 1 of Table 1. Females are significantly more risk-averse than males and individuals with higher cognitive abilities are significantly less risk-averse. Age is unrelated with risk aversion, once we control for cognitive abilities (see Table D-3 in the Supplemental Appendix). Impatience is significantly related with a lower degree of risk aversion.

Qualitatively, we obtain the same results using the simple measure of risk aversion (r^{CE}); see column 1 of Supplemental Appendix Table D-6. The mean (median) value of the simple risk aversion measure, expressed in standard deviations, is $r^{CE} = 0.36$ (0.27), with 0 indicating risk neutrality (p-value < 0.001, Wilcoxon signed-rank test). For 62 percent of our sample this measure is positive, implying risk aversion.

C. *Prudence and Its Determinants*

The significantly positive mean estimate (p-value < 0.001, Wilcoxon signed-rank test) of the Crainich-Eeckhoudt measure expressed in standard deviations is $p = 0.59$ (median: 0.22). For 67 percent of our sample, we estimate a positive Crainich-Eeckhoudt measure, implying prudence (see Figure D-1 in Supplemental Appendix D for the full distribution). The regression shown in column 2 of Table 1 reveals that females are more prudent than males. However, neither cognitive abilities nor age are significantly related with prudence.

As was the case for risk aversion, qualitatively, we obtain the same results using the simple measure of prudence (p^{CE}), see column 2 of Supplemental Appendix Table D-6. The mean (median) value of the simple prudence measure, expressed in standard deviations, is $p^{CE} = 0.80$ (0.36), with 0 indicating prudence neutrality (p-value < 0.001, Wilcoxon signed-rank test). For 88 percent of our sample, this measure is positive.

D. *Temperance and Its Determinants*

For the Denuit-Eeckhoudt measure of temperance, our mean estimate is $t = 0.30$ (median: 0.02), again expressed in standard deviations (p-value < 0.001, Wilcoxon signed-rank test). Having a positive Denuit-Eeckhoudt measure, 57 percent of our sample are classified as temperate (see Supplemental Appendix

Figure D-1 for the distribution). Column 3 of Table 1 shows that females are more temperate than males, and students with higher cognitive abilities and larger impatience are less temperate. We find no age effect.

E. *Relation between Elicited Risk Preference Measures*

The (sophisticated) measures of risk aversion, prudence, and temperance are significantly correlated in our sample (p-value < 0.001 for all pairwise correlations): $\rho = 0.56$ for risk aversion and prudence, $\rho = 0.87$ for risk aversion and temperance, and $\rho = 0.65$ for prudence and temperance.[14] Similarly, for our simple, nonprocessed measures r^{CE} and p^{CE} we find $\rho = 0.59$ for the relation of risk aversion and prudence (p-value < 0.001).

IV. Results II: Experimental Measures of Higher-Order Risk Preferences and Field Behavior

Now we turn to the empirical relationship between risk preferences and field behavior. Before discussing the results in Table 2, we note that we standardize our measures of risk preferences for comparison reasons. Hence, coefficients report the change in the field outcomes (in standard deviations) that is associated with a one-standard deviation increase in these measures. Since our sophisticated measures of (higher-order) risk preferences are estimates (generated regressors) and thus involve some degree of error, we use bootstrapped standard errors. Yet, our results remain almost identical if we use multiple imputation (e.g., Rubin 1996) instead. Also recall from Section III that our measures of risk preferences are considerably correlated. To ease interpretation of coefficients and to avoid multicollinearity, we orthogonalize these measures.[15] To this end, either the risk aversion, the prudence, or the temperance measure remains as is, and we pick the measure that matters according to theory for the respective index (as described in Section IC and stated in the notes to Supplemental Appendix Tables D-9 to D-19). If the model from which we derive a hypothesis does not suggest one measure as the most important, we take prudence because it has the lowest correlation with the other two measures. For example, if risk aversion is the measure we expect to matter, we remove the variation from the other measures that is already explained by risk aversion (and additionally prudence in the case of temperance). We do so by first running an OLS regression and then predicting prudence on the individual level, using the risk aversion measure as an explanatory variable. Next, we subtract this prediction from the actual measure to obtain the variation not yet explained by risk aversion, which becomes our orthogonalized measure of prudence (i.e., the residuals from the regression). Then, the same is done for temperance (using risk aversion and prudence as explanatory variables in this example). In the regressions, we then include

[14] Ebert and Wiesen (2014), the only published study to date that has measured the intensity of higher-order risk preferences, report almost the same Spearman correlations for prudence and risk aversion and prudence and temperance as we find in our data (note that we report Pearson coefficients above).

[15] Ridge regression (e.g., Hoerl and Kennard 1970) is another way of handling multicollinearity. Results are very similar, but as Ridge regression yields biased coefficients (Gruber 2017) and for ease of interpretation of coefficients, we present results from orthogonalized measures in the main text.

Table 2—Relevance of Risk Preferences in Predicting Field Behavior

Coefficient of risk preferences (columns show estimated coefficients)	Simple measures (one regression each)			Simple measures (joint est.)	
	Simple risk (r^{CE})	Simple prudence (counting measure)	Simple prudence (intensity p^{CE})	Risk aversion (r^{CE})	Prudence (p^{CE})
	(1)	(2)	(3)	(4)	(5)
Panel A. Health-related behavior					
Unhealthy behavior	−0.031	−0.085	−0.112	0.045	−0.110
	(0.033)	(0.033)	(0.034)	(0.044)	(0.033)
Addictive behavior	−0.036	−0.085	−0.119	0.044	−0.117
	(0.035)	(0.032)	(0.032)	(0.045)	(0.032)
Smartphone addiction	−0.038	−0.071	−0.132	0.051	−0.129
	(0.037)	(0.032)	(0.030)	(0.048)	(0.030)
Panel B. Financial decision-making					
Saving (with debt)	0.067	0.011	0.086	0.043	0.087
	(0.043)	(0.023)	(0.036)	(0.048)	(0.037)
Risky investment	−0.056	0.010	−0.017	−0.056	−0.019
	(0.036)	(0.038)	(0.036)	(0.044)	(0.036)
Insurance demand	−0.025	0.005	−0.028	−0.014	−0.029
	(0.020)	(0.022)	(0.025)	(0.021)	(0.025)
Panel C. Prevention and eco-friendly behavior					
Prevention (short term)	0.020	−0.027	−0.067	0.084	−0.065
	(0.040)	(0.033)	(0.038)	(0.043)	(0.038)
Prevention (long term)	0.124	−0.008	−0.020	0.197	−0.014
	(0.030)	(0.038)	(0.034)	(0.028)	(0.033)
Eco-friendly behavior	0.094	−0.011	0.021	0.119	0.025
	(0.038)	(0.028)	(0.027)	(0.042)	(0.028)
Panel D. Planning behavior and preference for competitive income					
Cautious planning	0.062	−0.044	0.100	0.007	0.100
	(0.035)	(0.042)	(0.031)	(0.042)	(0.031)
Preference for competitive income	−0.075	0.008	−0.001	−0.075	−0.003
	(0.022)	(0.021)	(0.025)	(0.022)	(0.024)

(*continued*)

the untouched measure and the so-residualized ones. As a robustness check, we run regressions on correlated measures. Note that due to multicollinearity coefficients and standard errors may then be imprecise. However, the results (in Supplemental Appendix Tables D-20 and D-21) are very similar to those discussed here.

We noted that, to date, the large majority of studies that relate risk measures to field behavior do not consider prudence and temperance, but are confined to experimental or survey measures of risk aversion. Therefore, we start our empirical investigation of the relation between risk preferences and field behavior by only taking into account experimentally elicited risk aversion, thus following the standard approach in the literature (e.g., Galizzi, Machado, and Miniaci 2016; Charness et al. 2020; Samek et al. 2021).

In all regressions shown in Table 2, we control for a large series of possible influence factors (including age, cognitive abilities, gender, impatience, parents' education, religion, students' grades in school, or the amount of pocket money). In column 1 of this table, mimicking the standard approach in the literature, we

Table 2—Relevance of Risk Preferences in Predicting Field Behavior (*continued*)

Coefficient of risk preferences (columns show estimated coefficients)	Measures from Section II (joint estimation)			Survey measure
	Risk aversion (r, Arrow-Pratt)	Prudence (p, Crainich-Eeckhoudt)	Temperance (t, Denuit-Eeckhoudt)	Survey measure (risk tolerance)
	(6)	(7)	(8)	(9)
Panel A. Health-related behavior				
Unhealthy behavior	0.024	−0.140	−0.011	0.080
	(0.086)	(0.025)	(0.047)	(0.048)
Addictive behavior	0.007	−0.146	−0.002	0.084
	(0.087)	(0.025)	(0.047)	(0.048)
Smartphone addiction	−0.023	−0.160	0.017	0.080
	(0.086)	(0.025)	(0.045)	(0.052)
Panel B. Financial decision-making				
Saving (with debt)	0.096	0.058	0.087	−0.057
	(0.105)	(0.034)	(0.035)	(0.040)
Risky investment	−0.066	−0.045	−0.032	0.135
	(0.070)	(0.043)	(0.016)	(0.043)
Insurance demand	−0.018	−0.023	−0.001	0.044
	(0.044)	(0.024)	(0.046)	(0.026)
Panel C. Prevention and eco-friendly behavior				
Prevention (short term)	0.120	−0.119	0.017	0.173
	(0.063)	(0.033)	(0.039)	(0.028)
Prevention (long term)	0.167	0.010	0.078	−0.032
	(0.060)	(0.042)	(0.034)	(0.048)
Eco-friendly behavior	0.201	0.022	0.034	−0.013
	(0.081)	(0.027)	(0.057)	(0.033)
Panel D. Planning behavior and preference for competitive income				
Cautious planning	−0.063	0.028	0.170	−0.082
	(0.050)	(0.051)	(0.045)	(0.038)
Preference for competitive income	−0.070	0.054	0.032	0.039
	(0.023)	(0.024)	(0.047)	(0.027)

Notes: Positive coefficients imply increasing preference for the respective type of behavior. All measures in standard deviations. Column 1 shows the regression coefficients when only using the simple risk aversion measure (r^{CE}) as measure of risk preferences (see Section IIA). Column 2 shows coefficients when using a "counting measure" of prudent decisions in the behavioral classification tasks due to Eeckhoudt and Schlesinger (2006) instead (see "Behavioral Prudence Task" in Section IB). Column 3 shows coefficients when replacing the counting measure of prudence with our raw intensity measure p^{CE} (see Section IIA). Columns 4 and 5 show results of a regression using jointly the raw intensity measures of risk aversion r^{CE} and prudence p^{CE} (see Section IIA). Columns 6 to 8 show results of a regression using jointly the continuous utility-based intensity measures of risk aversion, prudence and temperance as resulting from our sophisticated method (see Section II). Column 9 shows regression coefficients when risk preferences are measured with the one-item survey question on willingness to take risk in general. In all regressions, we control for age, cognitive abilities, gender, impatience, and the influence factors listed in the notes of Table 1, except for Saving and Risky investment, where we additionally control for measures of income risk, as the theory that we rely on here is about precautionary saving and investment in presence of background risk (see the considerations on the empirical strategy in Supplemental Appendix E.2 for details). Unfortunately, in the first school, we had to work with printed questionnaires which did not include most of our health questions. For this reason, the number of observations for which we obtained information on health-related behavior is slightly lower ($N = 561$) than for Insurance demand ($N = 643$) and Preference for competitive income ($N = 649$), where some individuals skipped the corresponding questions, and for all the other domains ($N = 658$). See Supplemental Appendix Tables D-9 to D-19 for more detailed regression results and further notes. Standard errors in parentheses result from bootstrapping (1,000 rep.), and are clustered at the session level.

present the relation of the simple risk measure (r^{CE}) to the different domains of field behavior. Overall, the coefficients of the raw risk measure (r^{CE}) seem to suggest, at first sight, that risk preferences are only sporadically related to the domains of field

behavior we study. While there are some significant coefficients in panels C and D with respect to, for example, prevention, panels A and B on health-related behavior and financial decision-making even seem to suggest that risk preferences are not at all related to this kind of field behavior.

Yet, the picture looks already different in column 2, where we use the simplest counting measure of prudence building on the behavioral classification tasks by Eeckhoudt and Schlesinger (2006) as predictor for field behavior. We see that prudence is related to health-related behavior (see panel A), while for the other domains of field behavior (in panels B to D) we see no relationship. This is just a first indication that higher-order risk preferences should be considered when linking risk preferences to field behavior in order not to overlook important relations (like here for health-related behavior). This first indication is corroborated by looking at column 3, where we use our raw measure of prudence (p^{CE}) as the only measure of risk preferences. Contrary to the (discrete) measure in column 2, this continuous measure has a clear interpretation as intensity measure of prudence (even though it is only an approximation of the utility-based prudence measure p). We see the first relations from the count measure in column 2 confirmed, and additionally see a relation to saving, short-term prevention, and cautious planning.

In columns 4 and 5 of Table 2 we then present a next step to provide a more encompassing view of the relation of risk preferences to field behavior. Here we consider *both* the raw measure of risk aversion (r^{CE}) and our raw measure of prudence (p^{CE}) in the *same* regression. For risk aversion, adding the raw measure of prudence does not change a lot. It turns significant for short-term prevention, but loses significance for cautious planning. More importantly, we confirm in these columns that prudence has explanatory power beyond the domain of health-related behavior also for savings, short-term prevention, and cautious planning—independently of controlling for risk aversion.

In columns 6 to 8 of Table 2 we then take the final step to show the importance of considering higher-order risk preferences when relating risk preferences to field behavior. In these three columns, we present the coefficients when *jointly* considering our sophisticated utility-based measures of risk aversion, prudence and temperance. We see again a strong relation of prudence to health-related behavior, but again no such relation for risk aversion. Regarding financial decision-making, temperance is relevant for risky investment, as predicted, and both prudence and temperance are related to savings. Yet again, we see no relation of financial decision-making with risk aversion.[16] For panels C and D on prevention and eco-friendly behavior, planning and a preference for competitive income, higher-order risk preferences matter, too. We see a strong and significantly negative relationship between prudence and short-term prevention. As was the case for panels A and B above, from using a simple measure of risk aversion as in column 1, one would, however, have inferred that short-term prevention is unrelated to risk preferences. For cautious planning,

[16] Of course, our findings should be taken with a grain of salt as our adolescents may only have limited exposure to and experience in certain domains of financial decision-making. In this light, it might not be surprising that, e.g., temperance seems to be more important for saving than prudence, although theory posits an unambiguously positive relationship only for the latter (which we also find). Nevertheless, the fact that we find the theoretically expected relations already among adolescents despite potential limitations suggests that higher-order risk preferences (but not risk aversion) are robustly related to financial behavior.

the simple risk aversion measure from column 1 was significant (10 percent level), but the measure in column 6 is no longer significant when taking into account also higher-order risk preferences. In this case it turns out that temperance is more relevant (which also matters for long-term prevention). Lastly, regarding a preference for a competitive income, using all utility-based measures jointly confirms the relevance of risk aversion, but the additional finding of a significant effect of prudence shows a more nuanced picture.

To summarize the main insights from columns 1 to 8 in Table 2, we note that risk preferences matter for all domains of behavior that we study. Yet, especially for health-related behavior and financial decision-making, it is not risk aversion that is actually relevant, but prudence and temperance, in line with the theoretical models mentioned in Section IC. This observation mirrors the pattern emerging from our literature review in Supplemental Appendix C, where we survey papers that relate risk aversion to the domains of field behavior that we study in this paper: Experimental measures of risk aversion are rarely successful in establishing a relation to field behavior, in particular with respect to financial decision-making and addictive behavior. Our column 1 of Table 2 replicates these results, and our sophisticated utility-based measures of higher-order risk preferences in columns 6 to 8 offer an explanation for this pattern: Particularly in the domains of panels A and B, the higher-order risk preferences prudence and temperance matter, but not risk aversion.

In Supplemental Appendix B we provide a multitude of robustness checks for our main results in Table 2. Since we have seen so far that our sophisticated method increases predictive quality compared to the raw measures (but does not lead to different qualitative results), we focus in these checks on the sophisticated measures. In particular, we explore whether alternative explanations could rationalize the relationships that we find between our measures of higher-order risk preferences and field behavior. We examine potential biases in the estimates of (higher-order) risk preferences (with and without "noisy" data), as well as potential measurement error in risk aversion. We also check whether simple patterns of decision-making could have been misinterpreted as prudence or temperance. Moreover, we consider the potential impact of deviations from expected utility, including an analysis of the potential role of probability weighting and reference dependence. Finally, we discuss the consistency in and the identification of the utility curvature. In all of these considerations and robustness checks, our main results presented in this section persist.

V. Results III: On Survey Measures of Risk and Field Behavior

In our literature survey in Supplemental Appendix C, we show that survey measures of risk aversion are more often related to field behavior in a significant way than experimental measures of risk aversion are, in particular with respect to addictive behavior and financial decision-making. In this final results section, we offer an explanation for this observation. Recall from Section IB that we also elicited subjects' responses to the standard question on the willingness to take risk in general, taken from the German Socio-Economic Panel (SOEP). This survey measure is usually interpreted as capturing risk tolerance, or risk aversion when reverse coded (Wagner, Frick, and Schupp 2007; Dohmen et al. 2011; Falk et al. 2018). Our data supports

this: Using the raw and sophisticated measures for risk aversion (r and r^{CE}), the raw Pearson correlation coefficient with this survey measure is -0.27 ($p < 0.001$) in both cases (see Supplemental Appendix Figures D-19 and D-20). Thus, more risk aversion in our experimental measures means less risk tolerance in the SOEP-survey measure.

Yet, it is particularly noteworthy that also our measures of prudence and temperance are significantly related to this survey measure—the raw measure of prudence (p^{CE}) even to the same degree (coefficient of -0.27, $p < 0.001$, see Supplemental Appendix Figure D-20). For our adapted DOSPERT-survey measure (Weber, Blais, and Betz 2002, see Section IB), we find almost exactly the same pattern. Finally, in a regression of the SOEP-survey measure on our sophisticated utility-based measures (in the spirit of columns 6 to 8 of Table 2), we obtain (standardized) partial correlation coefficients of -0.44 for risk aversion, -0.20 for prudence (both $p < 0.001$), and -0.09 for temperance ($p < 0.05$). So, even when controlling for the other risk preference measures, the partial correlation coefficient of prudence is still about half as large and the coefficient of temperance about one-quarter as large as the one for risk aversion. This shows that general risk-taking behavior, as we understand it in everyday language (and as used for the survey measures), might only be insufficiently captured by risk aversion alone (as there is a contribution of the higher-order risk preferences in explaining the survey measure even when controlling for risk aversion).

The relation to prudence and temperance (beyond what is explained by risk aversion) may then be one of the reasons why survey measures of risk tolerance are often related to field behavior, while experimental measures of risk aversion (alone) are not. In Supplemental Appendix B.5, we illustrate formally (and confirm with simulations) how this can be rationalized, even though the correlations between the experimental measures (risk aversion and prudence) are larger than their correlation with the survey measure. In short, if the share of variation in the survey measure that cannot be explained by the experimental risk-aversion measure captures any of the dimensions that the higher-order risk measures are capturing (which is what the partial correlations tell us), we can expect to see the survey measure being a less precise, but more often predictive measure of risk-related field behavior compared to, for example, an experimental risk aversion measure that is restricted to only a single dimension of risk.

In fact, this is what we find in our sample. In our discussion of columns 1 to 8 of Table 2, we have already seen that the experimental measures of risk aversion completely fail to relate to health-related behavior and financial decision-making. Column 9 of Table 2 presents the relation between the SOEP-survey measure and our various domains of field behavior. Disregarding for the moment the low precision of some results that are significant only at the 10 percent level, the survey measure actually is more often successful in predicting field behavior (six significant estimations) than the simple measure of risk aversion (four significant relations in column 1 of Table 2). Thus, the survey measure seems to capture more dimensions of risk preferences than the experimental risk aversion measure does. Therefore, caution is warranted in interpreting the established relations with the survey measure.

In fact, looking at the six cases where the survey measure on risk tolerance is significantly related to field behavior, the simple measure of risk aversion (r^{CE}) fails to relate to field behavior in all but one case (cautious planning). Much in

contrast, already our simple measure of prudence (p^{CE}) successfully relates to four of these six field behaviors (health-domain, short-term prevention, and cautious planning; see columns 1, 3–5, and 9 of Table 2). Note that the survey measure assesses the willingness to tolerate risk, while the experimental measures are measures of risk aversion, prudence, and temperance, and hence, the predictions coincide in their direction, despite an opposite sign in Table 2. Next, we look at our set of sophisticated measures (that also includes a measure of temperance) in columns 6 to 8. Recall that theoretically, temperance relates to risky investment (one of the six behaviors that the survey measure relates significantly to). All four of the six cases where the significant relations of the simple prudence measure (p^{CE}) and the survey measure overlap are confirmed with the sophisticated measure p (although for cautious planning, it is no longer prudence that matters, but temperance). In addition, we see temperance (t) significantly relates to risky investment, and hence, in five of the six cases, where the survey measure relates significantly to field behavior, we see a sophisticated measure of one of the higher-order risk preferences pointing in the same direction with a significant coefficient. (For insurance demand, the sixth case where the survey measure is significantly predicting field behavior, we see a significant relation with prudence, too, but this is not robust to our full set of controls, see Supplemental Appendix Table D-14). With respect to risk aversion (r), in turn, once we control for higher-order risk preferences, only for short-term prevention we find that risk aversion and the survey measure are significantly related to field behavior (as cautious planning is no longer significantly predicted by risk aversion once we control for higher-order risk preferences, see column 4 of Table 2). Looking at the sign, however, we see that risk aversion is *positively* related to short-term planning, whereas from the survey measure, we infer the *opposite* direction. The significant coefficient of the prudence measure, however, is in line with that of the survey measure. Thus, it is rather prudence that is being captured by the survey measure in this case. In sum, we see strong support for survey measures capturing higher-order risk preferences in our data (which we see confirmed with the DOSPERT-survey measure).

Lastly, controlling for higher-order risk preferences in the regression shown in column 9 further supports this interpretation (see Supplemental Appendix Table D-22): Once we control for prudence and temperance, the survey measure is no longer significant in predicting health-related behavior as well as insurance demand. In the case of short-term planning, the coefficient of the survey measure decreases, with a considerably lower t-statistic, but continues to be a significant predictor. Similarly, for cautious planning, the coefficient of the survey measure decreases and is estimated with a lower precision once we add higher-order risk preferences. Here, the significance level changes from 5 percent to 10 percent. Lastly, also for risky investment the coefficient and t-statistic of the survey measure change notably.

To sum up, interpreting survey measures of risk tolerance as capturing the influence of risk aversion is problematic, and can be as misleading as focusing on experimental measures of risk aversion alone.

These observations provide a novel perspective on the inconsistent findings on risk preferences and their relation to field behavior mentioned in the introduction: While all the mentioned failures rely on experimental measures of risk aversion and neglect higher-order risk preferences (Sutter et al. 2013; Galizzi, Machado and

Miniaci 2016; Charness et al. 2020; Samek et al. 2021), all of the mentioned successes (except for Anderson and Mellor 2008) use the survey measure that is obviously capable of capturing additional dimensions of risk preferences on top of risk aversion alone. Our extensive literature review on papers relating measures of "risk aversion" to field behavior in Supplemental Appendix C confirms this pattern at a larger scale. The survey measure of risk tolerance is more often significantly related to field behavior than are experimental measures of risk aversion—especially with respect to health-related behavior and financial decision-making. We replicate this in our data—and show that accounting for higher-order risk preferences makes the difference.

VI. Conclusion

In this paper, we have used two novel methods to measure (higher-order) risk preferences: We complement a sophisticated method with a simple approach to obtain two sets of utility-based intensity measures of (higher-order) risk preferences. The sophisticated method goes back to Schneider, Ibanez, and Riener (2025) and has been extended here. Our simple approach to measuring utility-based prudence has first been introduced in this paper. In our sample of 658 sixth- to twelfth-grade students in German schools, we have found clear evidence for risk aversion, prudence, and temperance in the aggregate. These findings are in line with studies on adult populations (e.g., Deck and Schlesinger 2014; Ebert and Wiesen 2014; Noussair, Trautmann, and van de Kuilen 2014). We have found no significant age effects for any of our preferences. Females exhibit more risk-averse, more prudent, and more temperate behavior. Cognitive abilities are unrelated to prudence and negatively related to risk aversion and temperance.

The most important finding of our paper concerns the relationship of measures of risk preferences and field behavior. Based on several theoretical models, prudence was predicted to matter for health-related behavior, financial decision-making, eco-friendly behavior, prevention behavior, and planning. This is what we actually find in most cases. Temperance was expected to be related to risky investment, and so it is. Risk aversion, however, is rarely related to these domains of field behavior. While one could conclude from the latter finding that risk preferences do not matter for field behavior—as is often the conclusion when experimental measures of risk aversion are related to field behavior—such a conclusion would be misleading, as risk comes in different forms, and prudence and temperance do clearly matter in most domains. Neglecting higher-order risk preferences thus bears the risk of drawing premature conclusions about the (seeming) irrelevance of risk preferences for field behavior.

Arguably, the second-most important finding relates to the observation that the widely used survey question on the willingness to take risks (Dohmen et al. 2011; Falk et al. 2018) is often significantly related to field behavior. We observe the same in our data and can show that this is because the survey measure is not only related to risk aversion, but also to prudence and temperance. This suggests that risk tolerance, as used in everyday language, is only insufficiently captured by measures of risk aversion alone, extending related results by Jaspersen, Ragin, and Sydnor (2020) and Chapman et al. (2021). Rather, other dimensions matter too, such as prudence

and temperance. This helps understanding why survey measures are often very successful in relating to field behavior and adds another piece of evidence that considering higher-order risk preferences is crucial for understanding the relationship between risk preferences and field behavior.

REFERENCES

Abdellaoui, Mohammed, Han Bleichrodt, and Corina Paraschiv. 2007. "Loss Aversion Under Prospect Theory: A Parameter-Free Measurement." *Management Science* 53 (10): 1659–74.

Anderson, Lisa R., and Jennifer M. Mellor. 2008. "Predicting Health Behaviors with an Experimental Measure of Risk Preference." *Journal of Health Economics* 27 (5): 1260–74.

Andreassen, Cecilie S. 2015. "Online Social Network Site Addiction: A Comprehensive Review." *Current Addiction Reports* 2 (2): 175–84.

Bardsley, Nicholas, Robin Cubitt, Graham Loomes, Peter Moffatt, Chris Starmer, and Robert Sugden. 2010. *Experimental Economics: Rethinking the Rules.* Princeton University Press.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1): 453–78.

Bleichrodt, Han, David Crainich, and Louis Eeckhoudt. 2003. "The Effect of Comorbidities on Treatment Decisions." *Journal of Health Economics* 22 (5): 805–20.

Bramoullé, Yann, and Nicolas Treich. 2009. "Can Uncertainty Alleviate the Commons Problem?" *Journal of the European Economic Association* 7 (5): 1042–67.

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky. 2008. "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty* 37 (2-3): 237–69.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2021. "On the Relation Between Willingness to Accept and Willingness to Pay." Unpublished.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2023. "Econographics." *Journal of Political Economy Microeconomics* 1 (1): 115–61.

Charness, Gary, Thomas Garcia, Theo Offerman, and Marie C. Villeval. 2020. "Do Measures of Risk Attitude in the Laboratory Predict Behavior Under Risk in and Outside of the Laboratory?" *Journal of Risk and Uncertainty* 60 (2): 99–123.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree – An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97.

Clarke, Damian, Joseph P. Romano, and Michael Wolf. 2020. "The Romano-Wolf Multiple-Hypothesis Correction in Stata." *Stata Journal* 20 (4): 812–43.

Courbage, Christophe, and Beatrice Rey. 2006. "Prudence and Optimal Prevention for Health Risks." *Health Economics* 15 (12): 1323–27.

Crainich, David, and Louis Eeckhoudt. 2008. "On the Intensity of Downside Risk Aversion." *Journal of Risk and Uncertainty* 36 (3): 267–76.

Crosetto, Paolo. 2025. "METARET: A Meta Analysis of the external validity of Risk Elicitation Tasks. OSF doi.org/10.17605/OSF.IO/H2Z56 (accessed on August 7, 2025).

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47 (2): 448–74.

Dardanoni, Valentino, and Adam Wagstaff. 1990. "Uncertainty and the Demand for Medical Care." *Journal of Health Economics* 9 (1): 23–38.

Deck, Cary, and Harris Schlesinger. 2014. "Consistency of Higher Order Risk Preferences." *Econometrica* 82 (5): 1913–43.

Denuit, Michael M., and Louis Eeckhoudt. 2010. "Stronger Measures of Higher-Order Risk Attitudes." *Journal of Economic Theory* 145 (5): 2027–36.

DeWit, David J., Edward M. Adlaf, David R. Offord, and Alan C. Ogborne. 2000. "Age at First Alcohol Use: A Risk Factor for the Development of Alcohol Disorders." *American Journal of Psychiatry* 157 (5): 745–50.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100 (3): 1238–60.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jurgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–50.

Durbin, J. 1954. "Errors in Variables." Revue de l'Institut International de Statistique/Review of the International Statistical Institute 22 (1/3): 23–32.
Ebert, Sebastian. 2013. "Moment Characterization of Higher-Order Risk Preferences." *Theory and Decision* 74 (2): 267–84.
Ebert, Sebastian, and Daniel Wiesen. 2014. "Joint Measurement of Risk Aversion, Prudence, and Temperance." *Journal of Risk and Uncertainty* 48 (3): 231–52.
Eeckhoudt, Louis, and Christian Gollier. 2005. "The Impact of Prudence on Optimal Prevention." *Economic Theory* 26 (4): 989–94.
Eeckhoudt, Louis, Béatrice Rey, and Harris Schlesinger. 2007. "A Good Sign for Multivariate Risk Taking." *Management Science* 53 (1): 117–24.
Eeckhoudt, Louis, and Harris Schlesinger. 2006. "Putting Risk in its Proper Place." *American Economic Review* 96 (1): 280–89.
Ehrlich, Isaac, and Gary S. Becker. 1972. "Market Insurance, Self-Insurance, and Self-Protection." *Journal of Political Economy* 80 (4): 623–48.
Eilers, Paul H., and Brian D. Marx. 1996. "Flexible Smoothing with B-splines and Penalties." *Statistical Science* 11 (2): 89–102.
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): 1645–92.
Familien-wegweiser.de. 2018. *Taschengeld.* Familien-wegweiser.de. http://web.archive.org/web/20180605192037/http://www.familien-wegweiser.de/wegweiser/stichwortverzeichnis%2Cdid%3D38294.html (accessed June 11, 2018).
Fei, Wenan, and Harris Schlesinger. 2008. "Precautionary Insurance Demand with State-Dependent Background Risk." *Journal of Risk and Insurance* 75 (1): 1–16.
Frey, Renato, Andreas Pedroni, Rui Mata, Jörg Rieskamp, and Ralph Hertwig. 2017. "Risk Preference Shares the Psychometric Structure of Major Psychological Traits." *Science Advances* 3 (10): 1–13.
Fuchs-Schündeln, Nicola, and Matthias Schündeln. 2005. "Precautionary Savings and Self-Selection: Evidence from the German Reunification "Experiment"." *Quarterly Journal of Economics* 120 (3): 1085–1120.
Galizzi, Matteo M., Sara R. Machado, and Raffaele Miniaci. 2016. "Temporal Stability, Cross-Validity, and External Validity of Risk Preferences Measures: Experimental Evidence from a UK Representative Sample." Unpublished.
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): 1826–63.
Gollier, Christian, and John W. Pratt. 1996. "Risk Vulnerability and the Tempering Effect of Background Risk." *Econometrica* 64 (5): 1109–23.
Gruber, Marvin H. J. 2017. *Improving Efficiency by Shrinkage*. Routledge.
Harrison, Glenn W. 1986. "An Experimental Test for Risk Aversion." *Economics Letters* 21 (1): 7–11.
He, Qinghua, Ofir Turel, and Antoine Bechara. 2017. "Brain Anatomy Alterations Associated with Social Networking Site (SNS) Addiction." *Scientific Reports* 7: 45064.
Heinrich, Timo, and Jason Shachat. 2020. "The Development of Risk Aversion and Prudence in Chinese Children and Adolescents." *Journal of Risk and Uncertainty* 61 (3): 263–87.
Hoerl, Arthur E., and Robert W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1): 55–67.
Hormes, Julia M., Brianna Kearns, and C. Alix Timko. 2014. "Craving Facebook? Behavioral Addiction to Online Social Networking and its Association with Emotion Regulation Deficits." *Addiction* 109 (12): 2079–88.
Jaspersen, Johannes G., Marc A. Ragin, and Justin R. Sydnor. 2020. "Linking Subjective and incentivized Risk Attitudes: The Importance of Losses." Journal of Risk and Uncertainty 60 (2): 187–206.
Kimball, Miles S. 1990. "Precautionary Saving in the Small and in the Large." *Econometrica* 58 (1): 53–73.
Kimball, Miles S. 1992. New Palgrave Dictionary of Money and Finance, Chapter Precautionary Motives for Holding Assets, pp. 158–161. MacMillan, London.
Kimball, Miles S. 1993. "Standard Risk Aversion." *Econometrica* 61 (3): 589–611.
Leland, H. E. 1968. "Saving and Uncertainty: The Precautionary Demand for Saving." *Quarterly Journal of Economics* 82 (3): 465–73.
Maier, J., and M. Rüger. 2011. "Experimental Evidence on Higher-Order Risk Preferences with Real Monetary Losses." Unpublished.
Menegatti, Mario. 2009. "Optimal Prevention and Prudence in a Two-Period Model." *Mathematical Social Sciences* 58 (3): 393–97.

Menezes, C., C. Geiss, and J. Tressler. 1980. "Increasing Downside Risk." *American Economic Review* 70 (5): 921–32.

Modica, Salvatore, and Marco Scarsini. 2005. "A Note on Comparative Downside Risk Aversion." *Journal of Economic Theory* 122 (2): 267–71.

Noussair, Charles N., Stefan T. Trautmann, and Gijs van de Kuilen. 2014. "Higher Order Risk Attitudes, Demographics, and Financial Decisions." *Review of Economic Studies* 81 (1): 325–55.

Orben, Amy, and Andrew K. Przybylski. 2019. "The Association Between Adolescent Well-Being and Digital Technology Use." *Nature Human Behaviour* 3 (2): 173–82.

Paul, Seana L., Leigh Blizzard, George C. Patton, Terry Dwyer, and Alison Venn. 2008. "Parental Smoking and Smoking Experimentation in Childhood Increase the Risk of Being a Smoker 20 Years Later: The Childhood Determinants of Adult Health Study." *Addiction* 103 (5): 846–53.

Pratt, J. W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica* 32 (1/2): 122–36.

Przybylski, Andrew K., and Netta Weinstein. 2017. "A Large-Scale Test of the Goldilocks Hypothesis: Quantifying the Relations between Digital-Screen Use and the Mental Well-Being of Adolescents." *Psychological Science* 28 (2): 204–15.

Romano, Joseph P. and Michael Wolf. 2005a. "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing." *Journal of the American Statistical Association* 100 (469): 94–108.

Romano, Joseph P., and Michael Wolf. 2005b. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73 (4): 1237–82.

Romano, Joseph P., and Michael Wolf. 2016. "Efficient Computation of Adjusted P-Values for Resampling-Based Stepdown Multiple Testing." *Statistics & Probability Letters* 113: 38–40.

Rubin, Donald B. 1996. "Multiple Imputation after 18+ Years." *Journal of the American Statistical Association* 91 (434): 473–89.

Runge, C. 1901. "Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten." *Zeitschrift für Mathematik und Physik* 46: 224–43.

Samek, Anya, Andre Gray, Ashlesha Datar, and Nancy Nicosia. 2021. "Adolescent Time and Risk Preferences: Measurement, Determinants and Field Consequences." *Journal of Economic Behavior & Organization* 184: 460–88.

Schneider, Sebastian O. 2025a. "oTree: Ready-Made Apps for Higher-Order Risk Preference Elicitation Methods." Unpublished.

Schneider, Sebastian O. 2025b. *utilityFunctionTools: P-spline regression for utility functions and derived measures. R package,* available via CRAN: https://cran.r-project.org/package=utilityFunctionTools. Package version 1.0.0.

Schneider, Sebastian O., Marcela Ibanez, and Gerhard Riener. 2025. "Measuring Utility: An Application to Higher-Order Risk." Unpublished.

Schneider, Sebastian O., and Matthias Sutter. 2026. *Data and Code for: "Risk Preferences and Field Behavior: The Relevance of Higher-Order Risk Preferences."* American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E237164V1.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rützler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents Field Behavior." *American Economic Review* 103 (1): 510–31.

Sutter, Matthias, Claudia Zoller, and Daniela Glätzle-Rützler. 2019. "Economic Behavior of Children and Adolescents – A First Survey of Experimental Economics Results." *European Economic Review* 111: 98–121.

Trautmann, Stefan T., and Gjis van de Kuilen. 2018. "Higher Order Risk Attitudes: A Review of Experimental Evidence." *European Economic Review* 103: 108–24.

Wagner, G. G., J. R. Frick, and J. Schupp. 2007. "The German Socio-Economic Panel Study (SOEP) – Evolution, Scope and Enhancements." *Journal of Contextual Economics – Schmollers Jahrbuch* 127 (1): 139–169.

Weber, Elke U., Ann-Renée Blais, and Nancy E. Betz. 2002. "A Domain-Specific Risk-Attitude Scale: Measuring Risk Perceptions and Risk Behaviors." *Journal of Behavioral Decision Making* 15 (4): 263–90.