ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Danz, David

Conference Paper

The curse of knowledge increases self-selection into competition: Experimental evidence

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Behavioral Economics, No. D10-V2

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Danz, David (2014) : The curse of knowledge increases self-selection into competition: Experimental evidence, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Behavioral Economics, No. D10-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:
https://hdl.handle.net/10419/100543

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

The curse of knowledge increases self-selection into competition: Experimental evidence*

David Danz†

September 2014‡

Abstract

The psychology literature provides ample evidence that people have difficulties taking the perspective of less informed others. This paper presents a controlled experiment showing that this "curse of knowledge" can cause comparative overconfidence and overentry into competition. In a broader context, the results provide an explanation for the overconfidence of nascent entrepreneurs and the substantial rate of failure among new businesses.

Keywords: curse of knowledge, hindsight bias, information projection, overconfidence, sorting, incentive schemes, competition, beliefs, experiments

JEL classification numbers: C91, D80, D82, D83, D84

*I am grateful for helpful comments by Marie-Pierre Dargnies, Dietmar Fehr, Jana Friedrichsen, Holger Gerhardt, Rustamdjan Hakimov, Frank Heinemann, Steffen Huck, Radosveta Ivanova-Stenzel, Dorothea Kübler, Morimitsu Kurino, Kristóf Madarász, Lydia Mechtenberg, Muriel Niederle, Julia Schmid, Roel van Veldhuizen, Stephanie Wang, Georg Weizsäcker, and seminar participants at HWR Berlin, WZB-TU Colloquium, BBE Workshop 2013, CRC 649 Workshop 2013, IMEBESS 2014, and CIRPÉE-TU-WZB Workshop 2014. I further thank Jennifer Rontganger for proofreading and Ronald A. Rensink for generously providing change-blindness tasks. Financial support from the Deutsche Forschungsgemeinschaft (DFG) through SFB 649 "Economic Risk" is gratefully acknowledged.

†WZB Berlin Social Science Center, 10785 Berlin, Germany and Technische Universität Berlin, Department of Microeconomics. E-mail: david.danz@wzb.eu.

‡A previous version of this paper appeared in my dissertation (Danz 2013).

1 Introduction

The empirical literature documents a substantial rate of failure among new businesses. Merely half of new businesses still exist after four years, and about two-thirds of closed businesses are explicitly considered failures by the entrepreneurs (Headd 2003). Furthermore, in most cases the income from self-employment is below earnings from dependent employment (Hamilton 2000). At the same time, there is ample evidence of overconfidence among entrepreneurs (Cooper et al. 1988; Wu and Knott 2006).[1] Moreover, the self-reported confidence of nascent entrepreneurs and objective measures for the expected survival of their businesses are negatively correlated (Koellinger et al. 2007) and, regarding ex post explanations of business failure, variables associated with a lack of managerial skills are among the most salient factors (Patterson et al. 1983; Gaskill et al. 1993). As new business entries constitute a considerable fraction of economic activity,[2] efficiency losses through suboptimal entry decisions are deemed to be high.

This paper investigates whether the *curse of knowledge* can explain overconfident self-selection into competitive environments. The curse of knowledge refers to a host of psychological evidence showing that people have difficulties ignoring their private information when trying to take a less-informed perspective. This "less-informed perspective" may refer to one's own past self (*hindsight bias*) or the perspective of someone else.[3] In the seminal experiments by Fischhoff (1975), subjects had to judge the likelihood of potential outcomes of historical events in light of the information that was available *before* those events. Providing subjects with the "true"[4] outcome increased their assessments of the *ex-ante* likelihood of that outcome—both when stating their own ex-ante likelihood assessments and

[1] High failure rates might be expected if entrepreneurs exhibited a comparatively high willingness to take risks as envisaged in the classical theoretical literature (Knight 1921). However, the empirical literature does not provide evidence for this conjecture. In fact, entrepreneurs do not seem to differ from wage earners in their willingness to take risks (see Wu and Knott 2006 for a review).

[2] For example, on an annual basis, in the US 10%–12% of all firms are new entrants (see Dennis 1997).

[3] Like Loewenstein et al. (2006) and Birch and Bloom (2007), I consider the hindsight bias to be "a within-person version of the curse of knowledge." In contrast to the hindsight bias in the narrow sense, the broader notion of the curse of knowledge seems also more consistent with the seminal study by Fischhoff (1975), who employed between-subject designs in his experiments, where subjects, among others, "were asked to respond as other student judges who had not known the true outcome." See also Biais and Weber (2009) for a review of associated experimental designs.

[4] The effect was found irrespective of which of the potential outcomes was labeled as true.

when estimating past judgments of others who did not know the outcome. Thus, knowing the outcome of a historical event "produces an unjustified increase in its perceived predictability, for it seems to have appeared more likely than it actually was" (Fischhoff 1975).

Following Fischhoff (1975) and Fischhoff and Beyth (1975), the curse of knowledge has been shown to manifest itself in a wide range of contexts such as the predictability of political events (Bryant and Guilbault 2002), the predictability of experimental results (Slovic and Fischhoff 1977), the evaluation of entrepreneurial success (Bukszar and Connolly 1988), the comprehensibility of messages in communication (Newton 1990; Keysar and Henly 2002; Kruger et al. 2005), and diagnoses by practicing physicians (Arkes et al. 1981). Experimental efforts to moderate the curse of knowledge, e.g., via feedback or even by raising subjects' awareness of the bias have been—if anything—of limited success (see Christensen-Szalanski and Willham 1991, and Guilbault et al. 2004 for meta-analyses).

How can the curse of knowledge explain comparative overconfidence and excessive business entry? Nascent entrepreneurship is certainly characterized by ability uncertainty in many dimensions, and feedback regarding one's entrepreneurial success is presumably rare. When a nascent entrepreneur assesses his ability relative to others, he might evaluate the past decisions of incumbents, or ponder how he would have decided in their stead. Here informational asymmetries arise rather naturally. For example, a nascent entrepreneur knows the demand for previously released products and services. He knows whether past investments turned out to be profitable or not. He knows whether a hiring decision was "wise." This kind of information, however, (i) was not available to the established entrepreneur when making her decisions and (ii) will not be available to the nascent entrepreneur once he faces similar decision making problems.

An agent who exhibits the curse of knowledge cannot disregard such superior information when considering the established entrepreneurs at the time of their decision making or when picturing himself in the same situation. In turn, he underestimates the uncertainty the established entrepreneurs faced when making their decisions, and hence overestimates the ex-ante probability of profitable managerial decision making and entrepreneurial success. Such biased judgment of a task's degree of difficulty (see, e.g., Loewenstein et al. 2006) can lead to biased comparative ability judgments: When the biased agent observes the *actual* performance of his

competitors, their performance seems "unexpectedly low." In turn, the competitors' ability—as inferred from their past decisions (or performance)—is underestimated on average (see Madarász 2012). Consequently, the biased agent exhibits overconfident comparative ability judgments once he observes the performance of others.

Of course, such biased ability judgments can cause suboptimal choices in the context of self-selection where productivity matters. In particular, when agents decide whether to enter competitive environments, biased agents will enter too often. Moreover, provided that the curse of knowledge affects an individual's inference about her own (absolute) ability, the effect on willingness to compete is expected to be more pronounced for individuals with comparatively low skills.

I test these predictions in a laboratory experiment where subjects choose between a tournament and a piece rate compensation for their performance in a real effort task. At the outset of the experiment, subjects receive sample tasks, knowing that their designated competitors worked on these tasks under tournament conditions. Subjects learn how many of the sample tasks were solved by each of their designated competitors. Subsequently, subjects' willingness to compete is elicited. The two main treatments differ with respect to the information subjects receive about the sample tasks. In the *informed treatment*, subjects receive the solutions to the sample tasks, which allows the curse of knowledge to affect beliefs and choices of the subjects. In the *uninformed (control) treatment*, no solutions are provided.

The main findings are the following. First, subjects in the informed treatment significantly overestimate their performance on average—both in absolute terms and relative to their competitors. No such overconfidence is observed in the uninformed treatment. Second, in line with these observations, subjects in the informed treatment exhibit a significantly higher willingness to compete than subjects in the uninformed treatment. Third, the treatment effect on tournament entry is primarily driven by subjects with comparatively low performance levels who should not enter the tournament.

Although the curse of knowledge is well documented in the psychology literature,[5] only a few studies have investigated the economic consequences thereof. The present paper adds to the experimental branch of this literature by establishing

[5] "One of the most widely studied biases in the judgment literature is the hindsight bias" (Rabin 1998).

a link between the curse of knowledge and excessive self-selection into competition. The experimental study on the curse of knowledge closest to this paper was conducted by Loewenstein et al. (2006). In their experiments, subjects were paid according to the accuracy of their predictions of the performance of others. In one of their experiments, a substantial fraction of the subjects paid for information—the solutions to the tasks—which biased their predictions and hence reduced their average earnings. Camerer et al. (1989) demonstrate the persistence of the curse of knowledge in competitive markets, though market forces alleviated the bias by roughly 50%. Biais and Weber (2009) show, theoretically and experimentally, that the curse of knowledge can prevent traders from learning about risks. When investors are hindsight-biased they are less surprised by unexpectedly large positive or negative returns (since they "knew it all along") and hence they adjust their volatility estimates too conservatively. Finally, it has been shown that the hindsight bias can lead to suboptimal delegation decisions (Danz et al. 2013).

The paper also contributes to the literature investigating the sources of overconfidence. Following Moore and Healy (2008) I refer to overly optimistic beliefs about one's absolute performance level as *overestimation*, and to overly optimistic beliefs about one's performance relative to others as *overplacement*.[6] Niederle and Vesterlund (2007) show that men enter into tournaments more often than women. Men exhibit more pronounced overplacement as well as stronger preferences for competition compared to women, and, compared to payoff-maximizing choices, men with low productivity enter too often, while highly productive women enter too little. Dohmen and Falk (2011) confirm that variable payment schemes (especially tournaments) are chosen more often by men. Camerer and Lovallo (1999) find excess entry into competitive markets with capacity constraints if there is scope for overplacement, i.e., when payments are determined by skill rather than chance. Similarly, Alicke (1985) and Moore and Cain (2007) show that the controllability of a task affects relative self-assessments. Camerer and Lovallo (1999) further show that the failure to recognize the composition of one's competitors due to self-selection (*reference-group neglect*) can increase excessive market entry.[7]

[6] The third notion of overconfidence where agents are overly confident in their beliefs (i.e., underestimate their confidence intervals) is referred to as *overprecision* by Moore and Healy (2008).

[7] Further determinants of overconfidence that have been found in more general contexts include individual traits, such as genes and social background (Cesarini et al. 2009), and psychological traits (Schaefer et al. 2004), the degree of difficulty of the task (Moore and Cain 2007),

This paper adds to this literature by showing that the curse of knowledge can cause comparative overconfidence.

The paper proceeds as follows. The next section presents the experimental design. Section 3 reports the results, section 4 concludes.

2 Experimental design

The experiment was designed to test the effect of the curse of knowledge on willingness to compete. The experiment consisted of three treatments. In the first sessions, subjects participated in the *reference treatment*, where they performed real effort tasks in tournaments. These subjects served as "competitors" for the two main treatments: In the *informed treatment* and the *uninformed treatment*, subjects decided whether to compete against subjects from the reference treatment or not. The two main treatments were designed such that the curse of knowledge could occur in the informed treatment, but not in the uninformed (control) treatment.

2.1 Experimental task

Subjects in all treatments worked on change-detection tasks, where they had to spot the difference between two nearly identical images (see Figure 1 for an example).[8] Each task was presented in a short video clip of 14 seconds, where the two images were displayed alternately with short interruptions.[9] Afterwards, subjects had 40 seconds to submit an answer. To this end, the image containing the object of interest was displayed together with a grid of 70 fields, where subjects could enter one of the grid numbers.[10]

This type of task has been used frequently to study how humans process visual information in the context of *change blindness*, which refers to people having difficulties in detecting changes to an object or a scene (Rensink et al. 1997;

and the availability and quality of feedback (Nisbett and Ross 1980).

[8] A part of the tasks utilized in this experiment were generously provided by Rensink et al. (1997). I adopted the tasks in their "marginal interest" condition excluding those where colors of objects changed. The remaining tasks designed for this study share the characteristics of these tasks. Photos were taken from private libraries and http://www.1a-photoshop.de/.

[9] Each image was displayed for one second followed by a blank screen of 150 ms. The presentation of the task is similar to the extended flicker design by Rensink et al. (1997) with longer display times as in Loewenstein et al. (2006).

[10] See the experimental instructions in Appendix C for an example.

Figure 1. Example of a task used in the experiment. Image A (containing the object of interest) and A' (without the object) were presented in order $A, A', A, A', \ldots$.

Simons and Levin 1997). As with many real-life problems, the solution to such tasks seems obvious once it is known. In this respect the task was previously shown to induce a curse of knowledge (see Loewenstein et al. 2006).

2.2 Reference treatment

In the reference treatment, subjects performed 60 tasks that had been assigned to six rounds of 10 tasks.[11] At the beginning of each round, subjects were randomly matched in groups of four. They were informed that in each round their rank relative to their competitors would be determined based on their performance (ties were broken randomly). No feedback was provided until the end of the experiment. At the end of the experiment, two of the six rounds were randomly selected for payment. Subjects earned €1 for each correctly solved task if they ranked first, and zero otherwise. In addition to the payments from the tournaments, subjects received a €5 show-up fee and a fixed payment of €7 for working on the tasks.

2.3 Main treatments

Of the 60 tasks from the reference treatment, 40 were selected for the main treatments. The selection was conducted to achieve an overall intermediate degree of difficulty and to exclude tasks with an extremely low or high degree of difficulty. The order of the tasks in the two main treatments was varied between sessions to

[11] Additionally, six sample tasks were given at the outset of the experiment to familiarize subjects with the task.

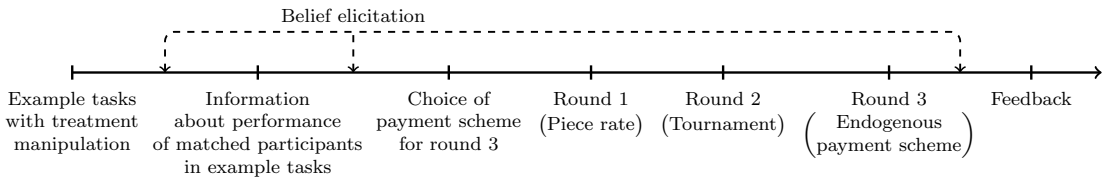

Figure 2. Timeline of the informed and the uninformed treatment.

assess whether subjects are calibrated correctly on average.[12] The timing of the two main treatments is illustrated in Figure 2.

2.3.1 Matching to reference treatment

At the outset of the sessions in the two main treatments, each participant was matched to three subjects from the reference treatment (their *competitors*). Participants were informed that this assignment would remain the same for the entire experiment, and that all tasks of the experiment—including those in the example stage—had been performed by their competitors in tournaments.

2.3.2 Example stage

The subjects first participated in an example stage where they observed 10 tasks in the same way as in the subsequent payoff-relevant rounds: After each image sequence of 14 seconds, subjects had at most 40 seconds to submit an answer. As in the payoff-relevant rounds, subjects did not receive feedback on their answers. The example stage served the purpose of familiarizing subjects with the task, enabling them to form (or update) beliefs about their productivity (and that of others), and introducing the treatment manipulation. The beginning of each task was announced by a countdown. During this countdown, participants in the informed treatment were given the solution of the task by showing them the image containing the object of interest circled in red to indicate its location (see Figure A.1 in the appendix). In the uninformed treatment, no solution was provided. This manipulation constitutes the only difference between the two main treatments.

[12] For more details see section A.2 in the Appendix.

2.3.3 Information about competitors

After the example stage, subjects received information about the past performance of their competitors. Specifically, they were told how many of the 10 sample tasks had been solved by each of their competitors. At this time, subjects were reminded that their example tasks were performed by their competitors in tournaments.

In the experiment, three *types of matching* were employed. The observed performance of the three competitors was $(4, 5, 5)$ for the first type of matching, $(4, 5, 6)$ for the second type of matching, and $(5, 5, 6)$ for the third type of matching.[13] For example, with the second type of matching, subjects in the main treatments observed that 4, 5, and 6 of their 10 example tasks had been solved by their three competitors, respectively.[14] Each type of matching was equally represented within each session. The matching was controlled for several reasons. First, controlling the patterns of observed performances assures the comparability of the two main treatments.[15] Second, the main interest was in sufficient variation of sorting choices *conditional* on the observed performance of the competitors. This variation was expected to be highest for "representative" performances of competitors.[16] Third, choosing values for the observed performance that occur frequently in the reference treatment, improves the reliability of the analysis of the optimality

[13] The matchings were determined as follows. Let (y_{01}, y_{02}, y_{03}) denote the observed performances for a given type of matching. For each session in the two main treatments, the task order was predetermined (see Table A.1). To form a matching group in a particular session, three subjects from the reference treatment were randomly selected, conditional on having solved y_{01}, y_{02}, and y_{03} out of the 10 tasks that were assigned to the example stage, respectively. For a better comparability of ranks between treatment (and gender), the random draws were used repeatedly: for each subject in the uninformed treatment, another subject in the informed treatment was matched to the same participants of the reference treatment. The same replication was conducted with respect to gender within each session. That is, within each pair of sessions, a particular triad from the reference treatment was assigned to one subject of each treatment-gender subset, which makes the actual rankings comparable between treatment and gender up to randomly broken ties. In this way, for every pair of sessions in the main treatments, 27 participants in the reference treatment were randomly selected conditional on their performance in the example tasks.

[14] The values were displayed in random order within treatment-gender groups.

[15] Besides experimental control per se, a further advantage of matching participants to past performances is that the participants' sorting choices do not affect the payoffs of their competitors. Thus, other-regarding preferences can be excluded as a determinant for subjects' sorting choices.

[16] The average performance per round in the reference treatment as well as the two main treatments is not significantly different from the average observed performance of the competitors (per design in the reference treatment; for the two main treatments, a t-test against 5 yields $p = 0.647$).

of sorting choices.[17] The slight variation in the patterns of observed performances was implemented to reduce the likelihood of the results being driven by a specific type of matching.

2.3.4 Sorting choice and payoff relevant rounds

In the two main treatments, subjects were informed that they would perform the tasks in three rounds of 10 tasks each, and that one of the three rounds would be randomly selected for payment at the end of the experiment. The subjects were informed that the payment schemes in rounds 1 and 2 were exogenously imposed. In round 1, the payment scheme was a piece rate, i.e., the payoffs were independent of the performance of their competitors. In round 2, the payment scheme was a tournament, where the subjects competed against their competitors. In round 3, the payment scheme was either a piece rate (as in round 1) or a tournament (as in round 2), depending on subjects' sorting choice (see Niederle and Vesterlund 2007).

The sorting choice was made after subjects learned about the past performance of their competitors but before they performed the tasks in an incentivized way. Regarding the payments in the tournament condition, subjects knew that if a corresponding round was selected for payment they would earn €2 for each correctly solved task if they ranked first relative to their competitors, and zero otherwise. The piece rate, however, was not known to the subjects at the time of their sorting choice. Subjects were informed that the piece rate would be publicly announced after each participant submitted his/her sorting choice. The instructions made clear that the piece rate had been determined before the experiment and stowed in a sealed envelope at the front door. This procedure ought to have prevented subjects from speculating that their sorting choice might affect the piece rate (see Bohnet and Zeckhauser 2004).

Regarding the sorting choice, subjects' *minimum acceptable piece rates* (MAP) were elicited, using a variation of the method by Bohnet and Zeckhauser (2004). Subjects were instructed that if round 3 was selected for payment, their compens-

[17] In this respect, consider the empirical probability that a subject from the reference treatment solves, say, seven tasks in the tournament stage, given that he/she solved y of the 10 example tasks. Choosing values of y that are observed frequently in the reference treatment, gives more reliable estimates of such conditional performance distributions for other rounds (see Table B.2 in the appendix).

ation scheme would depend on their stated MAP and the piece rate as follows. Should the piece rate in the envelope be greater than or equal to their stated MAP, they would be paid according to the piece rate. Conversely, if the piece rate was (strictly) below their stated MAP, they would be paid according to the tournament scheme.[18]

As pointed out by Bohnet et al. (2008), this variation of the Becker-DeGroot-Marschak mechanism is incentive-compatible with truth-telling being strictly dominant as long as subjects assign positive subjective probabilities to piece rates in the neighborhood of their MAP. By means of this elicitation procedure, I aimed at (i) more informative sorting choices compared to a setting where subjects make binary choices between payment schemes, and (ii) retaining control over the piece rate in contrast to the case where the piece rate is drawn from a known distribution.

The piece rates that were implemented were €0.50 and €0.54.[19] The goal was to ensure a sufficient number of subjects participated in either payment scheme. To this end, the piece rates were chosen such that (perfectly calibrated) subjects with an average performance level would be indifferent between the compensation schemes.[20] Furthermore, the chosen piece rates allow for a comparison of sorting choices with those in similar settings such as in Niederle and Vesterlund (2007).

[18] A slider interface was used to facilitate the decision. Here, two interactive bars indicated the resulting payoff scheme, depending on the currently selected MAP and each possible piece rate in the envelope (see Figure A.2 in the appendix; the interface was inspired by that of Andreoni and Harbaugh 2010 for the elicitation of individual risk attitudes). Additionally, for any currently selected MAP, two lines of verbal description beneath the interface indicated the resulting payment scheme for the case that the piece rate was (i) greater than or equal to the current MAP, and (ii) lower than the current MAP. For each subject, the initial value of the MAP in the interface was randomly determined by the computer.

[19] Variation of the piece rate between sessions serves to increase the incentive compatibility of the mechanism in case subjects communicate between sessions. For the comparability of sessions, the variation was minimal and no subject would have participated in a different payment scheme if the piece rates had been exchanged between sessions.

[20] Let w_P denote the piece rate and w_T denote the payment per correct task in the tournament if a subject ranked first. For a given level of performance x, an expected-payoff–maximizing subject is indifferent between the payment schemes if $w_T \Pr(R = 1|y_0, x)x = w_P x$, where $\Pr(R = 1|y_0, x)$ denotes the probability of ranking first, conditional on the subject's own performance x and the observed performance of the matched participants y_0. For $w_T = 2$ and $w_P \in \{0.50, 0.54\}$, this holds for winning probabilities of 0.25 and 0.27, respectively. Given the data of the reference treatment, these values lie between the most frequently observed performance levels of 5 and 6 (see Table B.2).

2.3.5 Belief elicitation

During the course of the experiment, subjects' beliefs were elicited at three particular points. The first elicitation followed the example stage, the second followed after subjects had learned of the performance of their competitors, and the third took place after subjects had completed the last task of the experiment but before they received feedback of any kind.

At each elicitation, subjects stated their beliefs with respect to the tournament stage (round 2). At each elicitation they were asked to assess (i) how many tasks they would solve (had solved), (ii) how many tasks a randomly selected other participant (RSOP) of the ongoing session would solve (had solved), and (iii) their rank in the tournament stage. The type of the questions was not known to the subjects ex ante. However, the instructions indicated that nine questions would be asked during the course of the experiment and that one of the answers would be randomly selected for payment.

Paying subjects for both their stated beliefs and their sorting choice creates an opportunity for hedging.[21] To mitigate hedging, subjects were informed that at the end of the experiment, one of their answers would be randomly selected and, if the answer was correct, they would receive €1 (see Niederle and Vesterlund 2007).

2.3.6 Elicitation of risk attitudes

With the randomized assignment of subjects to the two main treatments, the expected distribution of individual specific characteristics is the same between treatments. Accordingly, a consideration of subjects' risk attitudes is not crucial for the experimental hypotheses tests, since they rely on between-treatment comparisons (see below). However, individual risk attitudes are likely to explain a significant proportion of the variance in sorting choices (see for example Dohmen and Falk 2011). Therefore, subjects' risk attitudes were elicited at the beginning of the main treatments, using a multiple price list similar to Holt and Laury (2002). In 21 cases, the subjects had to decide whether they preferred a safe payment or a lottery. In each case, the lottery paid €6 or €0 with equal probability. The safe payoff was increased in each case from 0 to €6 in steps of €0.3. Subjects knew

[21] Blanco et al. (2010) find hedging only when the possibility to hedge is very apparent, though.

that one of their choices would be randomly selected at the end of the experiment and, if applicable, the lottery would be played out by the computer.

2.3.7 Questionnaire and feedback

At the end of the experiment, subjects answered a questionnaire. Besides essential demographic data, including age and gender, subjects completed a brief version of the Big-Five Personality Test as well as the Life Orientation Test (LOT-R) in order to measure their generalized degree of optimism.[22] In addition to the elicitation of risk attitudes by the lottery task, I adopted the seven-item questionnaire on risk attitudes by the German Socio-Economic Panel (SOEP).[23] Furthermore, subjects reported their average high-school grade as well as their last math grade in high school.[24]

After the questionnaire, subjects received feedback about their absolute performance and their rank (relative to their competitors) in each round. In order to prevent distortions of sorting decisions due to curiosity, the experimental instructions indicated that feedback about one's rank and one's absolute performance in each round would be provided independently of sorting choices. The feedback screen also reported the realizations of the random draws regarding the lottery task, the belief tasks, and the payoff-relevant round together with the respective payoffs. At the end, subjects were guided to the next room where they received their experimental payments in private.

2.4 Procedures

The experiment was programmed with z-Tree (Fischbacher 2007). The experimental sessions were run at the Technische Universität Berlin in October and

[22] Schaefer et al. (2004) show that Big-Five factors predict overconfidence. The questions for the tests were taken from Rammstedt and John (2007) and Glaesmer et al. (2008), respectively.

[23] The questionnaire can be found at http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.40965.de/personen_2004.pdf. Dohmen et al. (2011) demonstrate the internal reliability of these questions and their predictive power for real-life choices such as smoking and investment behavior. Responses to the general risk question also predicted behavior in incentivized lottery choices.

[24] In addition, the questionnaire included a subset of Raven's Advanced Progressive Matrices (Kratzmeier and Horn 1980). However, due to time constraints, the data on Raven's test could be acquired for a few sessions only and was not utilized in the analysis. In their meta-study on the hindsight bias, Musch and Wagner (2007) find intelligence to be negatively associated with hindsight bias (and overconfidence).

December 2012. Participants were recruited using ORSEE (Greiner 2004). In the reference treatment, 92 subjects participated in four sessions.[25] The average duration of the sessions was 80 minutes, the average earning was €15.39.

For each of the two main treatments, four sessions have been conducted. Seventy-two subjects participated in each of the two main treatments (18 subjects per session, invitations were balanced with respect to gender). The sessions lasted, on average, 1 hour and 44 minutes. The average payoff was €21.13.

2.5 Experimental hypotheses

At the first elicitation, subjects have completed the sample tasks but they have not had access to any performance data yet. When subjects in the informed treatment reason about the degree of difficulty of a sample task, they may not be able to disregard their superior information, i.e., the solution to the task. In accordance with experimental evidence on the curse of knowledge, I expected that subjects in the informed treatment would overestimate the probability that a sample task could be solved by others (see Loewenstein et al. 2006). Assuming that such biased expectations affect assessments of the average performance in further tasks, I surmised that subjects in the informed treatment would overestimate the performance of a randomly selected other participant of the same session (RSOP) in the upcoming tournament stage. Analogously, I expected that subjects in the informed treatment would overestimate the number of sample tasks *they* would have been able to solve, and hence, that they would overestimate their own performance in subsequent rounds.

At the second belief elicitation, subjects have learned the performance of their competitors in the sample stage. If subjects regard this data as being informative, to some extent, of their own performance (and that of others), they will update their performance estimates toward the observed performance.[26] However, given "cursed expectations" about the degree of difficulty of the task, some overestimation of future performance levels is expected to persist in the informed treatment. I will test to which extent the provision of performance data affects expectations

[25] In one of the sessions a client's connection to the server was suspended during the first round and reestablished in the fifth round. The data of this subject was excluded from the entire analysis.

[26] For example, Owens et al. (2012) find that subjects update overconfident performance expectations in a way consistent with Bayesian updating.

about one's own performance and expectations about the performance of RSOPs. In the following, I summarize the hypotheses regarding subjects' estimates of absolute performance levels.

Hypothesis 1 ***(Increased estimates of RSOPs' performance)*** *Subjects' stated beliefs about the absolute performance of a randomly selected other participant of the same session (RSOP) are more optimistic in the informed treatment than in the uninformed treatment.*

Hypothesis 2 ***(Increased estimates of own performance)*** *Subjects' stated beliefs about their own (absolute) performance are more optimistic in the informed treatment than in the uninformed treatment.*

I turn to subjects' assessments of their performance *relative* to others, in particular relative to their competitors. If subjects in the informed treatment underestimate the degree of difficulty of the experimental tasks, they may perceive the past performance of their competitors as "unexpectedly low." In turn, they may underestimate the ability of their competitors on average (see the theory of *information projection* by Madarász 2012). Consequently, informed subjects are expected to be overconfident *relative to their competitors* once they observe their competitors' past performance.

Hypothesis 3 ***(Increased estimates of own performance relative to competitors after observing their past performance)*** *After observing the past performance of their competitors, anticipated tournament ranks are more optimistic for subjects in the informed treatment than for subjects in the uninformed treatment.*

Following *Hypothesis 3*, subjects' anticipated payoffs in the tournament—relative to their anticipated payoffs with piece-rate compensation—are higher in the informed treatment than in the uninformed treatment. Hence, subjects in the informed treatment should exhibit a higher propensity to choose the competitive payment scheme than subjects in the uninformed treatment.

Hypothesis 4 ***(Increased willingness to compete)*** *Subjects in the informed treatment (i) exhibit higher minimum acceptable piece rates than subjects in the uninformed treatment, and (ii) participate more often in the tournament for the implemented piece rates.*

If the curse of knowledge indeed biases subjects' expectations about their *own* ability, the effect of the curse of knowledge on sorting choices should depend on the actual ability of subjects. For example, imagine a subject with outstanding skills who can solve any of the experimental tasks. Such a subject's inference about her skill cannot be biased by the curse of knowledge since for her it does not matter whether she is able to disregard her superior information or not: she will always (correctly) believe that she would have been able to accomplish the sample tasks and would thus (legitimately) entertain very optimistic estimates of her future performance. Conversely, for subjects with low skills, there is plenty of room for the curse of knowledge to increase their perceived likelihood of success for the given kind of tasks. Therefore, the effect of the curse of knowledge on tournament entry might be stronger for subjects with comparatively low skills than for highly skilled subjects.[27]

Hypothesis 5 ***(Interaction with productivity)*** *The treatment difference in tournament entry is larger for subjects with comparatively low productivity than for highly productive subjects.*

3 Results

I first investigate the performance of subjects. Then, the analysis follows the order of the above hypotheses. For the sake of brevity, WMP denotes Wilcoxon's signed-rank test for matched pairs, while MWU denotes Mann-Whitney U test for independent samples. If not stated otherwise, the conclusions drawn from t-tests do not differ from those of non-parametric tests.

3.1 Performance of subjects

The average number of tasks solved in each treatment and round is presented in Table 1. In the following, three observations are made that are important for the

[27] Following this reasoning, the curse of knowledge will have adverse effects on productivity-based sorting, i.e., the extent to which highly skilled subjects are more likely to enter the competitive environment than subjects with comparatively low skills. For example, if all subjects are fully biased, i.e., they believe they can solve all of the sample tasks, then any subject's inference about her skill is independent of her actual skill (assuming that individual priors about skill are independent of actual skill). Here, subjects with inferior skills are as likely to enter the competitive environment as highly skilled subjects, i.e., there would be no productivity-based sorting at all.

Table 1. Average performance by treatment, round (and compensation scheme), and choice of compensation scheme for round 3.

	Treatment	Choice of payment scheme for round 3	n	Round and compensation scheme: Round 1 (Piece Rate)	Round 2 (Tournament)	Round 3 (Endogenous)
(1)	**Uninformed**	**Aggregate**	**72**	**4.4**	**5.2**	**5.2**
				(0.205)	(0.194)	(0.199)
(2)		Piece rate	14	4.4	5.2	5.4
				(0.500)	(0.408)	(0.309)
(3)		Tournament	58	4.4	5.6	5.1
				(0.227)	(0.221)	(0.236)
(4)	**Informed**	**Aggregate**	**72**	**5.0**	**5.3**	**5.3**
				(0.212)	(0.188)	(0.212)
(5)		Piece rate	7	5.7	6.3	5.9
				(0.747)	(0.680)	(0.595)
(6)		Tournament	65	4.9	5.2	5.2
				(0.220)	(0.192)	(0.226)

Note: Values in parentheses represent standard errors.

main analysis following in the next subsection.

Treatment effects If the additional information provided in the informed treatment affected performance, this has to be considered in the tests of *Hypotheses 1–4*. A comparison of rows 1 and 4 in the table suggests a treatment difference in performance in round 1. While subjects in the informed treatment solved 5.0 tasks on average in the first round, the mean performance in the uninformed treatment is only 4.4 tasks. This difference is marginally significant (MWU yields $p = 0.062$).[28] Regarding the rounds of primary interest, i.e., round 2 (belief statements) and round 3 (sorting choices), no significant treatment differences in performance are observed (MWU yields $p = 0.492$ and $p = 0.654$, respectively).

Learning and incentive effects An assessment of an individual's productivity in either incentive scheme is required for the test of *Hypothesis 5.* Rows 1 and 4 of the table reveal a performance increase between the first and second round for both treatments, which is more pronounced in the uninformed treatment. Testing these differences yields mixed results. While the initial increase is significant for

[28] A t-test indicates a significant treatment difference in the average round-1 performance ($p = 0.036$). Chi-squared tests do *not* reject independence between the distributions of individual performance and treatment for any round ($p = 0.245$ for round 1, $p = 0.802$ for round 2, and $p = 0.883$ for round 3).

the uninformed treatment only (WMP yields $p = 0.001$ for the uninformed and $p = 0.301$ for the informed treatment), there is no significant interaction between the treatment and changes in performance over time, which suggests that the dynamics are not different between treatments ($p = 0.183$).[29] The performance levels in rounds 2 and 3 are very similar and not significantly different.[30]

Sorting effects I finally test whether subjects in the uninformed treatment sorted into the payment schemes according to their productivity. Rows 2 and 3 of Table 1 show the average performance levels in the uninformed treatment, conditional on subjects' chosen payment scheme for round 3. If subjects correctly anticipate their productivity, higher performance levels are expected for subjects who sorted into the tournament compared to those who sorted into the piece-rate scheme. However, in the uninformed treatment, there is no such relation.[31] This does not seem to be due to random performance, since individual performances are significantly correlated across rounds (tests of Pearson's ρ against zero yield $p < 0.01$ for each of the three pairwise comparisons). Subjects have difficulties in predicting their performance, which is also the case for other tasks in the experimental literature (see, e.g., Niederle and Vesterlund 2005).

Interpretation of performance data The increase in performance between the first and the second round could be due to learning or incentive effects.[32] I suggest an interpretation as learning, as the data shows no further indications of incentive effects: If incentive effects mattered then, within round 3, the performance under the tournament should be higher than under the piece-rate compensation as well. However, this is not observed. Furthermore, for the informed treatment, round-1 and round-2 performances are much more similar than in the uninformed

[29] I run a probit regression of the probability of solving a task correctly on a treatment dummy, a dummy for round 2, and an interaction of both dummies together with subject-specific random effects.

[30] The performance in rounds 2 and 3 is not significantly different, neither on the aggregate, nor conditional on treatment, on subjects' sorting choices, or conditional on both. WMP gives $p \geq 0.494$ for each of nine comparisons.

[31] MWU tests between the performance of uninformed subjects who opted for the tournament and those who opted for the piece rate yield $p = (0.840, 0.971, 0.618)$ for round 1, 2 and 3, respectively.

[32] To disentangle both explanations, the order of incentive schemes in rounds 1 and 2 could have been reversed in further sessions. The main interest in the present sessions was to change the order of tasks to ensure a proper analysis of whether stated beliefs are correct on average and an analysis of the optimality of sorting choices.

treatment, which is not what incentive effects would predict.

The treatment difference in initial performance (together with the observation that uninformed subjects catch up to the performance level of informed subjects in the second round) suggests that the information provided in the example stage was not entirely irrelevant for subjects' productivity at the outset of the experiment. Although subjects are incentivized to base their choices on their anticipated performances in round 2 (belief elicitation) and round 3 (sorting choice), subjects might expect higher profits in the informed treatment than in the uninformed treatment if they mistakenly base their choices on their (correctly) anticipated round-1 performance. This might result in treatment differences in stated beliefs and sorting choices that are not due to the curse of knowledge. Accordingly, throughout the tests of the hypotheses, I conduct a juxtaposed analysis where I control for individual performance in each separate round: Tests reported in the main text are conducted with reference to subjects' round-2 performance (as incentivized for the belief elicitation). If not stated otherwise, the results are the same for analogous tests based on round-1 and round-3 performance.[33] Throughout, I refer to a result as "not robust", when tests based on round-1, round-2, and round-3 performance yield different results.

3.2 Stated beliefs

Figure 3 shows the average stated beliefs before and after subjects received information about their competitors together with the average actual performances in the tournament stage.[34]

3.2.1 Stated beliefs about absolute performance levels

I first test whether informed subjects are too optimistic when estimating the performance of others, as suggested by *Hypothesis 1*. The triangle markers in Figure 3 illustrate that subjects in the informed treatment significantly overestimate the performance of randomly selected other participants of the same session (RSOP)

[33] Analogous tests based on round-1 and round-3 performance are relegated to the Appendix.

[34] As expected, the performance in the main treatments is not significantly different from the performance in the reference treatment (see Appendix B.1.1). Notably, the depicted average ranks in the tournament stage (right panel) deviate from 2.5. An average rank of 2.5 is expected if the performance in the main treatments and the reference treatment is the same. However, the observed deviation from 2.5 is subject to noise. A t-test does not reject the hypothesis that the average rank over all rounds (2.425) is 2.5 ($p = 0.227$).

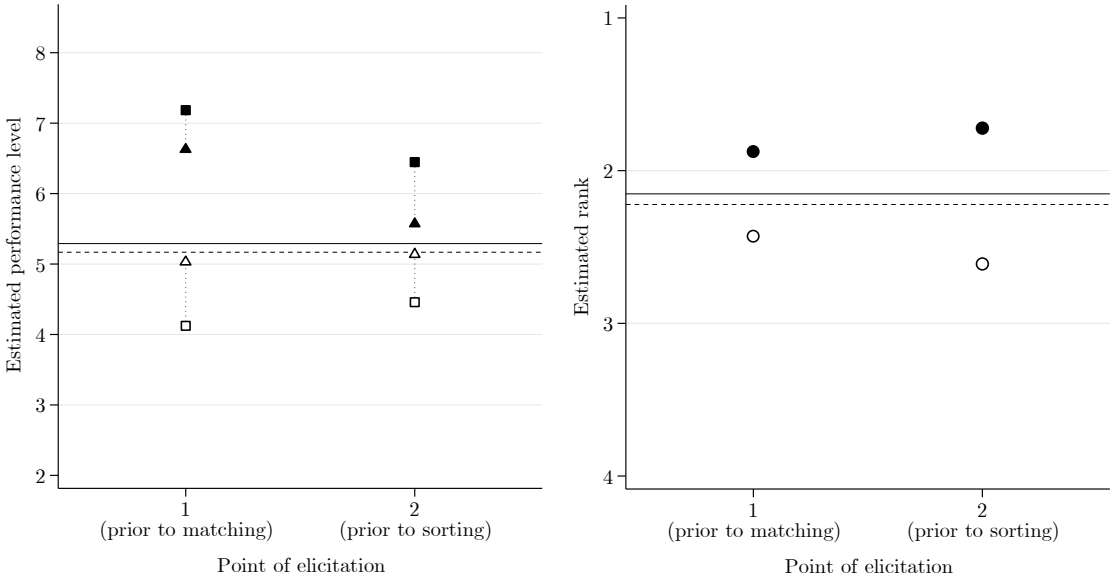

Figure 3. Average stated beliefs of uninformed subjects (white markers) and informed subjects (black markers) at two points in time regarding their own performance level (left panel, square markers), the performance of a randomly selected other participant (left panel, triangle markers), and their rank in the tournament stage (right panel). Horizontal lines represent the average actual performance levels (left panel) and actual ranks (right panel) in the tournament stage for uninformed (dashed line) and informed subjects (solid line).

at the *first* elicitation (permutation test yields $p < 0.001$).[35] In the uninformed treatment, there is no robust evidence for a bias in subjects' estimates of others' performance ($p = 0.637$). Accordingly, at the first elicitation, individual differences between the predicted and actual performance of RSOPs are significantly larger in the informed treatment than in the uninformed treatment (one-sided permutation test yields $p = 0.002$). These findings confirm *Hypothesis 1*.

Result 1.a ***(Overestimation of the performance of RSOPs without access to performance data)*** *Without access to performance data, predictions of the performance of randomly selected other participants of the same session (RSOPs) are significantly more optimistic in the informed treatment than in the uninformed treatment. In the informed treatment, subjects significantly overestimate the performance of RSOPs, while there is no robust evidence for biased predictions in the uninformed treatment.*

After subjects in the informed treatment have learned of the performance of

[35] A description of the permutation tests and corresponding bootstrap techniques is given in Appendix B.3.1.

their competitors in the sample tasks, their overly optimistic expectations regarding the performance of RSOPs vanishes. No changes are observed for uninformed subjects, which is not surprising given that they correctly assessed the performance of RSOPs beforehand.[36] Thus, at the second elicitation, no significant difference is found between treatments (one-sided permutation test: $p = 0.219$). Together these findings suggest that the curse of knowledge with regard to predictions of *others'* performances can be eliminated by providing a rather limited sample of individual performances.

Result 1.b ***(Unbiased estimation of the performance of RSOPs in the presence of performance data)*** *After providing performance data, no treatment difference in the stated beliefs about the performance of randomly selected other participants of the same session (RSOPs) is observed. No robust bias in the estimation of the performance of RSOPs is found in either treatment.*

It was hypothesized that the curse of knowledge would also increase subjects' expectations regarding their *own* performance level (*Hypothesis 2*). The square markers in Figure 3 show that this is the case. At the first elicitation, informed subjects significantly overestimate their performance on average (WMP yields $p < 0.001$), while uninformed subjects exhibit—if anything—underestimation ($p < 0.001$).[37] Contrary to the findings above, this pattern persists after subjects receive performance data. Although overestimation in the informed treatment is moderated by providing performance data, informed subjects remain overconfident ($p < 0.001$). In the uninformed treatment, subjects' expectations about their own performance are more optimistic after receiving performance data. However, indications for underestimation persist ($p = 0.005$).[38]

Accordingly, for both the first and the second belief elicitation, individual differences between anticipated and actual performance are significantly more optimistic in the informed treatment than in the uninformed treatment (MWU yields $p < 0.001$). Together, these findings confirm *Hypothesis 2*:

[36] In both treatments, there is no robust difference between subjects' stated beliefs and the actual performance of RSOPs (see Appendix B.3.1).

[37] Underestimation in the uninformed treatment is not robust. See Table B.4 in the Appendix for analogous tests based on round-1 and round-3 performance.

[38] Again, underestimation in the uninformed treatment is not robust (see Table B.4 in the Appendix).

Table 2. Tests of the treatment effect on anticipated tournament ranks (prior to sorting).

Dependent variable	Anticipated tournament rank (1–4)				
(Ordered probit)	(1)	(2)	(3)	(4)	(5)
Treatment[§]	-1.025^{***}	-1.028^{***}	-1.021^{***}	-1.023^{***}	-1.035^{***}
(1-informed)	(0.190)	(0.194)	(0.190)	(0.190)	(0.194)
Performance X_1		0.005			0.022
(Piece rate)		(0.055)			(0.058)
Performance X_2			-0.026		-0.018
(Tournament)			(0.059)		(0.062)
Performance X_3				-0.045	-0.046
(Endogenous)				(0.055)	(0.059)
N	144	144	144	144	144
$\log L$	-174.816	-174.811	-174.715	-174.488	-174.392
χ^2_{k-1}	33.567	33.577	33.769	34.223	34.416

[§]Stars represent one-sided p-values for tests of the treatment-dummy coefficient and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors. Estimated cutoffs, dummy coefficients for task order, type of matching, and their interactions are omitted. The task order and the type of matching had no significant effects.

Result 2 ***(Overestimation of own performance)*** *Stated beliefs about own (absolute) performance are significantly more optimistic in the informed treatment than in the uninformed treatment. On average, informed subjects overestimate their actual performance; no overestimation is observed for uninformed subjects. A provision of performance data alleviates, but does not eliminate overestimation of own performance in the informed treatment.*

3.2.2 Relative self-assessments

I turn to an investigation of subjects' relative self-assessments, specifically their anticipated rank in the tournament (right panel of Figure 3). Are informed subjects more optimistic than uninformed subjects with respect to anticipated tournament ranks as stated in *Hypothesis 3*? At the second elicitation, after subjects have observed the performance of their competitors in the sample tasks, a clear treatment effect on anticipated tournament ranks is observed. While 15.3% of the subjects in the uninformed treatment guess that they will win the tournament, 48.6% do so in the informed treatment. The average anticipated tournament rank in the uninformed treatment is 2.6, while in the informed treatment, the average anticipated rank is about one rank higher (1.7).

In order to test the treatment effect on anticipated tournament ranks, controlling for individual performance, type of matching, and task order, Table 2 reports the results of ordered probit regressions. The regressions indicate significantly higher probabilities of more optimistic rank guesses in the informed treatment than in the uninformed treatment. Similarly to the missing relation between individual performance and sorting choices (see previous section), individual performances are not significantly related to anticipated tournament ranks.[39] Given that performance is virtually orthogonal to stated beliefs, individual differences between performance and beliefs are subject to considerable noise. Nevertheless, testing these differences between treatments confirms that informed subjects are significantly more optimistic than uninformed subjects (a between-treatment MWU test of individual differences between anticipated ranks and actual ranks yields $p < 0.01$). Within-treatment comparisons of stated beliefs and actual ranks reveal significant overplacement in the informed treatment (WMP yields $p = 0.015$), and, if anything, underplacement in the uninformed treatment ($p = 0.033$).[40] These findings confirm *Hypothesis 3*.

Result 3.a ***(Overplacement relative to competitors after observing their performance)*** *After subjects have observed performance of their competitors, anticipated tournament ranks are significantly more optimistic in the informed treatment than in the uninformed treatment. On average, informed subjects exhibit overplacement; no overplacement is found for uninformed subjects.*

At the first elicitation, anticipated tournament ranks also tend to be more optimistic in the informed treatment than in the uninformed treatment. However, the evidence is weaker here, since one-sided tests of the treatment difference yield overall only marginally significant results.[41] Within treatments, informed subjects have a tendency to overplacement, while the results for uninformed subjects are mixed. Between the first and second elicitation, anticipated tournament ranks become more optimistic in the informed treatment (WMP yields $p = 0.045$), while they become more pessimistic in the uninformed treatment ($p = 0.032$). That is, the treatment difference in anticipated tournament ranks became more

[39] Alternative regressions with dummy coefficients for each performance level yield very similar results (see Table B.6 in the appendix).

[40] Underplacement in the uninformed treatment is not robust (see Table B.5 in the Appendix).

[41] See Table B.5 in the Appendix for a detailed overview of the results reported in this section.

pronounced after subjects had received information about the past performance of their competitors.

Interestingly, overplacement of informed subjects is not only found in comparison with their competitors but also in comparison with randomly selected other participants of the same session (RSOPs; see left panel of Figure 3). At both the first and second elicitation, individual differences between stated beliefs regarding own performance and stated beliefs regarding the performance of RSOPs are significantly larger in the informed treatment than in the uninformed treatment (MWU yields $p < 0.001$ for both elicitations). In the informed treatment, subjects' estimates of their own performance are significantly more optimistic than their estimates of the performance of RSOPs, while in the uninformed treatment, subjects considered themselves as underdogs compared to the RSOPs (WMP yields $p < 0.002$ for both treatments and elicitations).

Result 3.b ***(Overplacement relative to RSOPs)*** *Individual differences between subjects' estimates of their own performance and their estimates of the performance of randomly selected other participants of the same session (RSOPs) are significantly larger in the informed treatment than in the uninformed treatment. In the informed (uninformed) treatment, subjects' estimates of their own performance are significantly more optimistic (pessimistic) than their estimates of the performance of RSOPs.*

3.3 Sorting choices

After subjects had completed the sample tasks and learned the performance of their competitors, they made their sorting choice with respect to round 3. The previous analysis confirms that self-assessments of future performance relative to others are overly optimistic, if there is scope for the curse of knowledge. Does this translate into higher willingness to compete? If this is the case, then the minimum acceptable piece rates (MAPs) in the informed treatment should be higher than in the uninformed treatment (*Hypothesis 4*). Figure 4 depicts the empirical cumulative distribution functions (CDFs) of the MAPs in each treatment.

The figure reveals that the empirical CDF in the informed treatment first-order stochastically dominates the CDF in the uninformed treatment. The average MAP in the informed treatment amounts to €1.13, which is 19% higher than the average

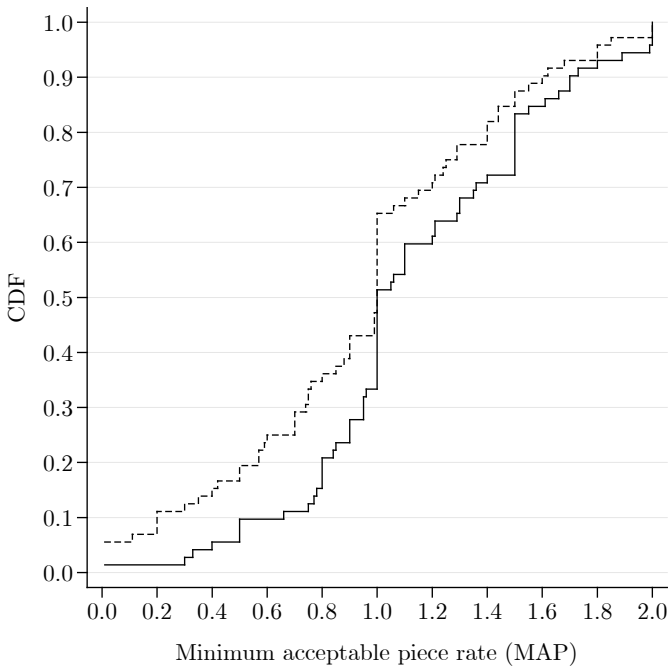

Figure 4. Cumulative distribution functions of minimum acceptable piece rates in the uninformed (dashed line) and the informed treatment (solid line).

MAP in the uninformed treatment (€0.95). For tests of the difference in MAPs between treatments, Table 3 reports the results of Tobit regressions with controls for individual performance, type of matching, and task order.

The regressions confirm a significant treatment effect on MAPs. The variation in individual performance does not account for the treatment effect on sorting choices, although the effect is somewhat weaker when controlling for performance in round 1. The results are corroborated by regressions with fixed effects for each performance level and further controls such as self-reported willingness to take risks, gender, age, high-school grade, and psychological traits (Big-Five and optimism; see Table B.7 in the appendix).[42]

Thus far, the analysis has confirmed a treatment effect on both subjects' anticipated tournament ranks and their willingness to compete. Naturally, a significant relationship between subject's anticipated ranks and their sorting choices is expected. Table 4 presents the results of Tobit regressions with the same specifications as in Table 3 except that subjects' anticipated ranks have been added as an explanat-

[42] MAPs are significantly increasing in self-reported willingness to take risks (SOEP) and optimism (LOT-R; partially significant only) and significantly decreasing in neuroticism (Big-Five). Additionally, for some specifications, MAPs are marginally increasing in openness (Big-Five), age, and decreasing in high-school grade.

Table 3. Tests of the treatment effect on minimum acceptable piece rates.

Dependent variable	Minimum acceptable piece rate (MAP)				
(Tobit regressions)	(1)	(2)	(3)	(4)	(5)
Treatment[§]	0.193***	0.188**	0.195***	0.192***	0.188**
(1-informed)	(0.080)	(0.082)	(0.080)	(0.080)	(0.081)
Performance X_1		0.008			0.011
(Piece rate)		(0.024)			(0.025)
Performance X_2			−0.021		−0.027
(Tournament)			(0.026)		(0.027)
Performance X_3				0.006	0.010
(Endogenous)				(0.024)	(0.025)
Constant	0.840***	0.804***	0.967***	0.811***	0.901***
	(0.145)	(0.182)	(0.211)	(0.189)	(0.233)
N	144	144	144	144	144
χ^2_{k-1}	11.304	11.413	11.985	11.362	12.429
Pseudo R^2	0.051	0.051	0.054	0.051	0.056

[§]Stars represent one-sided p-values for tests of the treatment-dummy coefficient and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors. Estimated dummy coefficients for task order, type of matching, and their interactions are omitted. The task order and the type of matching had no significant effects.

Table 4. Tests of the treatment effect on minimum acceptable piece rates controlling for anticipated tournament ranks.

Dependent variable	Minimum acceptable piece rate (MAP)				
(Tobit regressions)	(1)	(2)	(3)	(4)	(5)
Treatment[§]	0.006	−0.000	0.008	0.006	−0.002
(1-informed)	(0.082)	(0.083)	(0.081)	(0.082)	(0.082)
Anticipated rank	−0.209***	−0.209***	−0.210***	−0.209***	−0.211***
(Tournament)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Performance X_1		0.010			0.016
(Piece rate)		(0.022)			(0.023)
Performance X_2			−0.026		−0.031
(Tournament)			(0.024)		(0.025)
Performance X_3				−0.001	0.003
(Endogenous)				(0.022)	(0.023)
Constant	1.312***	1.268***	1.471***	1.318***	1.416***
	(0.161)	(0.189)	(0.216)	(0.199)	(0.234)
N	144	144	144	144	144
χ^2_{k-1}	36.145	36.344	37.364	36.148	37.914
Pseudo R^2	0.163	0.164	0.168	0.163	0.171

[§]Stars represent one-sided p-values for tests of the treatment-dummy coefficient and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors. Estimated dummy coefficients for task order, type of matching, and their interactions are omitted.

ory variable. There are two clear observations. First, the estimated coefficients on anticipated tournament ranks confirm a significant relationship between subjects' relative self-assessments and their sorting choices. Second, once subjects' anticipated ranks are added as controls, the treatment difference in MAPs is virtually zero, or, to put it differently, conditional on the level of relative self-assessment, average sorting choices are the same for both treatments.[43]

Result 4 ***(Increased willingness to compete)*** *Minimum acceptable piece rates are significantly higher in the informed treatment than in the uninformed treatment. This treatment difference can be explained by differences in subjects' anticipated tournament ranks.*

Do higher MAPs in the informed treatment entail more overentry—or less underentry? To investigate this question, I divide subjects according to their (round 2) tournament performance x, where I classify subjects with $x \geq 6$ as *high-performers* and those with $x < 6$ as *low-performers*.[44] This specific threshold is chosen because it (i) yields a relatively balanced number of subjects in both groups and (ii) is informative regarding the optimality of the sorting in the experiments: For a given performance, an expected payoff maximizer is indifferent between the piece rate and the tournament if $w_P x = w_T \Pr(R = 1 \mid y_0, x)x$, where w_P denotes the (unconditional) piece rate, $w_T = 2$ denotes the payment per correct answer in the tournament in the event of winning, and $\Pr(R = 1 \mid y_0, x)$ denotes the probability of winning the tournament conditional on the past performance of the competitors y_0 and own performance x. Given the values of w_P and w_T that were implemented in the experiments and the empirical winning probabilities, expected payoffs from tournament entry exceeded those with piece-rate compensation whenever a subject solved at least six tasks—irrespective of the type of matching and task order (see Table B.2 in the appendix). Hence, I consider sorting choices to be optimal if low-performers sort into the piece rate and high-performers sort into the tournament, while I refer to the reverse cases as *overentry* and *underentry*, respectively.

[43] Beliefs were elicited with respect to the tournament stage (round 2) while sorting choices were made with respect to round 3. I implicitly assume comparability of both, i.e., that MAPs (or anticipated ranks) would not have been systematically different for round 2 (round 3).

[44] The classification based on individual performance in the tournament stage (round 2) is significantly correlated with classifications based on round-1 and round-3 performances (chi-squared tests of independence yield $p = 0.029$ and $p = 0.001$, respectively).

Table 5. Tests of the treatment effect on tournament participation conditional on individual productivity (based on round-2 performance).

Dep. variable (Probit)	Tournament entry					
	Aggregate data		Low-performers		High-performers	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment[§]	0.436**	0.512**	1.055**	1.270**	0.103	−0.029
(1-informed)	(0.265)	(0.281)	(0.484)	(0.602)	(0.359)	(0.415)
Risk-Taking		0.127*		0.189*		0.135
(SOEP)		(0.066)		(0.107)		(0.100)
Gender		0.493*		0.254		0.671
(1-male)		(0.289)		(0.441)		(0.416)
Highschool gr.		0.036		−0.673		0.701*
(1–4,1-best)		(0.246)		(0.460)		(0.416)
Age		0.030		0.055		0.005
		(0.033)		(0.061)		(0.044)
Constant	0.862***	−0.771	0.859***	0.090	0.865***	−1.584
	(0.169)	(0.884)	(0.224)	(1.626)	(0.259)	(1.149)
N	144	144	77	77	67	67
$\log L$	−58.431	−53.292	−24.806	−21.624	−31.451	−25.720
χ^2_{k-1}	2.777	13.054	5.932	12.295	0.082	11.545

Note: The regressions in columns 3–6 are based on the classification of subjects according to their tournament performance (round 2). [§]Stars represent one-sided p-values for tests of the treatment-dummy coefficient and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors.

In order to assess the treatment effect on sorting choices for low- and high-performers, Table 5 shows the results of probit regressions of the probability of tournament entry on the treatment dummy and further controls. The regressions were run for the aggregate data and separately for high- and low-performers, as defined above.

The regressions for the aggregate data in columns 1 and 2 corroborate the previous finding that informed subjects are significantly more likely to enter the tournament than uninformed subjects (result 4).[45] The remaining columns suggest that the treatment effect is mainly due to low-performers: The treatment effect is significant for low-performers, but not for high-performers. Furthermore, when subjects are classified according to their round-2 performance (as in Table 5), the treatment effect is significantly larger for low-performers than for high-performers (see Table B.9 in the appendix).[46] However, this interaction effect is not significant

[45] The results are similar for regressions with fixed effects for each performance level and further controls (see Table B.8 in the appendix).

[46] Table B.9 in the appendix contains probit regressions of tournament entry on a treatment

in regressions, where the classification of subjects as high- or low-performers is based on their round-1 or round-3 performance. Nevertheless, separate tests of the treatment effect for low-performers and high-performers yield a consistent pattern: Irrespective of the round considered to classify subjects as low- or high-performers, the treatment effect is always significant for low-performers, while it is never significant for high-performers (see Table B.9 in the appendix). For the given performance threshold to separate high- and low-performers, these results can be rephrased in terms of sorting optimality: for the incentives in the experiment, the curse of knowledge entails a significant increase in overentry while there is no significant decrease in underentry. Regarding *Hypothesis 5*, I conclude:

Result 5 ***(Overentry)*** *The increased self-selection into the tournament in the informed treatment is mainly due to low-performers. While low-performers sort significantly more often into the tournament in the informed treatment than in the uninformed treatment, no significant treatment effects are observed for high-performers.*

4 Conclusion

The literature on judgment and decision making shows that both overconfidence and the curse of knowledge are robust biases that occur in a wide range of contexts. While the economic consequences of overconfidence have been studied in many settings, the sources of overconfidence are only partially understood. Loewenstein et al. (2006) have shown that the curse of knowledge can cause overly optimistic expectations regarding the performance of others. The present experiment complements their findings by showing that subjects, when "cursed by knowledge," also overestimate their *own* performance—even more than they overestimate the performance of others. Such comparative overconfidence in the presence of the curse of knowledge is observed in two dimensions. First, in line with the theory of *information projection* by Madarász (2012), subjects are overconfident relative to their competitors, once they observe their competitors' past performance. Second, in the presence of the curse of knowledge, subjects are overconfident relative to randomly selected other participants of the same session, about whom they receive no information at all. Together with the observation that the curse of knowledge

dummy, a dummy for high-performers, a dummy for their interaction, and further controls.

decreases the *perceived* degree of difficulty of a task, this finding can be reconciled with previous studies that have found a negative relationship between task difficulty and the extent of comparative overconfidence (see Kruger 1999, and Moore and Cain 2007).[47]

This paper also contributes to the literature on self-selection and incentive schemes. To the best of my knowledge, this is the first experiment showing that the curse of knowledge can cause overentry into competitive environments. With respect to the empirical literature on business formation, the present findings offer a potential explanation for entrepreneurial overconfidence and the high failure rates of new businesses (see Cooper et al. 1988; Wu and Knott 2006; Koellinger et al. 2007). Nascent entrepreneurship is certainly characterized by ability uncertainty in many dimensions, and feedback regarding one's entrepreneurial success is presumably rare. Informational asymmetries as the key ingredient for the curse of knowledge are likely to be the norm rather than the exception in this context. For example, a nascent entrepreneur might know about the success of past investments, product innovations, or personnel decisions of established entrepreneurs. Agents with the curse of knowledge cannot disregard such superior information when taking the perspective of the established entrepreneur at the time of her decision making or when picturing themselves in the same situation. As a consequence, they underestimate the uncertainty that established entrepreneurs faced when making their decisions, and overestimate the ex-ante probability of profitable managerial decision making and entrepreneurial success. As most entrepreneurs start their business in the field of their prior employment, direct comparisons with established entrepreneurs, and hence biased comparative ability judgments are likely to play a role in addition to biased absolute ability assessments. The urge of nascent entrepreneurs to make well-informed decisions based, for example, on detailed market analyses, may fuel the curse of knowledge by widening the informational gap between what *was* known at the time the established entrepreneurs made their decisions and what *is* known when the nascent entrepreneur engages in (comparative) ability judgments.

More research is needed to validate these claims and to assess whether, and to which extent, the curse of knowledge accounts for entrepreneurial overconfidence

[47] Kruger (1999) and Moore and Cain (2007) find *better-than-average* effects for easy tasks, i.e., tasks with comparatively high success rates, but *worse-than-average effects* for hard tasks. The authors refer to this relation as the *hard-easy effect*.

and the high failure rate of new businesses. Further research may assess whether the curse of knowledge affects choices in other kinds of self-selection problems. For example, overly optimistic comparative ability judgments induced by the curse of knowledge may affect self-selection into revenue-sharing schemes. Similarly, in the context of sorting choices where only one's own absolute ability is relevant, cursed agents with comparatively low ability may suboptimally choose payment schemes designed to attract highly skilled agents.

Bibliography

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, 49(6):1621–1630.

Andreoni, J. and Harbaugh, W. T. (2010). Unexpected utility: Experimental tests of five key questions about preferences over risk. University of Oregon Economics Department Working Paper 2010-14, University of Oregon Economics Department.

Arkes, H. R., Wortmann, R. L., Saville, P. D., and Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology*, 66(2):252–254.

Biais, B. and Weber, M. (2009). Hindsight bias, risk perception, and investment performance. *Management Science*, 55(6):1018–1029.

Birch, S. A. J. and Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18(5):382–386.

Blanco, M., Engelmann, D., Koch, A., and Normann, H.-T. (2010). Belief elicitation in experiments: Is there a hedging problem? *Experimental Economics*, 13(4):412–438.

Bohnet, I., Greig, F., Herrmann, B., and Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, 98(1):294–310. ArticleType: research-article / Full publication date: Mar., 2008 / Copyright © 2008 American Economic Association.

Bohnet, I. and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4):467–484.

Bryant, F. B. and Guilbault, R. L. (2002). "I knew it all along" eventually: The development of hindsight bias in reaction to the Clinton impeachment verdict. *Basic and Applied Social Psychology*, 24(1):27–41.

Bukszar, E. and Connolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. *Academy of Management Journal*, 31(3):628–641.

Camerer, C., Loewenstein, G., and Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97(5):1232–1254. ArticleType: research-article / Full publication date: Oct., 1989 / Copyright © 1989 The University of Chicago Press.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1):306–318. ArticleType: research-article / Full publication date: Mar., 1999 / Copyright © 1999 American Economic Association.

Cesarini, D., Lichtenstein, P., Johannesson, M., and Wallace, B. (2009). Heritability of overconfidence. *Journal of the European Economic Association*, 7(2-3):617–627.

Christensen-Szalanski, J. J. and Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48(1):147–168.

Cooper, A. C., Woo, C. Y., and Dunkelberg, W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing*, 3(2):97–108.

Danz, D., Hüber, F., Kübler, D., Mechtenberg, L., and Schmid, J. (2013). 'I'll do it by myself as i knew it all along': On the failure of hindsight-biased principals to delegate optimally. SFB 649 Discussion Paper 9, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Danz, D. N. (2013). The curse of knowledge increases self-selection into competition. In *Three essays on irrational beliefs*, pages 61–128. Dissertation. Technische Universität Berlin, Germany.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

Dennis, W. J. (1997). More than you think: An inclusive estimate of business entries. *Journal of Business Venturing*, 12(3):175–196.

Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2):556–590.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3):288–299. Cited by 1736.

Fischhoff, B. and Beyth, R. (1975). I knew it would happen: Remembered probabilities of once–future things. *Organizational Behavior and Human Performance*, 13(1):1–16.

Gaskill, L. R., Van Auken, H. E., and Manning, R. A. (1993). A factor analytic study of the perceived causes of small business failure. *Journal of Small Business Management*, 31(4):18–31.

Glaesmer, H., Hoyer, J., Klotsche, J., and Herzberg, P. Y. (2008). Die deutsche Version des Life-Orientation-Tests (LOT-R) zum dispositionellen Optimismus und Pessimismus. *Zeitschrift für Gesundheitspsychologie*, 16(1):26–31.

Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen 2003, GWDG Bericht 63*, pages 79–93. Ges. für Wiss. Datenverarbeitung, Göttingen.

Guilbault, R. L., Bryant, F. B., Brockway, J. H., and Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, 26(2-3):103–117.

Hamilton, B. (2000). Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy*, 108(3):604–631. ArticleType: research-article / Full publication date: June 2000 / Copyright © 2000 The University of Chicago Press.

Headd, B. (2003). Redefining business success: Distinguishing between closure and failure. *Small Business Economics*, 21(1):51–61.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655. ArticleType: research-article / Full publication date: Dec., 2002 / Copyright © 2002 American Economic Association.

Keysar, B. and Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, 13(3):207–212. PMID: 12009039.

Knight, F. H. (1921). *Risk, Uncertainty and Profit.* Houghton Mifflin Company, New York.

Koellinger, P., Minniti, M., and Schade, C. (2007). "I think I can, I think I can": Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology*, 28(4):502–527.

Kratzmeier, H. and Horn, W. (1980). Raven-Matrizen-Test: Advanced Progressive Matrices (APM). *Weinheim: Beltz Test Gesellschaft.*

Kruger, J. (1999). Lake Wobegon be gone! The 'below-average effect' and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77(2):221–232.

Kruger, J., Epley, N., Parker, J., and Ng, Z.-W. (2005). Egocentrism over e-mail: Can we communicate as well as we think? *Journal of Personality and Social Psychology*, 89(6):925–936.

Loewenstein, G., Moore, D., and Weber, R. (2006). Misperceiving the value of information in predicting the performance of others. *Experimental Economics*, 9(3):281–295.

Madarász, K. (2012). Information projection: Model and applications. *Review of Economic Studies*, 79(3):961–985.

Moore, D. A. and Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, 103(2):197–213.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2):502–517.

Musch, J. and Wagner, T. (2007). Did everybody know it all along? A review of individual differences in hindsight bias. *Social Cognition*, 25(1):64–82.

Newton, E. (1990). Overconfidence in the communication of intent: Heard and unheard melodies. *Ph.D. Thesis, Stanford University, Stanford, CA*.

Niederle, M. and Vesterlund, L. (2005). Do women shy away from competition? Do men compete too much? Working Paper 11474, National Bureau of Economic Research.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101.

Nisbett, R. E. and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice-Hall.

Owens, D., Grossman, Z., and Fackler, R. (2012). The control premium: A preference for payoff autonomy. *Working Paper, Department of Economics, UC Santa Barbara, University of California at Santa Barbara, Economics Working Paper Series*.

Patterson, R. A., Kozmetsky, G., and Ridgway, N. M. (1983). Perceived causes of small business failures: A research note. *American Journal of Small Business*, 8(1):15–19.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 36(1):11–46. ArticleType: research-article / Full publication date: Mar., 1998 / Copyright © 1998 American Economic Association.

Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1):203–212.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5):368–373.

Schaefer, P. S., Williams, C. C., Goodie, A. S., and Campbell, W. (2004). Overconfidence and the Big Five. *Journal of Research in Personality*, 38(5):473–480.

Simons, D. J. and Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1(7):261–267.

Slovic, P. and Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4):544–551.

Wu, B. and Knott, A. M. (2006). Entrepreneurial risk and market entry. *Management Science*, 52(9):1315–1330.

Appendices

Appendix A Experimental design

A.1 Supplemental figures

Figure A.1. Solution to example task as provided in the informed treatment (task-id 13).

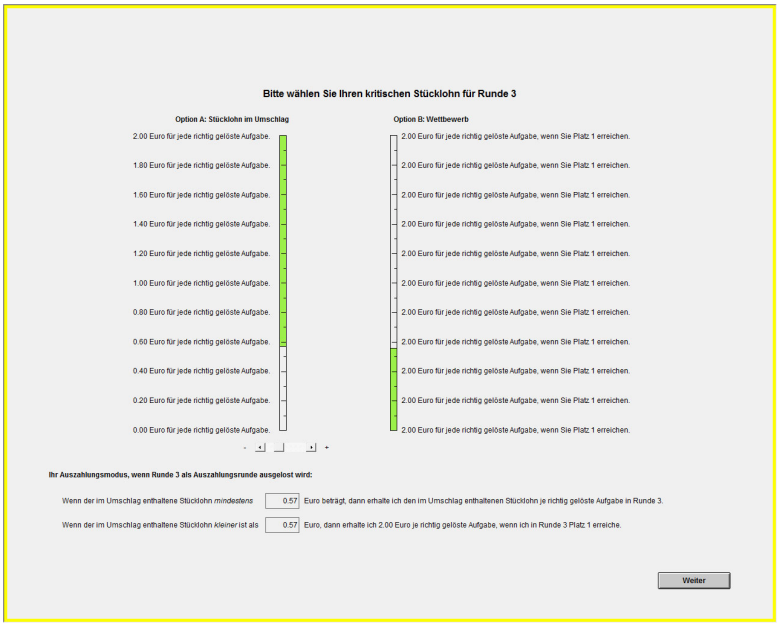

Figure A.2. Elicitation of minimum acceptable piece rates.

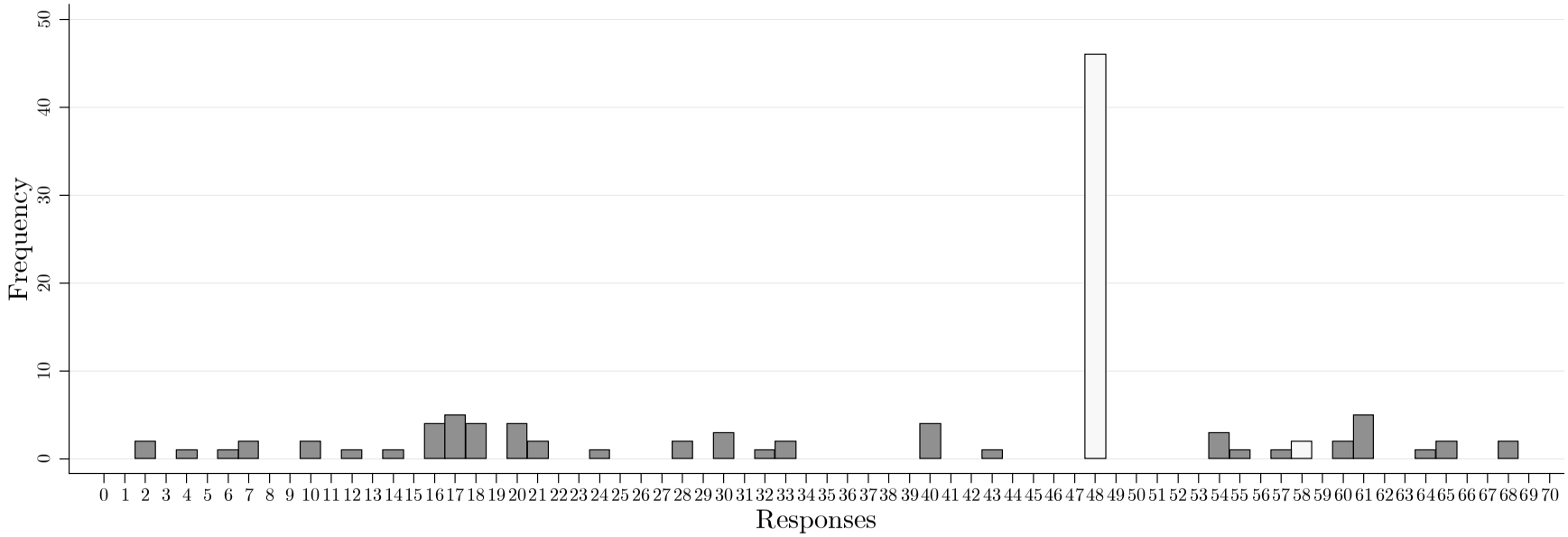

Figure A.3. Example of response pattern (task-id 13). White bars indicate valid answers, gray bars represent invalid answers. Answers coded as correct: 34, 38, 44, 48, 58. Overall success rate in the main treatments was 44.44% ($n = 108$).

A.2 Task order

In order to assess whether subjects' stated beliefs are correct on average and to assess the average optimality of sorting choices, the order of the tasks was counterbalanced between the sessions of the two main treatments as follows. First, to avoid severe variation in the degree of difficulty across rounds, the tasks for the main treatments were partitioned into four sets of tasks based on the success rates per task in the reference treatment.[48] The assignment was random, with the restriction that each set of tasks contains one task from each decile of the empirical difficulty distribution. This assignment to sets of tasks was held constant across all sessions. The *order* of the task sets was varied across sessions, such that comparisons of performance across rounds and comparisons of stated beliefs with actual performance are possible on the aggregate, i.e., over all sessions—even if the sets of tasks differ regarding their degree of difficulty (see Table A.1).

Table A.1. Assignment of task sets to rounds in the main treatments.

		Task set assigned to			
Task order	Applied to sessions	Example stage	Round 1 (Piece rate)	Round 2 (Tournament)	Round 3 (Self-selected)
1	1–2	1	2	3	4
2	3–4	4	3	2	1
3	5–6	2	1	4	3
4	7–8	3	4	1	2

Note: For each pair of sessions, with task set s assigned to the example stage and task set s' assigned to round r, there is one other pair of sessions with reversed task sets, i.e., with task set s assigned to round r and task set s' assigned to the example stage. The number of subjects was balanced in this respect.

[48] In order to assess the level of difficulty for each task without potential confounds through learning effects, the task order in the reference treatment was determined randomly in the first session, and systematically interchanged in subsequent sessions.

Appendix B Supplementary data analysis

B.1 Performance of subjects

B.1.1 Comparison of the main treatments and the reference treatment (competitors)

Subjects in the main treatments solved 5.05 out of 10 tasks on average which is not significantly different from the average performance of 5.19 in the reference treatment (MWU: $p = 0.563$). Table B.1 reports the success rates for each task in the reference treatment and the two main treatments.

Table B.1. Average performance per task in the reference treatment and the two main treatments.

		Solution rate				Solution rate	
Set of tasks	Task-ID[§]	Main treatments	Reference treatment	Set of tasks	Task-ID[§]	Main treatments	Reference treatment
	109	0.519	0.538		101	0.815	0.736
	18	0.333	0.352		112	0.472	0.429
	10	0.685	0.747		13	0.444	0.473
	16	0.759	0.780		25	0.343	0.341
1	28	0.333	0.407	3	38	0.407	0.352
	36	0.519	0.571		30	0.676	0.681
	57	0.713	0.670		58	0.157	0.264
	2	0.398	0.363		37	0.787	0.769
	11	0.204	0.242		19	0.556	0.615
	51	0.435	0.516		42	0.472	0.549
	9	0.620	0.670		6	0.648	0.549
	22	0.611	0.626		50	0.519	0.352
	114	0.259	0.341		41	0.611	0.626
	47	0.454	0.538		52	0.741	0.780
2	15	0.546	0.725	4	3	0.417	0.527
	31	0.778	0.758		54	0.657	0.714
	108	0.426	0.396		34	0.426	0.462
	21	0.583	0.538		49	0.185	0.198
	201	0.185	0.187		8	0.667	0.637
	27	0.435	0.429		12	0.389	0.319

[§]Tasks ordered as in sessions 1 and 2. 100-IDs indicate tasks provided by Rensink et al. (1997).

B.1.2 Performance by sets of tasks

The creation of sets of tasks that are interchangeable with respect to their degree of difficulty was mostly successful. Figure B.1 gives the histogram of individual performances for each set of tasks.[49] In order to test for between-set differences in the level of difficulty while controlling for repeated observations on the individual level, I run a probit regression of the probability of solving a task correctly on a constant and dummy variables for task sets 2, 3, and 4 with random subject effects. Testing this specification against a model with only a constant, I cannot reject the hypothesis of no differences in the degree of difficulty between task sets ($p = 0.330$). Also for pairwise comparisons between sets, I find insignificant differences in the performance distributions—except for the comparison between set 2 and set 4.[50] However, due to the experimental design, performance differences between the sets of tasks are not problematic for the analysis (see Table A.1).

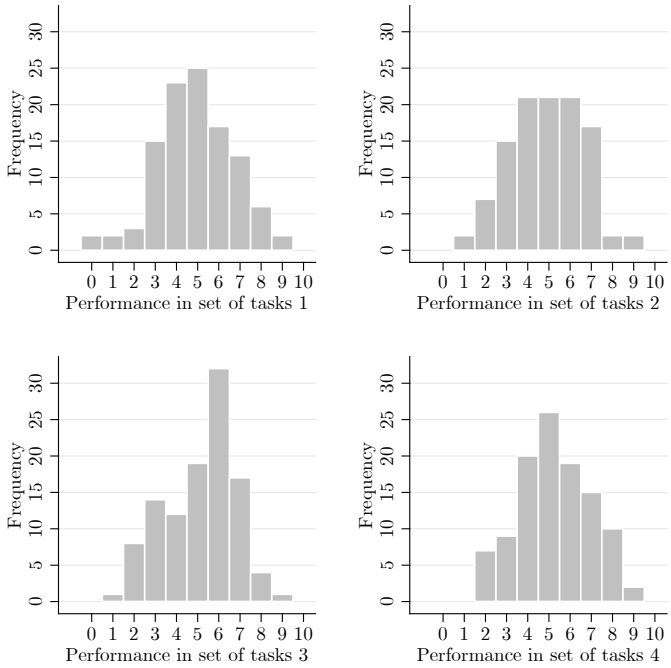

Figure B.1. Histograms of individual performance for each set of tasks in the main treatments.

[49] The average performances (ranges) are 4.90 (0-9), 4.90 (1-9), 5.13 (1-9), and 5.26 (2-9) for task set 1, 2, 3, and 4, respectively.

[50] Of the six comparisons between sets, based on the estimated coefficients, only one marginally significant difference is found between set 2 and set 4 ($p = 0.096$). Pairwise sign tests for subjects who participated in both blocks confirm these results. WMP-tests do not reject differences in the distributions for the six pairwise comparisons except for the comparison between set 2 and set 4 ($p = 0.044$).

B.1.3 Performance over time

Regarding the performance of subjects across rounds, Figure B.2 depicts the histograms of individual performance in the piece-rate compensation scheme (round 1), the tournament (round 2), and the self-selected compensation scheme (round 3). While the distributions exhibit a median performance of 5 in each round and rather stable modes of $5, 6$, and 5 for rounds 1, 2, and 3 respectively, the average performance levels in round 2 (5.3) and round 3 (5.2) are significantly higher than the average performance in round 1 (4.7, WMP yields $p = 0.003$ for both comparisons).[51]

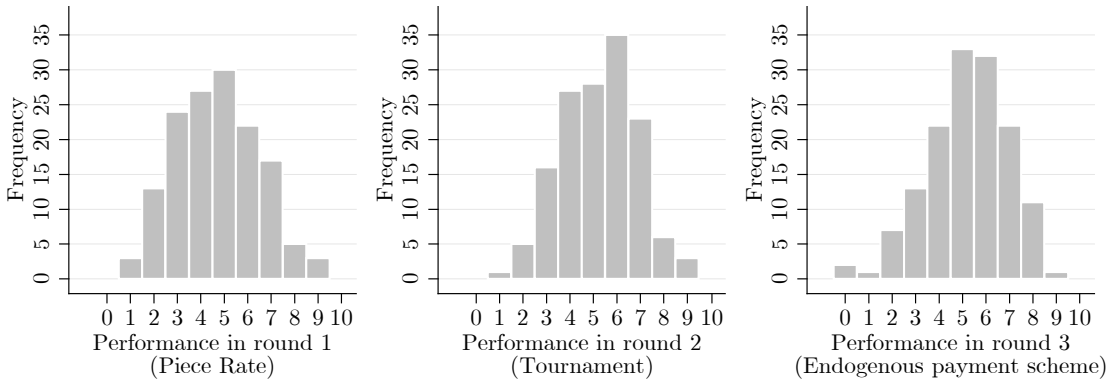

Figure B.2. Histograms of performance in the piece-rate stage (round 1), the tournament stage (round 2), and the stage with endogenous compensation scheme (round 3).

[51] A WMP test of the faint difference between rounds 2 and 3 yields $p = 0.633$.

B.2 Empirical probabilities of winning against the reference treatment

Table B.2. Empirical probabilities of winning against the competitors.

Round r	Task order (sessions/w_{PR})	$y_0 = (y_{01}, y_{02}, y_{03})$	$Pr\{R_r = 1 \mid y_0, X_r = k\}$ $k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
1	1 (1,2/0.50)	(4, 5, 5)	0.000	0.000	0.000	0.000	0.019	0.146	0.459	0.755	0.873	0.948	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.000	0.016	0.123	0.382	0.667	0.864	0.974	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.004	0.035	0.150	0.417	0.691	0.841	0.948	1.000
	2 (3,4/0.54)	(4, 5, 5)	0.000	0.000	0.000	0.002	0.021	0.172	0.521	0.768	0.920	1.000	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.000	0.013	0.111	0.345	0.598	0.853	1.000	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.000	0.011	0.127	0.426	0.654	0.853	1.000	1.000
	3 (5,6/0.54)	(4, 5, 5)	0.000	0.000	0.000	0.003	0.045	0.195	0.418	0.661	0.850	0.957	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.001	0.029	0.158	0.384	0.634	0.824	0.948	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.001	0.022	0.136	0.342	0.588	0.793	0.928	1.000
	4 (7,8/0.50)	(4, 5, 5)	0.000	0.000	0.000	0.001	0.013	0.128	0.413	0.685	0.897	1.000	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.001	0.021	0.124	0.365	0.628	0.864	1.000	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.001	0.018	0.162	0.484	0.717	0.890	1.000	1.000
2	1 (1,2/0.50)	(4, 5, 5)	0.000	0.000	0.000	0.004	0.049	0.232	0.574	0.870	0.972	1.000	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.002	0.027	0.155	0.435	0.698	0.882	1.000	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.004	0.064	0.241	0.501	0.741	0.906	1.000	1.000
	2 (3,4/0.54)	(4, 5, 5)	0.000	0.000	0.000	0.000	0.008	0.140	0.447	0.658	0.768	0.893	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.000	0.010	0.120	0.358	0.621	0.836	0.944	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.000	0.015	0.171	0.466	0.658	0.764	0.893	1.000
	3 (5,6/0.54)	(4, 5, 5)	0.000	0.000	0.000	0.002	0.032	0.211	0.543	0.789	0.944	1.000	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.002	0.029	0.148	0.361	0.621	0.887	1.000	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.002	0.030	0.171	0.407	0.690	0.938	1.000	1.000
	4 (7,8/0.50)	(4, 5, 5)	0.000	0.000	0.000	0.001	0.030	0.159	0.398	0.652	0.849	0.972	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.001	0.022	0.136	0.369	0.629	0.811	0.942	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.001	0.029	0.147	0.336	0.579	0.815	0.969	1.000
3	1 (1,2/0.50)	(4, 5, 5)	0.000	0.000	0.000	0.001	0.016	0.138	0.394	0.650	0.898	1.000	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.002	0.020	0.116	0.324	0.522	0.802	1.000	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.004	0.018	0.099	0.265	0.439	0.763	1.000	1.000
	2 (3,4/0.54)	(4, 5, 5)	0.000	0.000	0.000	0.000	0.018	0.170	0.408	0.599	0.845	1.000	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.000	0.025	0.155	0.326	0.511	0.745	0.933	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.000	0.012	0.125	0.342	0.545	0.745	0.933	1.000
	3 (5,6/0.54)	(4, 5, 5)	0.000	0.000	0.000	0.005	0.051	0.210	0.460	0.694	0.891	1.000	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.001	0.028	0.165	0.429	0.674	0.874	1.000	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.002	0.030	0.145	0.383	0.639	0.860	1.000	1.000
	4 (7,8/0.50)	(4, 5, 5)	0.000	0.000	0.000	0.000	0.017	0.125	0.372	0.631	0.756	0.898	1.000
		(4, 5, 6)	0.000	0.000	0.000	0.000	0.016	0.117	0.375	0.680	0.845	0.947	1.000
		(5, 5, 6)	0.000	0.000	0.000	0.001	0.017	0.104	0.330	0.646	0.801	0.898	1.000

Note: Table entries represent empirical probabilities of achieving rank $R_r = 1$ in round r conditional on (i) own performance $X_r = k$, (ii) the task order, and (iii) the observed performance of the three matched participants $y_0 = (y_{01}, y_{02}, y_{03})$, where y_{0j} refers to the number of sample tasks ("round 0" in the main treatments) that were solved by subject j from the reference treatment (here in tournaments).

B.3 Stated beliefs

B.3.1 Stated beliefs about a randomly selected other participant

In the following I describe the testing procedures regarding the belief tasks where subjects had to assess the performance of a randomly selected participant of the same session (RSOP). Here, permutation tests are required, since, to the best of my knowledge, standard tests are not applicable. For example, paired t-tests or Wilcoxon's signed-rank tests are not suitable since they are based on pairwise comparisons of an individual's guess and the individual's *own* performance. t-tests or Wilcoxon's signed-rank tests of stated beliefs against the average performance within each treatment are not applicable since (i) they do not account for the variation in the estimated level of actual performance, and (ii) the performance level that one tests against contains an individual's own performance.

Within treatments, the difference between the stated beliefs about an RSOP's performance and an RSOP's actual performance has been tested against zero. Between treatments, the corresponding difference in difference was of interest. For the sake of robustness as well as consistency, I employed both permutation tests and bootstrapping techniques.

Permutation tests The general procedure follows Davison and Hinkley (1997), p.141ff. I start with the tests of *within treatment* differences. Consider some stage of belief elicitation s, treatment and round of interest r. The p-values for tests of the average within-treatment bias against zero are obtained by the following permutation test. Under the null hypothesis, both guesses and actual performances of other participants, are drawn from the same population. For each of $n_{mc} = 1,000,000 - 1$ random permutations the procedure randomly draws (without replacement) guesses and actual performances from the entire set of guesses and performances where the guess and performance of each subject is treated as a paired observation. That is, if a subject's stated belief is randomly assigned to the group of permuted performances [stated beliefs], her performance is assigned to the group of permuted stated beliefs [performances]. For each simulation, the test statistic t is calculated as the average difference between the permuted guess of a subject and the permuted performance of a randomly selected *other* participant. Let $n_{t \geq \bar{t}}$ be the number of permutations with test statistic t greater than or equal to the test statistic in the original sample $\bar{t}$ (both in absolute

terms).[52] Then, the (two-sided) p-value is calculated as $p_{mc} = (n_{t \geq \bar{t}} + 1)/(n_{mc} + 1)$ (see Davison and Hinkley 1997, p.158, p.163).

Regarding *between treatment* differences, the permutation test differs from the one described above in the following way: Under the null hypothesis, the subjects from both treatments are drawn from the same population. Accordingly, in $(10,000)$ random permutations I reassign (without replacement) subjects' paired observations from both treatments to two artificially generated "new treatment groups." Within each group, the average difference between a subject's stated beliefs and the actual performance of a randomly selected other member of this group are calculated. The between-treatment difference of these averages yields the bootstrap test statistic. P-values are calculated as above.

Bootstrap confidence intervals Bootstrap confidence intervals are obtained by the basic percentile method (see Davison and Hinkley 1997, p.202ff). I start with the confidence intervals for the *within treatment* test-statistics. For each of $n_b = 1,000,000 - 1$ simulations within each treatment and round of interest the algorithm randomly draws subjects with replacement, with the number of draws being equal to that of the original sample size (72). Then, within each bootstrap sample, for each subject the difference between her stated belief and the performance of an (again) randomly drawn *other* subject of the same bootstrap sample is computed. The mean difference of the bootstrap sample gives the bootstrap test statistic t. The bootstrap confidence interval is obtained by the quantiles $(t_{((n_b+1)\alpha/2)}, t_{((n_b+1)(1-\alpha)/2)}), \alpha = 0.05$ of the empirical CDF of the bootstrapped statistic t.

To obtain confidence intervals for the test statistic *between treatments*, the same procedure is utilized. Here, the bootstrap test-statistic is the between-treatment difference of t (confidence intervals based on $10,000$ resamples).

Results Table B.3 gives the p-values obtained by the permutation tests together with the bootstrapped $(1-\alpha/2)$-percentile intervals for $\alpha = 0.05$. The table reveals that the test results based on the permutation test coincide with results based on the bootstrapped percentile intervals.

[52] In two cases repeated values of the test statistic between simulations are observed and ties occur in the sense that counts based on $t \geq \bar{t}$ differ from those based on $t > \bar{t}$ by more than one. However, results based on less conservative p-values obtained by counts of $t > \bar{t}$ are the same (see discussion of bounded p-values in Davison and Hinkley 1997, p.141).

Table B.3. Bootstrap percentile intervals and results of permutation tests for biases in stated beliefs regarding the tournament-performance (round 2) of randomly selected other participants. For the sake of robustness, tests against the actual performance in each of the three rounds are provided.

Treatment	Elicitation	Round	Point estimate	p-value (perm. test) two-sided	p-value (perm. test) left-sided	Bootstrapped 95%-confidence interval
Uninformed	1	1	0.639**	(0.031)	.	[0.042, 1.236]
		2	−0.139	(0.637)	.	[−0.722, 0.444]
		3	−0.139	(0.622)	.	[−0.694, 0.431]
	2	1	0.750***	(0.009)	.	[0.167, 1.333]
		2	−0.028	(0.936)	.	[−0.583, 0.528]
		3	−0.028	(0.895)	.	[−0.583, 0.542]
	3	1	0.875***	(0.003)	.	[0.306, 1.444]
		2	0.097	(0.678)	.	[−0.431, 0.625]
		3	0.097	(0.739)	.	[−0.458, 0.681]
Informed	1	1	1.611***	(0.000)	.	[0.889, 2.319]
		2	1.333***	(0.000)	.	[0.722, 1.958]
		3	1.375***	(0.000)	.	[0.708, 2.069]
	2	1	0.556*	(0.065)	.	[−0.069, 1.167]
		2	0.278	(0.310)	.	[−0.278, 0.861]
		3	0.319	(0.278)	.	[−0.264, 0.931]
	3	1	0.028	(0.937)	.	[−0.556, 0.597]
		2	−0.250	(0.287)	.	[−0.736, 0.250]
		3	−0.208	(0.453)	.	[−0.778, 0.389]
Treatment difference (Inf.−Uninf.)	1	1	0.972*	(0.053)	(0.026)	[0.042, 1.903]
		2	1.472***	(0.003)	(0.002)	[0.625, 2.319]
		3	1.514***	(0.002)	(0.001)	[0.639, 2.403]
	2	1	−0.194	(0.643)	(0.679)	[−1.056, 0.653]
		2	0.306	(0.440)	(0.219)	[−0.472, 1.111]
		3	0.347	(0.398)	(0.197)	[−0.472, 1.181]
	3	1	−0.847**	(0.041)	(0.980)	[−1.653,−0.042]
		2	−0.347	(0.407)	(0.781)	[−1.069, 0.375]
		3	−0.306	(0.468)	(0.764)	[−1.111, 0.514]

Note: Point estimates of biases obtained by average difference between individual belief statements and the performance of other participants in the original sample (average bootstrap test statistics are virtually the same). Elicitations: 1-prior to matching, 2-prior to sorting, 3-after performance. Calculations based on performance in round 2 (tournament stage) correspond to the belief elicitation as implemented in the experiment.

B.3.2 Stated beliefs about own (absolute) performance and ranks in the tournament

Table B.4. Comparison of subjects' estimate of their own performance in the tournament (round 2) and their actual performances (in all rounds).

	Treatment		Round for comparison 1	**2**	3
After sample stage (w/o information about competitors)	Uninformed	Stated belief		4.125	
		Performance	4.389	5.167	5.167
		Difference =: Δ_U^1	−0.264	−1.042	−1.042
		WMP (p-value)	0.308	0.000	0.000
		t-test (p-value)	0.325	0.000	0.000
	Informed	Stated belief		7.181	
		Performance	5.014	5.292	5.250
		Difference =: Δ_I^1	2.167	1.889	1.931
		WMP (p-value)	0.000	0.000	0.000
		t-test (p-value)	0.000	0.000	0.000
	DID	$\Delta_I^1 - \Delta_U^1$	2.431	2.931	2.972
		WMP (p-value)	0.000	0.000	0.000
		t-test (p-value)	0.000	0.000	0.000
Before sorting (w/ information about competitors)	Uninformed	Stated belief		4.458	
		Performance	4.389	5.167	5.167
		Difference =: Δ_U^2	0.069	−0.708	−0.708
		WMP (p-value)	0.792	0.005	0.004
		t-test (p-value)	0.789	0.005	0.003
	Informed	Stated belief		6.444	
		Performance	5.014	5.292	5.250
		Difference =: Δ_I^2	1.431	1.153	1.194
		WMP (p-value)	0.000	0.000	0.000
		t-test (p-value)	0.000	0.000	0.000
	DID	$\Delta_I^2 - \Delta_U^2$	1.361	1.861	1.903
		WMP (p-value)	0.001	0.000	0.000
		t-test (p-value)	0.001	0.000	0.000

Note: Bold columns indicate comparisons as incentivized (and as outlined in the main text). Payment schemes in the rounds for comparison: 1-piece rate, 2-tournament, 3-endogenous.

Table B.5. Comparison of subjects' estimate of their rank in the tournament (round 2) and their actual rank (in all rounds).

	Treatment		Round for comparison 1	**2**	3
After sample stage (w/o information about competitors)	Uninformed	Stated belief		2.431	
		Actual rank	2.806	2.222	2.389
		Difference =: Δ_U^1	−0.375	0.208	0.042
		WMP (p-value)	0.032	0.079	0.742
		t-test (p-value)	0.025	0.144	0.800
	Informed	Stated belief		1.875	
		Actual rank	2.611	2.153	2.347
		Difference =: Δ_I^1	−0.736	−0.278	−0.472
		WMP (p-value)	0.000	0.131	0.011
		t-test (p-value)	0.000	0.068	0.005
	DID	$\Delta_I^1 - \Delta_U^1$	−0.361	−0.486	−0.514
		WMP (p-value)	0.120	0.018	0.038
		t-test (p-value)	0.124	0.019	0.028
Before sorting (w/ information about competitors)	Uninformed	Stated belief		2.611	
		Actual rank	2.806	2.222	2.389
		Difference =: Δ_U^2	−0.194	0.389	0.222
		WMP (p-value)	0.226	0.033	0.279
		t-test (p-value)	0.286	0.027	0.230
	Informed	Stated belief		1.722	
		Actual rank	2.611	2.153	2.347
		Difference =: Δ_I^2	−0.889	−0.431	−0.625
		WMP (p-value)	0.000	0.015	0.002
		t-test (p-value)	0.000	0.011	0.001
	DID	$\Delta_I^2 - \Delta_U^2$	−0.694	−0.819	−0.847
		WMP (p-value)	0.007	0.001	0.003
		t-test (p-value)	0.007	0.001	0.001

Note: Bold columns indicate comparisons as incentivized (and as outlined in the paper). Payment schemes in the rounds for comparison: 1-piece rate, 2-tournament, 3-endogenous.

Table B.6. Regressions of anticipated tournament rank on treatment, individual performance, task order, and type of matching.

Dependent variable (Ordered probits)	Anticipated tournament rank on performance in round		
	$r = 1$	$r = 2$	$r = 3$
Treatment[§] (1-informed)	-1.052^{***}	-1.103^{***}	-1.036^{***}
	(0.198)	(0.197)	(0.194)
$\mathbf{1}_{X_r=1}$			-9.874
			.
$\mathbf{1}_{X_r=2}$	0.506	-1.945	-0.936
	(0.705)	(1.189)	(0.968)
$\mathbf{1}_{X_r=3}$	0.571	-1.923^{*}	-0.277
	(0.672)	(1.133)	(0.907)
$\mathbf{1}_{X_r=4}$	0.203	-1.214	-0.570
	(0.673)	(1.103)	(0.880)
$\mathbf{1}_{X_r=5}$	0.139	-1.509	-0.123
	(0.659)	(1.110)	(0.870)
$\mathbf{1}_{X_r=6}$	0.307	-1.582	-0.805
	(0.680)	(1.108)	(0.887)
$\mathbf{1}_{X_r=7}$	0.374	-1.949^{*}	-0.766
	(0.692)	(1.115)	(0.894)
$\mathbf{1}_{X_r=8}$	0.145	-1.827	-0.491
	(0.859)	(1.171)	(0.903)
$\mathbf{1}_{X_r=9}$	1.250	-0.535	-9.808
	(0.904)	(1.242)	.
$\mathbf{1}_{\text{task-order}=2}$	0.154	0.182	-0.005
	(0.469)	(0.470)	(0.472)
$\mathbf{1}_{\text{task-order}=3}$	0.326	0.243	0.437
	(0.460)	(0.470)	(0.461)
$\mathbf{1}_{\text{task-order}=4}$	0.605	0.622	0.441
	(0.460)	(0.473)	(0.457)
$\mathbf{1}_{\text{matching}=2}$	0.299	0.236	0.207
	(0.456)	(0.465)	(0.468)
$\mathbf{1}_{\text{matching}=3}$	0.558	0.582	0.579
	(0.475)	(0.480)	(0.472)

Table continued on next page.

Table B.6. Continued.

Dependent variable (Ordered probits)	Anticipated tournament rank on performance in round		
	$r = 1$	$r = 2$	$r = 3$
$\mathbf{1}_{\text{task-order}=2} \times \mathbf{1}_{\text{matching}=2}$	−0.116	−0.182	0.174
	(0.645)	(0.673)	(0.654)
$\mathbf{1}_{\text{task-order}=2} \times \mathbf{1}_{\text{matching}=3}$	−0.423	−0.442	−0.296
	(0.661)	(0.680)	(0.668)
$\mathbf{1}_{\text{task-order}=3} \times \mathbf{1}_{\text{matching}=2}$	0.069	0.248	0.144
	(0.645)	(0.655)	(0.658)
$\mathbf{1}_{\text{task-order}=3} \times \mathbf{1}_{\text{matching}=3}$	−0.401	−0.253	−0.622
	(0.656)	(0.657)	(0.663)
$\mathbf{1}_{\text{task-order}=4} \times \mathbf{1}_{\text{matching}=2}$	−0.978	−0.684	−0.493
	(0.673)	(0.650)	(0.659)
$\mathbf{1}_{\text{task-order}=4} \times \mathbf{1}_{\text{matching}=3}$	−0.375	−0.405	−0.115
	(0.656)	(0.694)	(0.670)
Cutoff value 1	−0.378	−2.321**	−1.256
	(0.714)	(1.157)	(0.926)
Cutoff value 2	0.646	−1.266	−0.204
	(0.713)	(1.149)	(0.928)
Cutoff value 3	1.412*	−0.470	0.590
	(0.725)	(1.145)	(0.924)
N	144	144	144
$\log L$	−172.450	−168.864	−169.101
χ^2_{k-1}	38.299	45.470	44.997

[§]Stars represent one-sided p-values for tests of the treatment-dummy coefficient and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors.

B.4 Sorting choices

B.4.1 Treatment effect on minimum acceptable piece rates

Table B.7. Regressions of minimum acceptable piece rate on treatment on treatment dummy, dummies for each performance level, task order, type of matching, and further explanatories.

Dependent variable (Tobit regressions)	Minimum acceptable piece rate on performance in round		
	$r=1$	$r=2$	$r=3$
Treatment[§]	0.172***	0.215***	0.194***
(1-informed)	(0.071)	(0.069)	(0.071)
Risk taking	0.057***	0.058***	0.054***
(SOEP)	(0.017)	(0.017)	(0.018)
Gender	−0.019	0.025	0.031
(1-male)	(0.079)	(0.076)	(0.078)
High-school grade	0.102	0.123*	0.109
(1-best,4-worst)	(0.065)	(0.066)	(0.069)
Age	0.014*	0.011	0.013
	(0.007)	(0.007)	(0.008)
Extraversion	0.071	0.020	0.018
(10-item Big Five)	(0.155)	(0.151)	(0.156)
Agreeableness	0.334	0.249	0.215
(10-item Big Five)	(0.227)	(0.227)	(0.233)
Conscientiousness	0.235	0.284	0.198
(10-item Big Five)	(0.191)	(0.186)	(0.191)
Neuroticism	−0.642***	−0.551***	−0.611***
(10-item Big Five)	(0.187)	(0.188)	(0.188)
Openness	0.267*	0.276*	0.264*
(10-item Big Five)	(0.154)	(0.148)	(0.154)
Optimism (LOT-R)	0.026***	0.012	0.016
	(0.010)	(0.009)	(0.010)
$1_{X_r=1}$			0.068
			(0.528)
$1_{X_r=2}$	0.153	−0.042	0.091
	(0.265)	(0.466)	(0.344)
$1_{X_r=3}$	0.121	−0.012	0.046
	(0.256)	(0.444)	(0.326)
$1_{X_r=4}$	0.270	−0.142	0.043
	(0.258)	(0.435)	(0.316)
$1_{X_r=5}$	0.264	−0.115	0.035
	(0.254)	(0.438)	(0.312)

Table continued on next page.

Table B.7. Continued.

Dependent variable (Tobit regressions)	Minimum acceptable piece rate on performance in round		
	$r = 1$	$r = 2$	$r = 3$
$\mathbf{1}_{X_r=6}$	0.174	−0.197	0.047
	(0.256)	(0.435)	(0.321)
$\mathbf{1}_{X_r=7}$	0.406	0.084	0.139
	(0.262)	(0.437)	(0.320)
$\mathbf{1}_{X_r=8}$	0.517	−0.413	−0.012
	(0.314)	(0.460)	(0.324)
$\mathbf{1}_{X_r=9}$	−0.384	−0.229	0.732
	(0.347)	(0.490)	(0.522)
$\mathbf{1}_{\text{task-order}=2}$	0.139	0.147	0.148
	(0.171)	(0.174)	(0.178)
$\mathbf{1}_{\text{task-order}=3}$	0.190	0.207	0.170
	(0.168)	(0.175)	(0.175)
$\mathbf{1}_{\text{task-order}=4}$	0.030	0.090	0.083
	(0.167)	(0.172)	(0.174)
$\mathbf{1}_{\text{matching}=2}$	−0.155	−0.149	−0.135
	(0.165)	(0.171)	(0.174)
$\mathbf{1}_{\text{matching}=3}$	0.122	0.152	0.111
	(0.171)	(0.176)	(0.176)
$\mathbf{1}_{\text{task-order}=2} \times \mathbf{1}_{\text{matching}=2}$	0.194	0.091	0.040
	(0.237)	(0.251)	(0.251)
$\mathbf{1}_{\text{task-order}=2} \times \mathbf{1}_{\text{matching}=3}$	−0.056	−0.115	−0.067
	(0.234)	(0.246)	(0.246)
$\mathbf{1}_{\text{task-order}=3} \times \mathbf{1}_{\text{matching}=2}$	0.305	0.248	0.237
	(0.237)	(0.243)	(0.250)
$\mathbf{1}_{\text{task-order}=3} \times \mathbf{1}_{\text{matching}=3}$	−0.335	−0.435*	−0.366
	(0.239)	(0.244)	(0.250)
$\mathbf{1}_{\text{task-order}=4} \times \mathbf{1}_{\text{matching}=2}$	0.322	0.113	0.093
	(0.240)	(0.237)	(0.247)
$\mathbf{1}_{\text{task-order}=4} \times \mathbf{1}_{\text{matching}=3}$	−0.044	−0.143	−0.046
	(0.240)	(0.258)	(0.256)
Constant	−0.665	−0.217	−0.277
	(0.437)	(0.534)	(0.514)
N	144	144	144
χ^2_{k-1}	67.091	65.271	58.446
Pseudo R^2	0.302	0.294	0.263

[§]Stars represent one-sided p-values for tests of the treatment-dummy coefficient and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors. Constant represents reference category with lowest observed performance level (1 for 1st and 2nd round, and 0 for 3rd round).

B.4.2 Treatment effect on tournament participation

Table B.8. Supplemental regressions of tournament participation on treatment dummy, dummies for each performance level, task order, type of matching, and further explanatories.

Dependent variable (Probit regressions)	Tournament entry Performance control (r-round)			
	all rounds	$r=1$	$r=2$	$r=3$
Treatment[§]	0.648**	0.671**	0.666**	0.707**
(1-informed)	(0.318)	(0.371)	(0.350)	(0.358)
Risk taking	0.114	0.192**	0.117	0.130
(SOEP)	(0.077)	(0.095)	(0.092)	(0.085)
Gender	0.531	0.466	0.702*	0.511
(1-male)	(0.342)	(0.388)	(0.397)	(0.372)
High-school grade	0.187	−0.050	0.124	0.228
(1-best,4-worst)	(0.322)	(0.369)	(0.373)	(0.361)
Age	0.019	0.025	0.010	0.020
	(0.036)	(0.040)	(0.038)	(0.046)
Extraversion	0.240	0.362	0.232	0.522
(10-item Big Five)	(0.684)	(0.757)	(0.735)	(0.779)
Agreeableness	1.092	2.143*	1.738	0.743
(10-item Big Five)	(1.028)	(1.240)	(1.135)	(1.093)
Conscientiousness	0.682	1.037	0.978	0.776
(10-item Big Five)	(0.814)	(0.915)	(0.928)	(0.854)
Neuroticism	−2.681***	−3.162***	−2.788***	−2.856**
(10-item Big Five)	(1.003)	(1.151)	(1.073)	(1.110)
Openness	0.783	0.764	0.936	0.836
(10-item Big Five)	(0.671)	(0.758)	(0.724)	(0.752)
Optimism (LOT-R)	0.043	0.106*	0.017	0.050
	(0.040)	(0.054)	(0.046)	(0.047)
$\mathbf{1}_{X_r=2}$		1.225		
		(1.200)		
$\mathbf{1}_{X_r=3}$		0.998	−1.648	
		(1.145)	(1.843)	
$\mathbf{1}_{X_r=4}$		1.457	−2.084	−4.470**
		(1.189)	(1.813)	(1.929)
$\mathbf{1}_{X_r=5}$		1.095	−1.309	−4.544**
		(1.188)	(1.833)	(1.913)
$\mathbf{1}_{X_r=6}$		0.463	−2.388	−4.008**
		(1.089)	(1.822)	(1.892)

Table continued on next page.

Table B.8. Continued.

Dependent variable (Probit regressions)	Tournament entry Performance control round r			
	all rounds	$r=1$	$r=2$	$r=3$
$\mathbf{1}_{X_r=7}$		1.358	−1.438	−4.558**
		(1.208)	(1.853)	(2.043)
$\mathbf{1}_{X_r=8}$		.	−2.668	−4.814**
		.	(1.923)	(2.073)
$\mathbf{1}_{X_r=9}$		−1.081	−2.943	.
		(1.336)	(2.010)	.
$\mathbf{1}_{\text{task-order}=2}$	0.735	1.469**	0.955	0.816
	(0.509)	(0.674)	(0.588)	(0.558)
$\mathbf{1}_{\text{task-order}=3}$	0.237	0.632	0.438	0.049
	(0.440)	(0.514)	(0.493)	(0.480)
$\mathbf{1}_{\text{task-order}=4}$	0.128	0.386	0.088	0.056
	(0.436)	(0.500)	(0.482)	(0.476)
$\mathbf{1}_{\text{matching}=2}$	−0.250	0.127	−0.185	−0.218
	(0.391)	(0.453)	(0.438)	(0.435)
$\mathbf{1}_{\text{matching}=3}$	−0.053	0.024	0.089	−0.180
	(0.404)	(0.453)	(0.451)	(0.443)
Constant	−1.459	−3.931*	0.250	2.751
	(1.704)	(2.353)	.	.
N	144	139	139	122
$\log L$	−44.910	−40.593	−40.044	−40.902
χ^2_{k-1}	29.819	36.847	37.945	30.253

[§]Stars represent one-sided p-values for tests of the treatment-dummy coefficient and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors.

B.4.3 Interaction of the treatment effect with productivity

By *Hypothesis 5,* the treatment effect on tournament entry is expected to be higher for low-productive subjects compared to high-productive subjects. Table B.9 shows probit regressions to test the interaction effect of individual productivity and treatment on tournament entry. As expected, the coefficients on the interaction dummies are negative throughout all regressions, which is in line with the smaller treatment effects for high-performers compared to low-performers. However, the interaction effects are significant only for the regressions where the classification of subjects is based on round-2 performance.[53] Nevertheless, separate tests of the treatment effect for low-performers and high-performers yield a very consistent pattern. Irrespective of the round considered to classify subjects as being low- or high-performers, the treatment effect is always significant for low-performers ($p < 0.041$ for each of the six regressions; see coefficients on the treatment dummy in the first row of the table), while the treatment effect is never significant for high-performers ($p > 0.151$ for each of the six regressions; obtained by joint tests of the coefficients on the treatment and the interaction dummy).

[53] For the classification based on round-1 (column 2), round-2 (column 4), and round-3 (column 6) performances, one-sided tests of the interaction effects yield $p = (0.163, 0.032, 0.210)$, respectively.

Table B.9. Tests of interaction effect of individual productivity and treatment on tournament participation.

Dependent variable (Probit regression)	Tournament entry Performance-cutoff based on					
	Round 1		Round 2		Round 3	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment[§]	0.668**	1.000**	1.235***	1.666***	0.741**	0.912**
(1-informed)	(0.384)	(0.454)	(0.518)	(0.652)	(0.415)	(0.452)
Productivity	−0.376	−0.057	−0.023	−0.000	−0.048	0.152
(1-high-performer)	(0.391)	(0.448)	(0.372)	(0.420)	(0.365)	(0.419)
Interaction[§]	−0.206	−0.687	−1.105**	−1.526**	−0.407	−0.536
(informed×high-p.)	(0.590)	(0.699)	(0.651)	(0.824)	(0.573)	(0.665)
Risk taking	0.134**	0.101	0.145**	0.126	0.130*	0.098
(SOEP)	(0.068)	(0.076)	(0.069)	(0.078)	(0.067)	(0.075)
Gender	0.540*	0.532	0.496*	0.522	0.549*	0.548
(1-male)	(0.295)	(0.338)	(0.294)	(0.345)	(0.296)	(0.349)
High-school grade	0.023	0.272	0.085	0.367	0.052	0.249
(1-best,4-worst)	(0.250)	(0.308)	(0.261)	(0.334)	(0.249)	(0.304)
Age	0.033	0.021	0.034	0.020	0.030	0.018
	(0.035)	(0.037)	(0.035)	(0.037)	(0.033)	(0.034)
Extraversion		0.144		0.309		0.147
(10-item Big Five)		(0.677)		(0.714)		(0.662)
Agreeableness		0.386		1.073		0.735
(10-item Big Five)		(0.982)		(1.046)		(0.961)
Conscientiousness		0.444		0.740		0.513
(10-item Big Five)		(0.786)		(0.832)		(0.790)
Neuroticism		−2.651***		−2.391**		−2.438**
(10-item Big Five)		(0.996)		(0.997)		(0.965)
Openness		0.815		0.697		0.692
(10-item Big Five)		(0.661)		(0.650)		(0.647)
Optimism (LOT-R)		0.040		0.059		0.039
		(0.040)		(0.043)		(0.040)
Constant	−0.754	−0.870	−1.046	−2.073	−0.832	−1.097
	(0.952)	(1.586)	(0.986)	(1.884)	(0.897)	(1.596)
N	144	144	144	144	144	144
$\log L$	−51.942	−45.277	−50.544	−42.939	−52.754	−46.056
χ^2_{k-1}	15.754	29.084	18.551	33.761	14.130	27.527

[§]Stars represent one-sided p-values for tests of the coefficient on the treatment- and the interaction dummy and two-sided p-values in all other cases: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Values in parentheses represent standard errors.

Appendix C Instructions for the main treatments (German)

Das Experiment, an dem Sie nun teilnehmen werden, ist Teil eines von der Deutschen Forschungsgemeinschaft (DFG) finanzierten Projekts. Es dient dazu, ökonomisches Entscheidungsverhalten zu analysieren.

Von jetzt an ist es während des gesamten Experiments nicht mehr erlaubt, mit anderen Teilnehmern zu sprechen oder irgendwelche Informationen auszutauschen bzw. weiterzugeben.

Sollten Sie eine Frage haben, dann heben Sie bitte Ihre Hand, und wir werden zu Ihnen kommen und Ihre Frage beantworten. Bitte stellen Sie Fragen auf keinen Fall laut. Sollten Sie gegen diese Regeln verstoßen, sind wir leider dazu gezwungen, das Experiment abzubrechen.

Für Ihr Erscheinen zu diesem Experiment erhalten Sie 5,– Euro. Zusätzlich können Sie während des Experiments Geld verdienen. Die Instruktionen sind für alle Experimentteilnehmer gleich.

Allgemeines

Das Experiment besteht aus mehreren Runden. In jeder Runde ist eine vorgegebene Anzahl von Aufgaben zu bearbeiten. Ihre Auszahlung am Ende des Experiments hängt davon ab, welche Entscheidungen Sie treffen, wie viele Aufgaben Sie innerhalb einer Runde richtig lösen und möglicherweise davon, wie viele Aufgaben Teilnehmer von vorangegangenen Experimenten richtig gelöst haben. Der genaue Ablauf des Experiments, die von Ihnen zu treffenden Entscheidungen und die Auszahlungsmodalitäten werden im Folgenden genau erklärt.

Aufbau der Aufgaben

Bei jeder Aufgabe wird Ihnen ein kurzes Video gezeigt. Jedes der Videos besteht aus zwei Bildern, die abwechselnd eingeblendet werden. Die beiden Bilder eines Videos sind bis auf einen Bereich identisch. Bei dem Unterschied handelt es sich stets um ein oder mehrere zusammenhängende Objekte, die auf einem der Bilder vorhanden sind und auf dem anderen Bild vollständig oder teilweise fehlen. Ihre Aufgabe ist es, den Unterschied zwischen den Bildern zu finden.

Abbildung 1 zeigt ein Beispiel. Bild A des Beispiels enthält einen Kajakfahrer auf der linken Bildseite. In Bild B ist der Kajakfahrer nicht vorhanden.

Bild A Bild B

Abbildung 1. Beispiel eines Bildpaares.

Während des Videos wird jedes Bild für etwa eine Sekunde eingeblendet. Nach einer kurzen Pause folgt das nächste Bild. Der Ablauf der Bilder ist also: Bild A, Bild B, Bild A, Bild B, ... und so weiter. Die Gesamtdauer jedes Videos beträgt 14 Sekunden.

Nach Ablauf des Videos haben Sie 40 Sekunden Zeit, um eine Antwort einzugeben. Während dieser Eingabezeit wird ein Antwortraster eingeblendet (vgl. Abbildung 2). Das Antwortraster besteht aus einem Gitter mit nummerierten Feldern und dem Bild, welches das zu identifizierende Objekt enthält (Bild A im Beispiel).

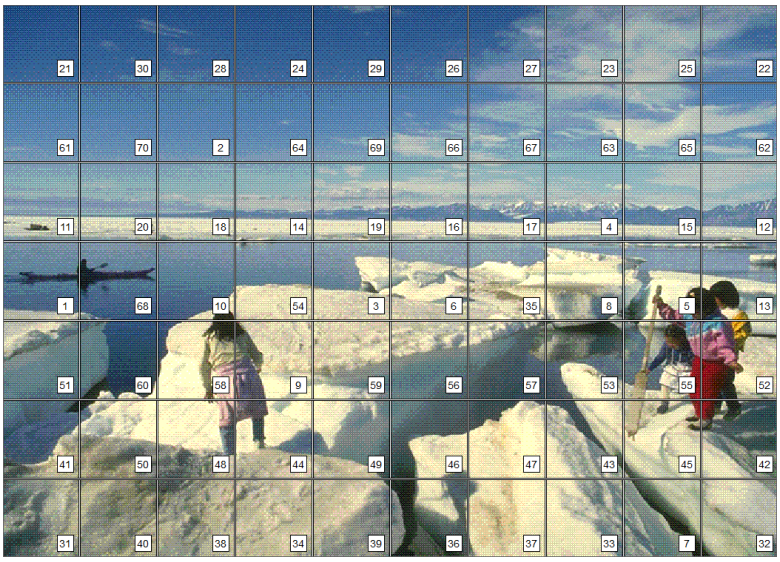

Abbildung 2. Beispiel eines Bildes mit Antwortraster

Unterhalb des Antwortrasters erscheint ein Eingabefenster. Hier können Sie eine der Feldnummern eingeben. Zur Lösung der Aufgabe, muss das Feld mit der eingegebenen Nummer den Unterschied zwischen den Bildern enthalten. Falls sich der Unterschied auf mehrere Felder bezieht, wird die Nummer jedes der Felder, die den Unterschied enthalten, als korrekt gewertet. Im obigen Beispiel (Abbildung 2)

geben Sie zur Lösung der Aufgabe also entweder “1” oder “68” ein. *Bitte beachten Sie, dass die gleichzeitige Eingabe von mehreren Feldnummern nicht gewertet wird.*

Zum Fortfahren klicken Sie auf “Weiter”. Sofern Sie eine Eingabe gemacht haben, wird diese hiermit gespeichert und anschließend zur Bestätigung auf dem Bildschirm angezeigt. Jetzt können Sie auf “Nächste Aufgabe” klicken, um die Aufgabe abzuschließen. Alternativ haben Sie mit dem Klick auf “Korrektur” die Möglichkeit, Ihre Eingabe innerhalb der vorgegebenen Eingabezeit zu korrigieren.

Bitte beachten Sie, dass ausschließlich die zuletzt erfolgte Eingabe gewertet wird. Wenn Sie also eine Korrektur erfolgreich abschließen, das heißt, eine neue Feldnummer in das Eingabefenster eingegeben und mit “Weiter” gespeichert haben, wird nur diese und nicht Ihre vorherige Eingabe gewertet. Wenn Sie eine Korrektur innerhalb der Eingabezeit nicht abschließen sollten, wird Ihre vorherige Eingabe gewertet.

Das Experiment wird fortgesetzt, sobald alle Teilnehmer auf “Nächste Aufgabe” gedrückt haben oder wenn die Eingabezeit abgelaufen ist.

Zuordnung zu Gruppen mit Teilnehmern vorangegangener Experimente

Alle Aufgaben, die Sie in diesem Experiment bearbeiten werden, wurden bereits von anderen Teilnehmern in vorangegangenen Experimenten bearbeitet.

Zu Beginn des Experiments werden Ihnen drei Teilnehmer aus den vergangenen Experimenten zufällig zugeordnet. Sie und die Ihnen zugeordneten Teilnehmer bilden eine Gruppe.

Für die gesamte Dauer des Experiments bleibt diese Zuordnung bestehen, das heißt, dass die drei Ihnen zugeordneten Teilnehmer, für die gesamte Dauer des Experiments dieselben bleiben. Die zufällig ausgewählten Teilnehmer vorheriger Experimente sind jedoch nicht die gleichen für alle Teilnehmer dieses Experiments.

Sie wissen selbstverständlich nicht, wer diese früheren Teilnehmer sind, sodass deren Anonymität vollständig gewahrt bleibt. Selbstverständlich bleibt auch Ihre Anonymität während des gesamten Experiments gewahrt.

Aufbau der Runden

Eine Runde besteht aus zehn Aufgaben, also zehn Videos, bei denen der Unterschied zwischen zwei Bildern zu identifizieren ist. Am Ende jeder Runde wird Ihre Platzierung im Verhältnis zu den Ihnen zugeordneten Teilnehmern bestimmt.

Die Platzierung erfolgt entsprechend der Anzahl richtig gelöster Aufgaben innerhalb der jeweiligen Runde. Wenn Sie mehr Aufgaben als die drei Ihnen zugeordneten Teilnehmer gelöst haben, sind Sie auf Platz 1. Wenn Sie die zweithöchste Anzahl an Aufgaben gelöst haben, sind Sie auf Platz 2. Wenn Sie die dritthöchste Anzahl an Aufgaben gelöst haben, sind Sie auf Platz 3. Wenn Sie im Vergleich zu den Ihnen zugeordneten Teilnehmern die wenigsten Aufgaben gelöst haben, sind Sie auf Platz 4.

Wenn Sie die gleiche Anzahl an Aufgaben wie einer oder mehrere der Ihnen zugeordneten Teilnehmer gelöst haben, wird Ihre Platzierung im Verhältnis zu diesen Teilnehmern zufällig bestimmt.

Wenn Sie beispielsweise die gleiche Anzahl an Aufgaben gelöst haben wie einer der Ihnen zugeordneten Teilnehmer und die beiden anderen Ihnen zugeordneten Teilnehmer weniger Aufgaben als Sie gelöst haben, dann haben Sie die gleiche Chance, den 1. Platz zu erreichen, wie der Ihnen zugeordnete Teilnehmer mit der gleichen Anzahl richtig gelöster Aufgaben.

Natürlich erfolgt der Vergleich Ihrer Punktzahl mit denen der Ihnen zugeordneten Teilnehmer auf Grundlage derselben Videos. Das heißt, Ihre Anzahl an richtig gelösten Aufgaben in einer Runde wird mit der Anzahl an richtig gelösten Aufgaben der anderen Teilnehmer für genau dieselben Videos vorgenommen.

Bedingungen für die Ihnen zugeordneten Teilnehmer

Bei der Bearbeitung der Aufgaben hatten die Ihnen zugeordneten Teilnehmer *die gleichen Bedingungen wie Sie*. Das heißt, sie hatten die gleiche Zeit wie Sie für die Bearbeitung jeder Aufgabe, also 14 Sekunden zur Betrachtung jedes Videos und maximal 40 Sekunden für die Eingabe der Antwort.

Die Ihnen zugordneten Teilnehmer haben jedoch alle Aufgaben, die Sie in diesem Experiment sehen werden – einschließlich der Aufgaben Ihrer Beispielrunde – unter Wettbewerbsbedingungen bearbeitet. Das heißt, sie erhielten nur dann eine Bezahlung für eine Runde, wenn Sie innerhalb ihrer Gruppe Platz 1 belegten.

Ablauf und Bezahlung

Der Ablauf des Experiments ist folgender:

1. Entscheidungsphase 1
2. Beispielrunde (10 Aufgaben)
3. Informationen über die Ihnen zugeordneten Teilnehmer
4. Entscheidungsphase 2: Wahl Ihres kritischen Stücklohns

5. Bekanntgabe des Stücklohns im Umschlag
6. Runde 1 (10 Aufgaben): Stücklohn
7. Runde 2 (10 Aufgaben): Wettbewerb
8. Runde 3 (10 Aufgaben): Stücklohn oder Wettbewerb
9. Fragebogen
10. Zufallsziehungen zur Bestimmung Ihrer Auszahlung
11. Feedback und Auszahlung

1. **Entscheidungsphase 1**

Zu Beginn des Experiments werden Sie mit einer Reihe von Entscheidungsproblemen konfrontiert. Am Ende des Experiments wird eine Ihrer Entscheidungen zufällig ausgewählt, die dann auszahlungsrelevant ist. Genauere Erläuterungen zu den Entscheidungen und zum Auszahlungsmodus erhalten Sie während des Experiments.

2. **Beispielrunde**

Im Anschluss an Entscheidungsphase 1 nehmen Sie an einer Beispielrunde teil. Hier werden Ihnen zehn Aufgaben in gleicher Art und Weise wie in den folgenden drei Runden gezeigt. In der Beispielrunde können Sie jedoch kein Geld verdienen, das heißt, Ihre Eingaben werden nicht gewertet. [In der Beispielrunde erhalten Sie zu Beginn jeder Aufgabe ein Hinweisbild zur Lösung der Aufgabe. Während der folgenden auszahlungsrelevanten Runden werden Sie jedoch keine Hinweisbilder erhalten.][54]

Sowohl die Zuordnung der Videos zur Beispielrunde und den drei Runden als auch die Reihenfolge der gezeigten Videos innerhalb jeder Runde sind das Resultat einer Zufallsziehung. Bei dieser Zufallszuordnung wurde nicht berücksichtigt, in welcher Reihenfolge die Ihnen zugeordneten Teilnehmer die Aufgaben gelöst haben. Auch für diese Teilnehmer wurde die Anordnung der Videos zufällig bestimmt.

Da sich die Aufgaben jedoch im Schwierigkeitsgrad nicht vollständig gleichen, wurde bei der Zufallsziehung für dieses Experiment ein einheitlicher Schwierigkeitsgrad über die Runden hinweg sichergestellt. Das heißt sowohl die Beispielrunde als auch die darauffolgenden drei Runden enthalten den gleichen Anteil an leichteren bis schwereren Aufgaben.

Die Zufallsziehung der Videos ist für alle Teilnehmer dieses Experiments die gleiche, das heißt, jeder Teilnehmer dieses Experiments sieht zu jedem Zeitpunkt des Experiments das gleiche Video.

[54] These sentences have been excluded in the instructions for the uninformed treatment.

3. **Informationsphase**

Hier erhalten Sie Informationen über die drei Ihnen zugeordneten Teilnehmer in Bezug auf die Beispielaufgaben: Für jeden der drei Ihnen zugeordneten Teilnehmer erfahren Sie, wie viele der zehn Beispielaufgaben dieser Teilnehmer richtig gelöst hat. *Bitte beachten Sie, dass die Ihnen zugeordneten Teilnehmer Ihre Beispielaufgaben (sowie die Aufgaben in Runde 1-3) unter Wettbewerbsbedingungen bearbeitet haben*, das heißt, sie erhielten nur dann Auszahlungen für eine Runde, wenn sie innerhalb ihrer Gruppe Platz 1 belegten.

4.-8. **Entscheidungsphase 2, Bekanntgabe des Stücklohns und Runde 1-3**

Für die Bearbeitung der Aufgaben in den Runden 1 bis 3 erhalten Sie 7,– Euro. Zusätzlich wird am Ende des Experiments eine der Runden 1–3 zufällig gezogen. In den drei Runden gibt es zwei Varianten, nach der Ihre Auszahlung berechnet wird.

Runde 1: Stücklohn

Vor Beginn dieses Experiments wurde ein Stücklohn festgelegt, dessen Höhe in dem an der Tür befestigten Umschlag verzeichnet ist. Die Höhe dieses Stücklohns liegt zwischen 0,01 Euro und 2,– Euro. Der Umschlag wird nach Entscheidungsphase 2, also vor Runde 1 geöffnet und durch einen der Experimentleiter in einen Ihrer Computer eingegeben.

Wenn Runde 1 als Auszahlungsrunde ausgelost wird, erhalten Sie den im Umschlag enthaltenen Stücklohn für jede von Ihnen in dieser Runde richtig gelöste Aufgabe.

Runde 2: Wettbewerb

Wenn Runde 2 als Auszahlungsrunde ausgelost wird, wird Ihre Auszahlung folgendermaßen berechnet: Sie erhalten 2,– Euro für jede richtig gelöste Aufgabe, wenn Sie in dieser Runde Platz 1 erreichen, andernfalls erhalten Sie 0,– Euro.

Runde 3: Stücklohn oder Wettbewerb

Wenn Runde 3 als Auszahlungsrunde gewählt wird, hängt Ihre Auszahlung von Ihrer Entscheidung in der Entscheidungsphase 2 ab, in der Sie Ihren kritischen Stücklohn festlegen.

Die Wahl Ihres kritischen Stücklohns wirkt sich folgendermaßen auf Ihre Auszahlungsberechnung in Runde 3 aus:

- Falls der im Umschlag enthaltene Stücklohn kleiner ist als der von Ihnen gewählte kritische Stücklohn, dann wird Ihre Auszahlung entsprechend der

Wettbewerbsbedingungen berechnet. Sie erhalten dann also 2,– Euro für jede von Ihnen in dieser Runde richtig gelöste Aufgabe, wenn Sie in dieser Runde Platz 1 erreichen, andernfalls erhalten Sie 0,– Euro.

- Falls der im Umschlag enthaltene Stücklohn mindestens so hoch ist, wie der von Ihnen gewählte kritische Stücklohn, dann wird Ihre Auszahlung entsprechend dem im Umschlag enthaltenen Stücklohn berechnet. Das heißt, Sie erhalten den im Umschlag enthaltenen Stücklohn für jede von Ihnen in dieser Runde richtig gelöste Aufgabe.

Während des Experiments soll Ihnen eine graphische Oberfläche bei der Entscheidungsfindung behilflich sein. Der Startwert des kritischen Stücklohns in dieser graphischen Entscheidungshilfe wird für jeden Teilnehmer zufällig durch den Computer bestimmt.

Fragen während des Experiments

Im Verlauf des Experiments werden Ihnen an verschiedenen Stellen insgesamt neun Fragen gestellt. Hier können Sie zusätzlich Geld verdienen. Am Ende des Experiments wird eine dieser neun Fragen zufällig ausgewählt. Wenn Sie diese Auszahlungsfrage richtig beantwortet haben, erhalten Sie 1,– Euro. Instruktionen zu den genauen Auszahlungsmodalitäten der einzelnen Fragen erhalten Sie während des Experiments.

9. Fragebogen

Im Anschluss an die drei Runden erhalten Sie einen Fragebogen. Auch wenn Ihre Angaben in diesem Teil keinen Einfluss auf Ihre Auszahlung haben, möchten wir Sie bitten, die Fragen gewissenhaft zu beantworten.

10.-11. Zufallsziehungen, Feedback und Auszahlung

Am Ende des Experiments werden eine Ihrer Entscheidungen in Entscheidungsphase 1, eine Ihrer Antworten auf die neun Fragen während des Experiments sowie die Auszahlungsrunde (für alle Teilnehmer die gleiche) zufällig gezogen. Anschließend werden Ihre hieraus folgende Auszahlungen sowie Ihre Gesamtauszahlung auf dem Bildschirm angezeigt. Darüber hinaus erhalten Sie genauere Informationen zu allen drei Runden – insbesondere die Anzahl der von Ihnen in jeder Runde richtig gelösten Aufgaben sowie den Platz, den Sie in jeder Runde belegt haben.

Ihre Gesamtauszahlung für dieses Experiment besteht somit aus

1. Ihrem Erscheinungsentgelt (5,– Euro),

2. Ihrer Bezahlung für die Bearbeitung der Aufgaben (7,– Euro),
3. Ihrer Bezahlung entsprechend einer Entscheidung in Entscheidungsphase 1,
4. Ihrer Bezahlung entsprechend einer Ihrer Antworten auf die Fragen während des Experiments sowie
5. Ihrer Bezahlung entsprechend der zufällig ausgewählten Auszahlungsrunde.

Haben Sie Fragen? Wenn ja, zeigen Sie dies bitte per Handzeichen an. Wir werden Ihre Fragen dann einzeln beantworten.

Wir danken Ihnen für die Teilnahme an diesem Experiment!