The Biases of Others: Anticipating Informational Projection in an Agency Setting

David N. Danz† Kristóf Madarász‡ Stephanie W. Wang§

November 9, 2017

Abstract

Evidence shows that people fail to account for informational differences and instead project their information onto others in that they too often act as if others had access to the same information they did. In this study, we find that while people naively project their information onto others, they also anticipate the projection of their differentially informed opponents onto them. Specifically, we directly test the model of projection equilibrium, Madarasz (2014), a single-parameter extension of common prior BNE which posits a tight continuous relationship between the extent to which people project onto others and the extent to which they anticipate the projection of others onto them. Consistent with the theory, we find not only that on average better-informed principals exaggerate the extent to which lesser informed agents should act as if they were better-informed, but that on average lesser-informed agents anticipate but underestimate such misperceptions as revealed by their choice of incentive scheme and elicited second-order beliefs. Furthermore, we estimate the distribution of the extent to which players project onto others and the distribution of the extent to which players

*We are grateful for valuable comments by Colin Camerer, Gary Charness, John Duffy, Ignacio Esponda, Dorothea Kübler, Muriel Niederle, Demian Pouzo, Andrew Schotter, Adam Szeidl, Lise Vesterlund, Axel Werwatz, and Alistair Wilson. Financial support from the Deutsche Forschungsgemeinschaft (DFG) through CRC 649 “Economic Risk” is gratefully acknowledged.

†University of Pittsburgh, 4924 Wesley W. Posvar Hall, 230 South Bouquet Street, Pittsburgh, PA 15260 and WZB Berlin Social Science Center, danz@pitt.edu.

‡London School of Economics and Political Science, WC2A 2AE, UK, k.p.madarasz@lse.ac.uk.

§University of Pittsburgh, 4715 Wesley W. Posvar Hall, 230 South Bouquet Street, Pittsburgh, PA 15260, swwang@pitt.edu.
fail to recognize others’ projection onto them and find that the relationship between the two is remarkably consistent with that implied by projection equilibrium.

**Keywords:** information projection, defensive agency, curse of knowledge, hindsight bias, higher-order beliefs

1 Introduction

Over the past decades significant effort has been devoted to documenting domains where a person’s choice departs from classical notions of rationality. At the same time, less attention has been devoted to the issue of understanding the extent to which people anticipate such departures, in particular, little is known about whether people anticipate or mispredict such departures in others. In the context of an individual bias, one’s own departure from rationality can be studied separately from one’s anticipation of the same kind of departure by others. In the context of a general social bias, as considered by this paper, where the failure of rationality lies in a person’s belief about the beliefs of others, such a separation need not be possible.

In particular, this paper considers the essential psychological domain of the theory of mind, e.g., Piaget (1956), that is, the way people form beliefs about the beliefs of others. The phenomenon of informational projection (Madarasz 2012, 2016) formalizes a general shortcoming in this domain whereby people engage in limited perspective taking and act as if they attributed information to others that are too close to their own. For example, a theory teacher who exhibits the curse-of-knowledge and exaggerates the amount of background knowledge her students have tends to explain the material in terms that are too complex for her students to understand.

In many economically relevant situations what matters is not simply what beliefs a person has about the information of others, but also what she thinks about others’ think her beliefs and information is. In other words, specifying such second-order perceptions is key. A poker player may exaggerate the extent to which others know her hand, but does she also think that others wrongly assume that she might know their hands? A game theory teacher who exhibits the curse of knowledge too often thinks that her students should already know what a Nash equilibrium is. Does she, however, anticipate that her students too often expect her to know that they have very little idea about Nash equilibrium? Will this not directly contradict the first-order implications of the curse-of-knowledge? In all these, and
many other settings, specifying the extent to which one anticipates the mistakes of others becomes part of the very definition of one’s own mistake. In a social context, studying empirically the extent to which people anticipate each other’s biases is essential to study the nature and the form of the bias itself.

Answering the question whether or not people anticipate the biases of others is important not only to provide a coherent and empirically plausible account of projection per se, but has directly relevant implications for a variety of economic outcomes. For example, in the fundamental agency problem characterizing organizations, if an agent anticipates that monitoring with ex-post information will bias a principal’s assessment of her and adversely affect her reputation, she may try to manipulate the production of ex-ante information or simply dis-invest from an otherwise efficient relationship. Thus, classical tools that could reduce agency costs under the assumption of no such anticipation could offer perverse incentives and instead exacerbate inefficiencies.

A widely discussed example of such defensive agency can be found in the context of medical malpractice liability. The radiologist Leonard Berlin, in his 2003 testimony on the regulation of mammography to the U.S. Senate Committee on Health, Education, Labor, and Pensions, describes how ex-post information causes the public to misperceive the ex-ante accuracy of mammograms, implying that juries are “all too ready to grant great compensation.” Berlin references the role of information projection in such ex-post assessments, where ex-post information makes reading old radiographs much easier.¹ In response, physicians are reluctant to follow such crucial diagnostic practices: “The end result is that more and more radiologists are refusing to perform mammography [...] In turn, mammography facilities are closing.”

In the context of such systematic false beliefs that do not follow the tenants of common priors standard strategic models provide no particular guidance. A direct and simple response to the presence of higher-order considerations in contexts where projection may be key is to assume a binary classification of people into biased ones who also fail to anticipate the biases of others and unbiased ones who, in contrast, do anticipate the biases of others. While this still naive classification is still under-specified as it leaves key higher-order considerations open, e.g., a player’s belief about her opponent’s belief of her degree of projection, it resolves the apparent contradiction be-

¹As Berlin (2003) points out “Suffice it to say that research studies performed at some of the most prestigious medical institutions of the United States reveal that as many as 90% of lung cancers and 70% of breast cancers can at least be partially observed on studies previously read as normal.”
tween projecting information and anticipating the projection of others and may also prove to be a useful first approximation. In another vein, one may also suppose that the extent to which the average person project onto others in a context provides no prediction as to the average person’s anticipation of others’ projection in that context.

In contrast, the model of Madarasz (2014) offers a more continuous and fully specified equilibrium approach, *projection equilibrium*. This admits the behavioral types of the above binary specifications as two limiting extremes. This model provides a general and fully specified approach. At the same time, by tying together the extent to which people project their information onto others (first-order projection) and the extent to which they fail to anticipate others’ projecting onto them (second-order projection), it provides a still very parsimonious alternative. It offers a single-parameter extension of the common-prior Bayesian Nash Equilibrium and implies a measure under which the extent to which people are biased is directly and inversely related to the extent to which they anticipate the biases of others. Specifically, under the model of projection equilibrium, a person who projects to degree $\rho$ acts as if she (wrongly) believed that her opponent knew exactly her information, had correct beliefs about her beliefs about his information and so on with probability $\rho$ and, with the remaining probability $1 - \rho$, she has correct beliefs about her opponent’s information, correct beliefs about her opponent’s (wrong) beliefs about her information and so on. A tight implication of the model is that if a person acts as if she mistakenly believed that others knew her information with probability $\rho$, first-order projection, she also acts as if she underestimated the analogous beliefs of others by probability $\rho^2$, second-order projection. It predicts that the extent to which people project onto others is inversely related to the square-root of the extent to which they fail to anticipate how much others project onto them. The extent to which this single-parameter model provides greater explanatory power than a more naive alternative is ultimately an empirical question.

This paper develops a simple experimental design to directly test this theory. Moreover, the design not only allows one to test the equilibrium model against the perfect Bayesian equilibrium with correct priors, but is able to assess the relevance of the continuous specification over a more naive dichotomous classification of people into biased (unanticipating) and unbiased (fully anticipating) types. In addition, it is also able to tightly describe the extent to which a single-parameter parsimonious specification which directly links the first-order and the second-order mistake is able to capture key aspects of the data.
In our experiments, *principals* (evaluators) estimated the average performance of *agents* in a real-effort task. The principals received the solution to each task prior to the estimation in the *informed treatment* but not in the *uninformed (control) treatment*. Agents are randomly matched with either informed or uninformed principals depending on the treatment. If a principal exaggerates the extent to which agents act as if they had the same information as she did, then on average a principal in the informed, but not in the uninformed treatment, should overestimate the agents’ performance. The difference between the principals’ performance estimations in the treatment vs. control then allows us to identify the extent of the principals’ information projection in first-order beliefs. We do find strong evidence of first-order informational projection by principals in our experiments, consistent with previous results.

For each task, the agents could choose between a sure payoff and an investment whose payoff was decreasing in the principal’s expectation of the success rate in the task. If the agents anticipate that informed principals overestimate their performance on average, vis-a-vis the estimates in the control treatment, then they would prefer the sure payoff, which is not dependent upon the principals’ estimation, more in the informed than in the uninformed treatment. We find that 67.3% of agents matched with informed principals as opposed to 39.2% of those matched with uninformed principals chose the investment whose payoff was decreasing in the principal’s estimation.

To further explore the issue of anticipation in beliefs, we also directly elicited the agents’ both first-order (their own estimate of the success rate) and second-order beliefs (their estimate of the principals’ estimates of the success rate). Indeed, we find clear evidence that agents anticipate that principals are biased by private information. In particular, agents’ first-order beliefs are well-calibrated on average in both treatments. Furthermore, agents’ second-order beliefs do not differ from their first-order beliefs in the uninformed treatment, at the same time, their second-order beliefs are significantly higher than their own first-order beliefs in the informed treatment. These findings lend direct support to projection equilibrium.

We then use the estimates by informed principals and the first- and second-order beliefs of agents matched to informed principals to estimate projection equilibrium, which nests the Bayesian Nash equilibrium as a special case. Under projection equilibrium, a person simultaneously projects onto others, and to a lesser extent—where this lesser extent is also governed by projection itself—she anticipates others will project onto her. The fact that such anticipation under projection is limited stems from the fact that
a person who projects too often thinks others not only have the same basic perspective as she does, but also a truthful perspective (second-order belief) about her perspective as well. Consistent with the predictions of projection equilibrium, we find the exaggeration anticipated by the agents matched with informed principals is positive—but, as the model predicts, less than the actual exaggeration exhibited by the informed principals. We estimate the single-parameter version of projection equilibrium and find significant information projection. Crucially, we then relax the assumption that the principal and the agent must exhibit the same degree of projection, that is, that the degree to which people project their information on others is the square-root of the extent to which people underestimate others’ projection onto them, and nevertheless obtain very similar parameter estimates and log-likelihoods. Finally, we also compare the distribution of first-order projection in the population and the distribution of the under-anticipation, that is, the extent to which agents underestimate the principal’s overestimation. We find that these two distributions are remarkably close to each other.

The paper is structured as follows. In Section 2, we present the experimental design and procedures. Section 3 enumerates the model and the main predictions. Section 4 contains the results. In Section 5, we conclude with a discussion of related models and future directions.

2 Experimental Design

2.1 Experimental task

All participants worked on the same series of 20 information-projection stimuli: In each of 20 change-detection tasks, the subjects had to find the difference between two otherwise identical images (see Rensink et al., 1997; Simons and Levin, 1997). Figure 1 shows an example, where the difference is located in the upper-right corner of the images. While the difference between these images is quite hard to detect, it seems obvious once you know where to look. In turn, people who know the solution in the first place, cannot “unsee” it and hence tend to overestimate how likely others will find it (see, e.g., Loewenstein et al. 2006).²

We presented each task in a 14-second video clip in which the two images were displayed alternately with short interruptions.³ Afterwards, subjects

²Further information-projection stimuli typically used in experiments are prediction tasks (in various contexts), logic puzzles, and trivia questions.

³Each image was displayed for one second, followed by a blank screen for 150 milliseconds.
had 40 seconds to submit an answer, i.e., to indicate the location of the difference.\footnote{See the instructions in the Appendix for more details.}

\subsection{Principals}

Principals had to estimate the performance of others in the change-detection tasks. Specifically, the principals were told that subjects in previous sessions worked on the tasks and that these subjects (reference agents henceforth) were paid according to their performance.\footnote{The performance data of 144 reference agents was taken from Danz (2014). In their experiment, the subjects performed the tasks in winner-take-all tournaments, where they faced the tasks in exactly the same way as the subjects in the current experiment.}

In each of 20 rounds, the principals were first exposed to the task; that is, they watched the 14-second video clip and then had 40 seconds to submit a solution to the task. Afterwards, the principals stated their belief ($b^P_t$) about the fraction of reference agents who spotted the difference in that task (success rate $\pi_t$ henceforth). After each principal stated his or her belief, the next round started.\footnote{The principals first participated in three practice rounds to become familiar with the interface.}

For the principals the two treatments differed as follows. In the informed treatment, the principals received the solution to each task before they went through the change-detection task. Specifically, during a countdown phase that announced the start of each task, the screen showed the image (one of the two in the video) with the difference highlighted with a red circle (see Figure 2). In the uninformed treatment, the principals were not given
solutions to the tasks (the same image was shown on the countdown screen, but without the red circle and the note). Principals in both treatments then went through each task exactly as the reference agents did. The principals did not receive feedback of any kind during the experiment.

At the end of the sessions, the principals received €0.50 for each correct answer in the uninformed treatment and €0.30 in the informed treatment. In addition, they were paid based on the accuracy of their stated beliefs in two of the 20 tasks (randomly chosen): for each of these two tasks, they received €12 if \( b_t^P \in [\pi_t - 0.05, \pi_t + 0.05] \), that is, if the estimate was within 5 percentage points of the true success rate of the agents. We ran one session with informed principals, and one with uninformed principals with 24 participants in each.

### 2.3 Agents

Agents in both treatments were informed that in previous sessions (i) reference agents had performed the change-detection tasks being paid according to their performance and (ii) principals had estimated the average performance of the reference agents being paid according to the accuracy of their estimates. The agents where further told that they had been randomly matched to one of the principals at the outset of the experiment and that this matching would remain the same for the duration of the experiment. For the agents, the two treatments differed only with respect to the kind
of principal they where matched to: In the informed treatment, agents were randomly matched to one of the informed principals; in the uninformed treatment, agents where randomly matched to one of the uninformed principals. In both treatments the agents where made fully aware of the information their principal received (but, of course, not of the information the principals received in the other condition, i.e., the existence of another treatment).

In each of 20 rounds, the agents in both treatments first performed the task in the same way as the reference agents; that is, they watched the 14-second video clip and then had 40 seconds to submit an answer. Next, agents received feedback regarding the solution to the tasks. Specifically, the screen showed one of the two the images with the difference highlighted with a red circle; then, the video clip was shown again. Agents matched to informed principals were told this feedback corresponded to what the principal had seen for that task. Agents matched to uninformed principals were told the principals had not received the solution to the task. In both treatments, the agents did not receive information about the principal’s estimates.

Afterward, the agent chose between two options. One of the options provided a sure payoff of €4. The payoff of the other option depended on the success rate of the reference agents, \( \pi_t \), and the corresponding estimate of the principal matched to her. Specifically, the agent received €10 if the principal’s estimate \( b_p^t \) was not more than 10 percentage points higher than the success rate \( \pi_t \); otherwise, the payoff was €0. Choosing this option can be thought of an investment whose expected return is decreasing in the principals’ expectations of the agents’ likelihood of success. Similarly, choosing the sure payoff can be thought of buying insurance against overly optimistic expectations of the principals. Throughout the paper, we will refer to this choice as the agents’ investment decision.

In terms of our motivating example of defensive medicine, the principal in the informed treatment can be thought of a judge or juror who, in hindsight, forms expectations about the radiologist’s ex ante chance of correctly diagnosing a patient based on a radiograph. At the time of her evaluation, the juror has access to both the initial radiograph and further information about the patient’s condition revealed later on. The biased juror projects this data on the radiologist’s information set when diagnosing the patient and consequently overestimates her ex ante chances of success. In turn, the physician who anticipates this bias will exhibit a higher willingness to pay for insurance against malpractice claims or simply abandons career paths that are particularly susceptible to biased ex post performance evaluations.
The agents received €0.50 for each correct answer to the change-detection tasks. In addition they were also paid according to the outcome of one randomly selected investment decision. We ran two sessions of agents matched to informed principals (24 participants per session) and two session with agents matched to uninformed principals (23 participants per session).

To explore the agents’ beliefs that underlie their investment decisions, we ran additional sessions: one with agents matched to informed principals (24 participants) and one with agents matched to uninformed principals (23 participants). The sessions differed from the sessions without belief elicitation in that belief elicitation replaced the investment decisions for the first ten tasks of the experiment. Specifically, following each of the first 10 change-detection tasks (and the feedback regarding the solution), the agents stated (i) their belief about the fraction of reference agents that spotted the difference in that task (first-order belief $b_{A,1}^{1,t}$ henceforth) and (ii) their belief about their principal’s estimate of that success rate (second-order belief $b_{A,2}^{1,t}$ henceforth).

At the end of the experiment one round was randomly selected for payment. If this round involved belief elicitation, we randomly selected one of the agent’s stated beliefs for payment, namely, either her first- or second-order belief in that round.\footnote{This payment structure addresses hedging concerns (Blanco et al., 2010).} The subject received €12 if her stated belief was within five percentage points of the actual value (the actual success rate in case of a first-order belief and the principal’s estimate of that success rate in case of a second-order belief), and nothing otherwise.\footnote{We chose this elicitation mechanism because of its simplicity and strong incentives. In comparison, the quadratic scoring rule is relatively flat incentive-wise over a range of beliefs, and its incentive compatibility is dependent on assumptions about risk preferences (Schotter and Trevino, 2014). The Becker-DeGroot-Marschak mechanism can be confusing and misperceived (Cason and Plott, 2014). The beliefs we elicited were coherent and sensible.} If the round selected for payment involved an investment decision, the agent was paid according to her decision. Table 1 provides an overview of the treatments and sessions.

\subsection*{2.4 Procedures}

The experimental sessions were run at the Technische Universität Berlin in 2014. Subjects were recruited with ORSEE (Greiner, 2004). The experiment was programmed and conducted with z-Tree (Fischbacher, 2007). The average duration of the principals’ sessions was 67 minutes; the average earning was €15.15. The agents’ sessions lasted 1 hour and 45 minutes.
Table 1: Overview of treatments and sessions.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Informed</th>
<th>Uninformed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principals</strong></td>
<td>Elicitation of first-order beliefs regarding the success rate of reference agents in 20 change-detection tasks</td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>Solution to change-detection tasks</td>
<td>No solution to change-detection tasks</td>
</tr>
<tr>
<td>Sessions</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Subjects</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>Agents</strong></td>
<td>Choices between a sure payoff and investment whose expected return is decreasing in principals’ expectations;</td>
<td></td>
</tr>
<tr>
<td><em>Elicitation of first-order beliefs (estimate of the success rate) and second-order beliefs (estimates of the principals’ estimate)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>Principals have access to solutions</td>
<td>Principals do not have access to solutions</td>
</tr>
<tr>
<td>Sessions</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Subjects</td>
<td>48 (12 + 12 + 24*)</td>
<td>46 (11 + 12 + 23*)</td>
</tr>
</tbody>
</table>

Note: Asterisks indicate sessions with belief elicitation instead of investment decisions in the first half of the experiment.

The average duration of the sessions (the average payoff) in the treatments with and without belief elicitation was 115 and 96 minutes (€21.47 and €19.10), respectively. Two participants did not complete the comprehension questions and were excluded from the experiment.

3 Predictions

The theoretical framework for our predictions is based on the notion of a projection equilibrium introduced in Madarász (2014a). Below, we state the predictions of projection equilibrium for our design. Fix a task. Let there be
a finite state-space \( \Omega \), with generic element \( \omega \), and a common prior \( \phi \) over this space. Player \( i \)’s information is generated by an information partition \( P_i : \Omega \rightarrow 2^\Omega \). Player \( i \)’s action set is given by \( A_i \). Finally, player \( i \)’s payoff function is given by \( u_i(\omega, a) : \Omega \times A \rightarrow \mathbb{R} \). The true game is thus described by \( \Gamma = \{ \Omega, \phi, \{ P_i \}_i, \{ A_i \}_i, \{ u_i \}_i \} \).

Solving the basic task in our design is equivalent to picking a cell \( x \in D \) from the finite grid \( D \) on the visual image. A solution is a success if the chosen cell contains the difference.\(^{11}\) Because the payoff from the basic task and the payoffs from other decisions (these are always about population averages) do not directly interact, we can denote the former payoff by \( f(\omega, x) \). We normalize this to be 1 if the solution is a success, and 0 otherwise. Given this normalization, the value of subject \( i \)’s maximization problem,

\[
\max_{x \in D} E[f(\omega, x) \mid P_i(\omega)],
\]

corresponds to the probability with which player \( i \), given information \( P_i(\omega) \), successfully solves the task. Finally, the action set for the principal, which includes the estimation task, is \( A_P = D \times [0, 1] \). Similarly, the action set for the agent, which includes the estimation tasks and the investment task, is \( A_A = D \times [\text{Invest, Not Invest}] \times [0, 1]^2 \).

The information a subject obtains in the uninformed treatment contains the inference from watching the video. In the informed treatment, in addition to this exact information, the principal also receives the solution. Given a general state-space, we can allow for the fact that players make inferences not only about the solution to the basic task, but also more generally, about other characteristics of the task. Furthermore, we can also allow for the fact that different players may make different inferences from watching the video. In other words, each player may privately observe a different realizations of the same signal-generating process. Thus, even in the uninformed treatment, any two players \( j \neq k \) may obtain private signal realizations and thus have private information generated by watching the video. The identifying restriction will be that the prior distribution of these realizations, given the same fixed signal generating technology, i.e., the video, is the same for each randomly chosen subject ex-ante..

**Predictions** We can now turn to the predictions of \( \rho \) projection equilibrium for our design. It is important to emphasize that the predictions hold from an ex-ante expected perspective, that is, on average. This gives\(^{11}\)

---

\(^{11}\)We explain to the subjects prior to them solving the task, that when the difference is contained in multiple cells, identifying any of them is equivalent to a correct solution.
power for our design since it allows us to test the predictions of the model without making additional assumptions on the data generating process.\footnote{This fact also allows us to immediately separate the predictions of the model to potential alternative explanation in terms of cursed equilibrium, Eyster and Rabin (2005) or analogy-based expectations equilibrium, Jehiel (2005) or Jehiel and Koessler (2008) because all of these approaches are based on the identifying assumption that people have correct predictions about the average behavior of their opponents.}

A few remarks are in order.

1. First, under correct expectations, $\rho = 0$, the predictions are identical to that of the predictions of BNE. Here, the martingale property of Bayesian beliefs holds, which implies that the ex-ante expected posterior always equals the prior. Therefore, the treatment differences described below are expected to be zero.

2. Second, since a biased agent (or principal) projects her information, a player who does figure out the solution, versus one who does not, can have systematically biased conditional estimates of the information of others. However, below we do not focus on such conditional estimates, which are also affected by selection, but instead on the expected average estimates within a treatment. The predictions described thus express the \textit{ex-ante} expected differences across treatments.

Consider first the ex-ante probability with which a randomly chosen agent can solve the task. Denote this success rate by

$$E_\omega[\max_{x \in D} E[f(\omega, x) \mid P_A(\omega)] = \pi. \quad (2)$$

A key variable for our predictions is the ex-ante expected difference between this probability and the success probability of a randomly chosen principal in a given treatment. Formally, consider the following variable:

$$d = E_\omega[\max_{x \in D} E[f(\omega, x) \mid P_A(\omega)] - E_\omega[\max_{x \in D} E[f(\omega, x) \mid P_A(\omega)]] \quad (3)$$

In the uninformed treatment, both an agent and a principal has access to the video only. Given that roles are determined randomly by sampling from the same population, the above expected difference must be zero by construction. By contrast, in the informed treatment, where a principal has access to the solution of the basic task in addition to the video, the above variable will be positive (Blackwell 1953). Here, $d$ expresses the ex-ante expected value of having the solution, relative to the video only, for the basic task.
Below, we allow for role-specific degrees of projection, that is, the degree of projection is \( \rho = (\rho_P, \rho_A) \) where \( \rho_P \) denotes the degree to which the principal projects and \( \rho_A \) the degree to which the agent project. At the end of this section, we also restate the predictions of projection equilibrium for the case where projection is homogenous, i.e., \( \rho_P = \rho_A \), and thus the degree of projection is independent of the role a subject takes in the game.

Consider first a \( \rho_P \)-biased principal. The claim below describes the principal’s estimate of the agents’ success rate.

**Claim 1** (Principal’s estimate). *The principal’s expected estimate of the success rate is given by \( \pi + \rho_P d \).*

*Proof.* See Appendix. \( \square \)

In the uninformed treatment, the principal’s estimate of the success rate is correct on average. A principal who solves the task and one who does not might have systematically different and wrong conditional beliefs about the agent’s success probability. In the uninformed treatment, however, these conditional mistakes cancel out on average. This is true because the ex-ante probability with which a randomly chosen subject—indeed of the role assigned—figures out the solution is the same. In contrast, in the informed treatment, the principal always has access to the solution. Because a biased principal who knows the solution to the basic task acts as if she believed that with probability \( \rho_P \) the agent must know the solution for sure, the principal now exaggerates the agent’s success rate on average. This exaggeration is increasing both in the value of the principal’s additional information for the basic task, \( d \), and in the principal’s degree of projection, \( \rho_P \).

We can now turn to the key prediction of the model and state the partial anticipation feature of projection equilibrium in our setup. Consider now a \( \rho_A \)-biased agent. Given projection of degree \( \rho_P \) by the principal, the second claim implies a systematic difference between an agent’s estimate of the principal’s estimate and his own estimate of the average success rate in the informed treatment, but not in the uninformed treatment, on average. It predicts that the difference between such second-order and first-order estimates should be greater in the informed treatment than in the uninformed treatment.

**Claim 2** (Agent’s estimate). *The expected difference between the agent’s estimate of the success rate and the agent’s estimate of the principal’s mean estimate of the success rate is given by \( (1 - \rho_A)\rho_P d \).*

*Proof.* See Appendix. \( \square \)
The above claim establishes that the agent anticipates the principal’s projection but underestimates its extent on average. In more detail, the agent’s first-order estimate of $\pi$ is unbiased in both treatments on average. In the uninformed treatment, the same holds for the second-order estimate, that is, the agent’s expected estimate of the principal’s mean estimate as well. These facts are true for the same reason as in Claim 1: the agent and the principal here are in a symmetric situation, each seeing only the video. By contrast, in the informed treatment, because $d > 0$, the agent’s estimate of the principal’s estimate is higher than his own on average. Let us describe this logic in more detail.

The agent assigns probability $(1 - \rho_A)$ to the true distribution of the principal’s strategy. Hence in the informed treatment the agent anticipates that the principal exaggerates the success rate by degree $\rho_P$ on average. At the same time, the agent is also biased and assigns probability $\rho_A$ to the principal—and also all other agents—being the projected version and thus having the same information as he does. Since projection is all-encompassing, the projected version of the principal, who is real in a given agent’s mind, is believed to know not only this given agent’s strategy, but also share his beliefs about the strategies of all the other agents. Because an agent’s estimate of the strategies of the other agents is unbiased on average, the estimate attributed to the projected version of the principal will also be unbiased on average. In sum, the agent anticipates but underappreciates the extent to which an informed principal exaggerates the performance of the reference agents on average. Hence the agent underappreciates the informed principal’s exaggeration on average. The magnitude of this underappreciation is increasing in the agent’s own bias. If the agent is fully biased, $\rho_A = 1$, then no exaggeration is anticipated on average. By contrast, if the agent is unbiased, $\rho_A = 0$, the agent has correct beliefs about the principal’s expected estimate on average. The agent’s average underappreciation is thus increasing in $\rho_A$.

Finally, as long as the agent is not fully biased, the above claim on anticipation implies the following comparison of the agent’s propensity to choose the investment option over the sure payoff between the two treatments.

**Claim 3** (Agent’s investment choice). *The agent’s propensity to invest is lower when matched with an informed as opposed to an uninformed principal on average.*

This claim holds qualitatively as long as the principal is biased and the agent is not fully biased. The result follows from the fact that the value of the investment in the uninformed treatment first-order stochastically dominates
the value of the investment in the informed treatment. Because of the agent’s anticipation, it then follows that the agent has a greater incentive to invest in the relationship in the uninformed treatment than in the informed one.

**Homogenous Projection** We can now return to the case of homogenous projection, where the agent and the principal are equally biased. Let again $\rho_A = \rho_P = \rho$. Here the principal’s exaggeration in Claim 1 is governed by $\rho$, the perceived probability that an agent is a projected version. The agent’s anticipation is governed by $(1 - \rho)\rho$ – the agent’s perception of the probability that the principal assigning to the agent being the projected version. When estimating the model we will do so both under homogenous projection and under role-specific heterogeneous projection.

4 Results

We embark on the analysis by testing whether the principals’ expectations of the agents’ performance is affected by the information conditions (Claim 1). We then investigate whether the agents’ investment choices (Claim 3) as well as their second-order beliefs (Claim 2) reflect anticipation of different performance expectations by differentially informed principals. Finally, we explore to which extent the data can be accommodated by projection equilibrium (Madarász 2014a).

4.1 Principals

Figure 3 shows the distribution of individual performance estimates by informed and uninformed principals together with the actual performance of the reference agents. The figure clearly supports Claim 1. First, uninformed principals are, on average, very well calibrated: there is virtually no difference between their average estimate (39.76%) and the average performance of the reference agents (39.25%; $p = 0.824$). In contrast, informed principals overestimate the performance of the reference agents for the vast majority of tasks. Their average estimate of the success rate amounts to 57.45% which is significantly higher than both the actual performance of the

13We employed a $t$-test of the average estimate per principal against the average success rate (over all tasks). A Kolmogorov-Smirnov test of the distributions of average individual estimates between treatments yields $p = 0.001$. Unless stated otherwise, $p$-values throughout the result section refer to (two-sided) $t$-tests that are based on average statistics per subject.

14Figure 7 in the Appendix provides a plot of the principals’ average first-order belief per treatment over time.
reference agents \((p < 0.001)\) and the average estimate of uninformed principals \((p < 0.001)\).\(^{15}\) Informed principals overestimating the agents’ performance by more than 18 percentage points led to significantly lower expected earnings of informed principals (€2.40) than uninformed principals (€3.65; one-sided \(t\)-test: \(p = 0.034\)).\(^{16}\) We record our findings in support of Claim 1 in

**Result 1.** **Informed principals state significantly higher expectations regarding the performance of the reference agents than uninformed principals.**

\(^{15}\)Principals in the uninformed treatment who spotted the difference in the task also overestimated the success of the reference agents (60.93%). Principals who have the solution think the task was easier for the agents than it actually was, regardless of whether the solution was given right away as in the informed treatment or found by the principals themselves. The projection equilibrium framework predicts this finding. The magnitude of the uninformed principals’ overestimation cannot be directly compared to that of the informed principals, because of the self-selection issue. The uninformed principals who solved the task could be more skilled on average. Thus, we focus on the comparisons between the treatments rather than within treatments.

\(^{16}\)The average payoffs in the rounds randomly selected for payment were €1.50 and €2.50 in the informed and the uninformed treatment, respectively.
formed principals, but not uninformed principals, overestimate the performance of the reference agents on average.

Following Moore and Healy (2008), we can also examine the extent to which task difficulty per se plays a role in our analysis. If we divide our tasks into hard and easy tasks by using the median task difficulty as the cutoff (yielding 10 hard tasks with success rates of 0.42 and below and 10 easy tasks with success rates of 0.43 and above), we find that uninformed principals, on average, overestimate the actual success rate for hard tasks (p=0.047, sign test) but also underestimate the true success rate for easy tasks (p = 0.059). This reversal is well known as the hard-easy effect first established by Moore and Healy (2008). This reversal, however, is not observed for informed principals who significantly overestimate the true success rate for any task, irrespective of the degree of difficulty (p=0.012 and p=0.0069 for hard and easy tasks, respectively).

The systematically biased forecasts in Result 1 are not surprising given previous findings in the literature (Loewenstein et al., 2006) and it also confirms the basic premise of our design. We can now turn to the new aspect of our design which is the careful testing of the simultaneous anticipation of this phenomenon.

4.2 Agents

4.2.1 Investment decisions

We move on to the agents' investment choices. Remember, for the agents the only difference between the two treatments is that agents matched to informed principals were told that their principal had access to the solution while agents in the uninformed treatment were told that their principal had not seen the solution when forming his performance expectation. Figure 4 shows the distribution of individual investment rates in the informed and the uninformed treatment.

The figure reveals a clear pattern. Agents matched to informed principals invest at a considerably lower rate than agents matched to uninformed principals, but not uninformed principals, overestimate the performance of the reference agents on average.

If the solution to the task altered the informed principals' ex-ante perception of the task in other ways, the average estimate of the agents' performance across the two treatments should still be equal under the Bayesian framework even if the distributions are not exactly the same.
The treatment difference is not only statistically significant; the magnitude of the difference is also economically relevant. The average investment rate of agents matched to uninformed principals is 67.3%, whereas the average investment rate of agents matched to informed principals is only 39.2% ($p < 0.001$). The investment rate in each treatment is remarkably close to the fraction of cases where the investment pays off in each treatment, i.e., where the principals’ expectations do not exceed the agents’ success rate by more than 10 percentage points. Specifically, if agents knew the estimate of their principal (and best responded to this in-
Table 2: Regressions of individual investment rates on treatment, gender, and risk attitude.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(OLS)</th>
<th>Individual investment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1-informed)</td>
<td>−0.281***</td>
<td>−0.279***</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>−0.059</td>
</tr>
<tr>
<td>(1-female)</td>
<td></td>
<td>(0.075)</td>
</tr>
<tr>
<td>Treatment × Gender</td>
<td></td>
<td>−0.053</td>
</tr>
<tr>
<td>Coef. risk aversion</td>
<td></td>
<td>(0.151)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.673***</td>
<td>0.698***</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>N</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.134</td>
<td>0.140</td>
</tr>
</tbody>
</table>

Note: Values in parentheses represent standard errors. Asterisks represent p-values: *p < 0.1, **p < 0.05, ***p < 0.01.

formation), they would, on average, invest in 66.9% of the cases when being matched to an uninformed principal but only in 38.1% when being matched to an informed principal.

Table 2 reports the results of regressions of the investment rate per subject on a constant, a treatment dummy, a gender dummy, and individual risk attitudes. The estimation results corroborate the previous finding. First, the treatment effect remains significant and similar in size when we control for gender and individual risk attitude. Second, subjects with higher degrees of risk aversion tend to invest less often (columns 4 and 5) but this effect is not significant. Finally, the treatment effect is significant for both male and female subjects, with female subjects investing slightly (but not signifi-
significantly) less often than male subjects (columns 2, 3, and 5).\textsuperscript{20} We summarize our findings in

**Result 2.** *Agents invest significantly less often when they are matched to informed principals than when they are matched to uninformed principals.*

Result 2 is consistent with agents anticipating the projection of informed principals. The agents shy away from investing, that is, from choosing an option whose payoff depends negatively on the level of the principal’s performance expectations.

At this point other explanations for the treatment difference in investment choices might come to mind that do not posit that the agents anticipate the projection of informed principals. Specifically, models of stochastic choice with differential assumptions about the rate at which subjects make errors in their choices—while assuming that agents in both treatments entertain the same beliefs about the principals’ estimates on average—might explain the lower investment rate of agents in the informed treatment.\textsuperscript{21} As we will see in the following section, such alternative explanations are clearly ruled out by the analysis of the agents’ stated beliefs.

### 4.2.2 Stated beliefs

Collecting information about higher-order beliefs is essential to test equilibrium theories. We elicited agents’ first-order beliefs, i.e., their own estimates of the success rate, and their second-order beliefs, i.e., their estimate of their principal’s estimate of the success rate. According to Claim 2, the agents’ first-order beliefs are expected to be correct on average in both treatments, and the agents’ second-order beliefs are expected to reflect partial anticipation of the informed principals’ biased performance expectations. Figure 5 shows a bar chart of the average stated beliefs of the agents in each treatment together with the average performance of the reference agents and the corresponding estimates of the principals.

The figure summarizes our key findings. The left-hand bars on each panel (first-order beliefs) show that the agents are well calibrated in both

\textsuperscript{20}There is no significant interaction between the treatment and having completed a task successfully. The treatment effect on investment decisions is significant both for periods in which the agents solved the task and for periods in which the agents did not solve the task. For further details see Table 4 in the Appendix.

\textsuperscript{21}In principle, this idea can be captured by quantal response equilibrium (QRE) with different precision parameters between treatments—either on the side of the agents or the principals. Note, however, that QRE is not suited to explain the biased estimates of informed principals in the first place (or any anticipation thereof).
treatments: their estimates of the success rate are correct on average and not significantly different between treatments ($p = 0.956$). The middle bars reveals that the agents’ second-order beliefs are somewhat higher than their first-order beliefs in either treatment. That is, in both information conditions the agents tend to expect that the principals are somewhat more optimistic regarding the reference agents’ performance than themselves. However, as predicted by Claim 2, this effect is significantly stronger in the informed treatment than in the uninformed treatment: First, the second-order beliefs of agents matched to informed principals (51.14%) are significantly.

---

22 The average success rate in the first ten tasks (where we elicited the agents’ beliefs) is 40.0%. The agents’ average estimate of the success rates in the first ten tasks is 39.72% ($p = 0.917$) in the uninformed treatment and 39.91% ($p = 0.967$) in the uninformed treatment.

23 In the uninformed treatment, none of the elicited beliefs are, on average, significantly different from the true success rate: neither the agents’ first-order beliefs ($p = 0.917$), their second-order beliefs ($p = 0.140$), nor the principals’ estimates ($p = 0.337$).

---
higher than those of agents matched to uninformed principals (44.15%; one-sided \( t \)-test: \( p = 0.031 \)). Second, the individual differences between second- and first-order beliefs are significantly larger for agents matched to informed principals than for agents matched to uninformed principals (\( p < 0.001 \)).

A further key finding can be taken from Figure 5. Although the second-order beliefs of agents in the informed treatment are significantly higher than the actual success rate, they are, on average, also significantly lower than the informed principals’ estimates (one-sided \( t \)-test: \( p = 0.047 \)). That is, the agents appear to anticipate the projection of the informed principals, but not to its full extent. This finding is consistent with the predictions of projection equilibrium.

**Result 3** (Partial anticipation of information projection). The agent’s estimate of the success rate (first-order belief) is correct on average in both treatments. The average difference between the agent’s estimate of the principal’s estimate (second-order belief) and her own estimate (first-order belief) is significantly larger in the informed treatment than in the uninformed treatment. In the informed treatment, the agent’s estimate of the principal’s estimate of the success rate (second-order belief) is, on average, between her own estimate and the principal’s estimate.

We have presented evidence that is qualitatively consistent with the predictions of Section 3. In particular, informed principals, but not uninformed ones, exaggerated the success rate of agents (Claim 1). Agents matched with informed principals, but not with uninformed ones, partially anticipated that their principals overestimated the reference agents’ success rates on the tasks (Claim 2). Furthermore, agents were less likely to invest when matched with an informed principal than when matched with an uninformed principal (Claim 3). In the next section we conduct a quantitative test of the descriptive power of projection equilibrium.

---

24 This treatment difference is robust when controlling for individual characteristics (see Table 5 in the Appendix) and significant both for instances where the agents solved the task and where they did not solve the task (see Table 6 in the Appendix).

25 We find no significant interaction between the treatment and having completed a task successfully. The treatment effect on the individual differences between first- and second-order estimates is significant both for periods in which the agents solved the task and for periods in which the agents did not solve the task. For further details, see Table 6 in the Appendix.
4.3 Estimation of projection equilibrium parameters

How well can the data be accommodated by a single-parameter extension of Bayesian Nash Equilibrium? In the most parsimonious parametrization of projection equilibrium, both the principal’s exaggeration and the agent’s underestimation of the principal’s projection are governed by the same parameter $\rho$. In the following we identify the levels of projection on both sides that would be consistent with the beliefs we elicited. We then test whether the estimated projection is significantly different across player roles.

Claims 1 and 2 provide the equations to estimate the projection parameter of the principals and the agents in the informed treatment. We start with a flexible specification that allows for different degrees of projection bias between roles as well as within roles, where we denote the average degree of projection in the principal population by $\rho_P$ and that in the agent population by $\rho_A$. Following Claim 1, an informed principal $i$—who solves the task with probability $1$—has the following expected belief $b_{P1,i,t}$ of the success rate $\pi_t$ in task $t$

$$E(b_{P1,i,t} \mid \pi_t, \rho_P) = \pi_t + \rho_P (1 - \pi_t) =: \mu_{it}. \quad (4)$$

Following Claim 2, agent $j$’s expected second-order belief $b_{A2,j,t}$ about her informed principal’s mean estimate, conditional on her (unbiased) first-order belief, is given by

$$E(b_{A2,j,t} \mid b_{A1,j,t} = \pi_t, \rho_A) = b_{A1,j,t} + (1 - \rho_A) \rho_P (1 - b_{A1,j,t}) =: \mu_{j,t}. \quad (5)$$

To account for the fact that subjects’ stated beliefs fall in a bounded interval between zero and one, we employ a beta regression model. Specifically, we assume that subjects’ responses $b_{P1,i,t}$ and $b_{A2,i,t}$ follow a beta distribution with task- and individual-specific mean (4) and (5), respectively. With slight abuse of notation we write $b_{kit}$ for subject $i$’s stated belief in role $k$, where $b_{kit} = b_{P1,i,t}$ for principals ($k = P$) and $b_{kit} = b_{A2,i,t}$ for agents ($k = A$). With beta-distributed responses $b_{kit}$, the probability of observing belief $b_{kit}$ of subject $i$ in role $k$ can be written as:

$$f(b_{kit} \mid \mu_{kit}, \phi_b) = \frac{\Gamma(\phi_b)}{\Gamma(\phi_b \mu_{kit}) \Gamma(\phi_b(1 - \mu_{kit}))} (b_{kit})^{\phi_b \mu_{kit} - 1} (1 - b_{kit})^{\phi_b (1 - \mu_{kit}) - 1},$$

Following the derivation of Claim 1 and Claim 2, we calculate $d_t$—the difference between the ex-ante probability of solving task $t$ for an informed principal and for an agent—by subtracting the empirical success rate of the reference agents for each task from 1—approximately the success rate of informed principals. 

26 Following the derivation of Claim 1 and Claim 2, we calculate $d_t$—the difference between the ex-ante probability of solving task $t$ for an informed principal and for an agent—by subtracting the empirical success rate of the reference agents for each task from 1—approximately the success rate of informed principals.

27 See Ferrari and Cribari-Neto (2004). A link function as in standard beta regression models is not needed because (4) and (5) map $(0,1) \rightarrow (0,1)$ for $\rho \in [0,1)$ and $\pi_t, b_{kit} \in (0,1)$. Abstaining from using a link function has the advantage that we can interpret
where $\phi_b$ is a precision parameter that is negatively related to the noise in subjects’ response.

To capture unobserved individual heterogeneity in the projection parameters and to account for repeated observations on the individual level, we employ a random coefficients model. Specifically, we assume that the individual-specific degree of projection $\rho_{ki} \in [0, 1], k \in \{A, P\}$ follows a beta distribution with role-specific mean $\mu_{ki}$ and density

$$g(\rho_{ki}; \mu_{ki}, \phi) = \frac{\Gamma(\phi_{\rho})}{\Gamma(\phi_{\mu}\rho)\Gamma(\phi_{\mu}(1-\rho_{ki}))}(\rho_{ki})^{\phi_{\rho}(\mu_{ki})-1}(1-\rho_{ki})^{\phi_{\rho}(1-\mu_{ki})-1},$$

where the parameter $\phi_{\rho}$ is negatively related to the variance of projection bias in the principal and the agent population.

We now formulate the log-likelihood function. Conditional on $\rho_{ki}$ and $\phi_{\rho}$, the likelihood of observing the sequence of stated beliefs $(b_{ki}^k)_t$ of subject $i$ in role $k$ is given by

$$L^k_{ki}(\rho_{ki}, \phi_{\rho}) = \prod_t f(b_{ki}^k; \mu_{ki}(\rho_{ki}), \phi_{\rho}).$$

Hence, the unconditional probability amounts to

$$L^k_{ki}(\mu_{ki}, \phi_{\rho}) = \int \left[ \prod_t f(b_{ki}^k; \mu_{ki}(\rho_{ki}), \phi_{\rho}) \right] g(\rho_{ki}; \mu_{ki}, \phi_{\rho}) d\rho_{ki}. \tag{6}$$

The joint log likelihood function of the principals’ and the agents’ responses can then be written as

$$\ln L(\mu_P, \mu_A, \phi_{\rho}, \phi_{b}) = \sum_k \sum_i \log L^k_{ki}(\rho_{ki}, \phi_{\rho}, \phi_{b}). \tag{7}$$

We estimate the the parameters in (7) simultaneously by maximum simulated likelihood (Train 2009; Wooldridge, 2010).²⁸

Table 3 shows the estimation results for the unrestricted model as well as for the restricted model with $\mu_P = \mu_A$. We make three observations.²⁹

²⁸The estimation is conducted with GAUSS. We use Halton sequences of length $R = 100,000$ for each individual with different primes as the basis for the sequences for the principals and the agents.

²⁹The results are robust with respect to alternative starting values for the estimation procedure. All regressions for a uniform grid of starting values converge to the same estimates (both for the restricted and the unrestricted model). Thus, the likelihood function in (7) appears to assume a global (and unique) maximum at the estimated parameters.
Table 3: Maximum likelihood estimates of projection bias $\rho$ based on Claim 1 and 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Heterogeneous $\rho$</th>
<th>Homogeneous $\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Conf. interval</td>
</tr>
<tr>
<td>$\rho_P^P$</td>
<td>0.279***</td>
<td>[0.198, 0.359]</td>
</tr>
<tr>
<td>$\rho_A^A$</td>
<td>0.299**</td>
<td>[0.036, 0.562]</td>
</tr>
<tr>
<td>$\rho_P = \rho_A$</td>
<td>0.274***</td>
<td>[0.211, 0.338]</td>
</tr>
<tr>
<td>$\phi_P$</td>
<td>4.013***</td>
<td>[1.651, 6.375]</td>
</tr>
<tr>
<td>$\phi_b$</td>
<td>5.156***</td>
<td>[4.637, 5.675]</td>
</tr>
</tbody>
</table>

Note: Values in square brackets represent 95% confidence intervals. Asterisks represent $p$-values: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ Testing $H_0: \rho_P^P = \rho_A^A$ in column (1) yields $p = 0.8389$.

First, the principals’ average degree of projection is estimated to be 0.28 with a confidence interval of [0.20, 0.36]. This estimate clearly indicates the relevance of projection: The Bayesian Nash Equilibrium—which is the special case in which $\rho$ is zero—is clearly rejected. Second, the agents’ average degree of projection, the extent to which the agent under-appreciates that the principal has beliefs that are potentially different and also biased, is estimated to be 0.30 with a confidence interval of [0.04, 0.56]. Recall that if $\rho_A = 0$, the agent does not display any projection and fully anticipates the principal’s projection. Contrary, if $\rho_A = 1$, the agent exhibits full projection and does not anticipate any projection from the principal. The 0.30 estimate—which is significantly different from 0 and clearly different from 1—gives structure to our observation that agents do anticipate the projection of informed principals—but tend to under-anticipate the principals’ level of projection in their second-order estimates. Finally, and strikingly, the esti-

---

$^{30}$We also did the estimation separately for tasks below median difficulty and tasks above median difficulty. The parameters in these two estimations were similar to each other and to the pooled estimation. Furthermore, Vuong’s (1989) test of the model in column (1) against the standard model without projection yields $p < 0.0001$. For the standard model specification, we assume $b_{it}^k$ is beta distributed with mean $\mu_{it}^k = \pi_t$ for principals and $\mu_{it}^A = b_{it}^A_{1,j}$ for agents, i.e., $b_{it}^k \sim Beta(\mu_{it}^k, \phi_k)$. The estimated precision parameter for the standard model amounts to $\hat{\phi}_k = 2.580$ (s.e. 0.115); the log likelihood of the standard model is $-115.801$. 

26
estimated parameters of projection are not significantly different between the principals and the agents \((p = 0.839)\). Given the similarity between the estimated parameters in the unrestricted model, the log likelihood of the two models are very close, and standard model selection criteria clearly favor the single-parameter model of homogeneous projection over the model with two parameters \((\text{BIC})\).

### 4.4 Individual heterogeneity

The final part of the analysis is devoted to test of the predictions of projection equilibrium on the individual level in conjunction with a test of the econometric specification of model \((7)\). Specifically, we test whether the mean projection bias \(\hat{\rho}_\mu = 0.274\) estimated from \((7)\) is indeed generated by a beta distribution of \(\rho_i\) or, alternatively, whether individual heterogeneity in projection bias is better described by some other distribution. A misspecification in this matter would not only be relevant from an econometric point of view; it might also challenge the economic interpretation of our results. If, for example, the estimate of the mean projection bias was a result of a finite mixture of some agents not anticipating information projection at all \((\rho = 1)\) and the remaining agents fully anticipating information projection \((\rho = 0)\), then model \((7)\) would be misspecified and, more importantly, Claim 2 \((\text{partial} \text{ anticipation of information projection})\) would have little empirical bite on the individual level.

We base our specification test on non-parametric density estimates of individual projection bias in the principal and the agent population. To this end, we first obtain individual estimates of the projection bias parameter \(\rho\) for each principal and agent in the informed treatment, where we employ the most simple and straightforward estimation technique without imposing any restrictions on the parameter \(\rho\). Specifically, for each informed principal \(i\), we estimate his projection bias \(\hat{\rho}_i^P\) from

\[
b_{1,i,t}^P = \pi_t + \rho_i^P(1 - \pi_t) + \epsilon_{it},
\]

where \(b_{1,i,t}^P\) denotes the principal’s expectation of the success rate \(\pi_t\) in task \(t\), and \(\epsilon_{it}\) denotes an independent and normally distributed error term with mean zero and variance \(\sigma_i^2\). Analogously, for each agent \(j\) in the informed treatment we estimate her projection bias \(\hat{\rho}_j^A\) from

\[
b_{2,j,t}^A = b_{1,j,t}^A + (1 - \rho_j^A)\rho_\mu^P(1 - b_{1,j,t}^A) + \epsilon_{jt},
\]

where \(b_{1,j,t}^A\) denotes the agent’s estimate of the success rate in task \(t\) (first-order belief), \(b_{2,j,t}^A\) denotes her estimate of the principal’s estimate (second-
order belief), $\rho^P_\mu$ denotes the mean projection bias in the principal population, and $\epsilon_{jt}$ denotes an independent and normally distributed error with mean zero and variance $\sigma_j^2$. We estimate the parameters in (8) and (9) by linear least squares, where we substitute $\rho^P_\mu$ in (9) with the average estimate of $\rho^P_i$ obtained from the regressions in (8).  

![Figure 6: Cumulative distribution functions (CDF) of principals’ (solid) and agents’ (dashed) projection bias $\rho$ in the informed treatment. Black lines represent empirical CDFs; gray lines represent best-fitting beta CDFs.](image)

Figure 6 shows the empirical cumulative distribution functions (CDFs) of individual projection bias in the principal and the agent population. A casual inspection of the figure shows that the empirical CDFs of the principals’ and the agents’ $\rho$ are quite similar. In fact, a Kolmogorov-Smirnov test does not reveal any significant difference between the distributions ($p = 0.441$). That is, not only the average projection bias by the principals and the agents

---

41Because we use an estimate of $\rho^P_i$ in the estimation of the agents’ projection bias, the composite error term in equation (9) is heteroscedastic. We therefore base all inference on the individual level on heteroscedasticity-robust standard errors. Unlike in the simultaneous estimation of the agents’ and the principals’ projection bias from (7), the simple estimation approach applied here assures that the individual estimates of the principals’ projection bias are not informed by the data of the agents’ choices, a feature that is very desirable for our specification test below.
is the same; also the distributions of the principals’ and the agents’ projection bias are the same. These findings lend strong support for a key feature of projection equilibrium—that information projection and the anticipation thereof are governed by the same parameter.

Figure 6 conveys a second important message regarding the econometric specification in (7). The empirical CDFs of principals’ and the agents’ projection bias are very well described by beta distributions (the gray lines in the figure show the best-fitting beta distributions). Moreover, the joint empirical CDF of the principals’ and the agents’ projection bias is not significantly different from the beta distribution \( g(\rho_P = 0.274, \phi_P = 3.965) \) obtained from model (7) with homogeneous projection (see Table 3; Kolmogorov-Smirnov test: \( p = 0.063 \)). That is, our non-parametric estimate of the distribution of individual projection bias clearly supports the beta specification with respect to individual heterogeneity in (7).

A final crucial observation that can be taken from Figure 6 is that—in line with Claim 2—most agents appear to anticipate the information projection by the principals—but not too its full extent \( (\rho_j^A \in (0, 1) \text{ for most agents}) \). This observation is backed up by tests of the estimated projection parameters against zero and one. For both the principals and the agents, the most frequently observed category is the one where \( \rho \) is estimated to be significantly larger than zero but also significantly smaller than 1. That is, the most prevalent behavioral pattern is significant—but not full—information projection and at the same time significant—but not full—anticipation of information projection.

\[ 32 \text{Kolmogorov-Smirnov tests of the empirical CDF against the best-fitting beta distribution (as shown in Figure 6) yields } p = 0.941 \text{ for the principals and } p = 0.584 \text{ for the agents.} \]

\[ 33 \text{70.8\% of the principals and 50\% of the agents fall in this category. The second most common category in both populations is the one of } \rho \text{ being not significantly different from 0 but significantly different from 1, i.e., no information projection and at the same time full anticipation of others’ information projection (25\% of the principals and 37.5\% of the agents). Finally, a few agents (12.5\%) exhibit an estimated } \rho \text{ that is not significantly different from 1 but significantly different from 0, i.e., they do not anticipate the principals’ information projection at all. Although the estimated mass on the same category in the principal population (full projection) is zero, the difference is based on three observations only. In fact, in line with the previous findings there is no significant difference between the agents’ and the principals’ categorized distribution of projection bias (Fisher’s exact test: } p = 0.124 \text{).} \]

\[ 34 \text{We also performed a model-free estimation of each agent’s degree of anticipation } \alpha_j \text{ by estimating the weight in } b_{2,t}^A = (1 - \alpha_j)b_{1,t}^A + \alpha_j b_{1,t}^P + \epsilon_{jt} \text{, where } b_{2,t}^P = \sum_j b_{1,t}^P \text{ is the principals’ average estimate of the success rate in task } t. \text{ The categorized estimated distribution of } \alpha_j \text{ is qualitatively similar to the corresponding distribution of } (1 - \rho_j^A). \]
5 Conclusion

A host of robust findings demonstrate that people engage in limited informational perspective-taking. This study is the first to document people’s anticipation of such mistakes in the thinking of others. We find not only that better-informed principals project their information onto agents, but also that lesser-informed agents anticipate the principals’ bias as evinced by their decision to insure against the principals’ overestimation of success. The second-order estimates show that an agent in our experiment understands that the principal has a belief about the agent’s perspective which is systematically different from the agent’s true perspective.

The findings support the logic of projection equilibrium which integrates this basic phenomenon with such higher-order perceptions. While informational projection biases one’s belief about the perspectives of others, our anticipation result shows that a person is at least partially sophisticated about the fact that differentially informed others have biased beliefs about her perspective on average. Such sophistication is partial due to a person’s own projection: agents underappreciate the degree to which principals exaggerate their success rate due to the agents’ own mistake. In support of the logic of this solution concept, we find that the extent to which informed principals overestimate the agents’ success rate is very similar in magnitude to the extent to which lesser-informed agents under-appreciate such over-estimation.

5.1 Relation to Alternative Models and Mechanisms

Unlike a number of other behavioral game theory models of private information games, projection equilibrium focuses on the players misperceiving other players’ information per se, rather than misperceiving the relationship between other players’ information and their actions. In particular, the models of Jehiel and Koessler (2008) and Eyster and Rabin (2005) assume that people have coarse or misspecified expectations about the link

First, the majority of agents (54.2\%) exhibit an estimated $\alpha_j$ that is significantly larger than zero but also significantly smaller than 1. That is, their estimate $b_{1,t}^A$ of the principal’s estimate of the success rate is, on average, between their own estimate $b_{1,t}^A$ and the average estimate of the principals $b_{1,t}^P$. In other words, most agents partially anticipate the projection of informed principals. Further 25.0\% of the agents fall into the category of $\alpha_j = 0$ with no anticipation of information projection ($b_{2,t}^A = b_{1,t}^A$), and 16.7\% of the agents fall into the category of $\alpha_j = 1$ exhibiting full anticipation ($b_{2,t}^A = b_{1,t}^P$). The remaining 4.2\% correspond to an agent whose estimated $\alpha_j$ is not significantly different from both zero and one.
between others’ actions and their information, but these models are closed by the assumption that those expectations are correct on average. In turn, in the context of the current experiment, these models imply that a principal should never exaggerate the agent’s performance on average and the agent should never anticipate any mistake by the principal on average. Hence, just as the BNE, they predict a null affect for both hypothesis tested in this paper.\footnote{Note also that QRE predicts the null of no treatment difference since the principal’s incentives in the two treatments are exactly the same.}

Although we find no evidence that risk aversion matters for the subjects choices, note also that since more information helps unbiased principal’s make more accurate forecasts on average, under correct beliefs and risk aversion the agent should be choosing the lottery over the safe option more often when the principal is informed rather than when she is not. Instead we find the opposite.

Finally, note that overconfidence cannot explain the subjects choices either. If an agent believes that she is better than average, then she will simply exaggerate the return from choosing the investment option as opposed to the sure payment, but this will not differ across treatments. Similarly a principal may be over- and under-confident when inferring about other’s performance on a given task, but there is no reason for this to systematically interact with the treatment per se.

\section*{5.2 Future Directions}

Our simple experimental setup can be used to study a number of interesting questions related to the phenomenon of information projection. For example, to what extent might principals anticipate their biased estimation of the agents’ performance and opt out of receiving the solutions when given the choice? Additionally, in our study, the principals do not receive any feedback about the agents’ choices and the agents do not receive any feedback about the principals’ estimations. Future experimental work could study how different levels of feedback might affect how the principals adjust their estimations and how the agents change the extent of their anticipation.

Naiveté about one’s own tendency combined with partial sophistication about others’ tendency to project information has potentially important economic implications in agency and strategic settings more generally. Consistent with the discussion in Berlin (2003), our results imply that people not simply exaggerate the extent to which others have the same information as
they do, but anticipate that others’ belief about their perspective systematically differs from the perspective they truly have. Our investment decision can be interpreted as the agents’ valuation of the relationship with the principal. The greater is the information gap between the principal’s information when monitoring and the agent’s true information, the less the agent will value the relationship. If agents have to sort into various jobs in the labor market, such anticipation will affect the decision and may induce people to work without principals when principals monitor with ex-post information; here monitoring increases rather than decreases agency costs. Our results may help explain a firm’s desire to limit the use of subjective performance measures based on ex-post information in combination with objective performance measures that would otherwise be optimal in a Bayesian setting (Baker et al., 1994) and provides a cognitive reason why such evaluations might often be biased. This affects the Bayesian results on the optimal use of information and incentives (Holmström, 1979). In our setting, greater scrutiny directly biases learning leading to wrong judgments and systematically lower payoffs. Importantly, the type of reasons identified in the literature under rational expectations whereby more information backfires in agency settings with dynamic career concerns (e.g., Scharfstein and Stein, 1990; Cremer, 1995; Prat, 2005) will typically require different contractual solutions than those used for the anticipation of projection.

Our design could be extended to study other information and incentive schemes and further contribute empirically to the growing literature on incorporating behavioral assumptions into the understanding of decision-making in firms (e.g., Camerer and Malmendier, 2007; Malmendier and Tate, 2008; Spiegler, 2011). In context of negligence liability or patent law, where projection will systematically distort judgements made in hindsight (Harley, 2007), anticipating it may further distort the ex-ante incentives for effort or investment into innovation. Importantly, the fact that anticipation is partial means that the extent to which an ex-ante designer would want to correct for the presence of this mistake, for example in the design of legal institutions, might be insufficient.

The paper also lends support to the notion of projection equilibrium for the analysis of games with informational differences. Future research can investigate its predictions for a great variety of settings both theoretically and empirically. Here key implications may depend on the extent to which people anticipate the projection of others. For example, Madarász (2014b) applies the idea of informational projection to dynamic bargaining and shows that the anticipation of even a small amount of this mistake can have a significant impact on bargaining outcomes and the optimal selling mechanism.
(Myerson, 1981). It would be interesting to know the extent to which this mistake and its anticipation may play a significant role in communication or coordination problems.

Finally, our results speak to the importance of understanding the consequences of established forms of biased beliefs (DellaVigna, 2009; Möbius et al., 2014) being anticipated by others. Indeed, while there are multiple factors that might generate biased beliefs, it is the patterns in higher-order beliefs, beliefs about those biases, that give us more insight into the leading factors. For example, if informed players exaggerated the performance of uninformed players because everyone thinks others are just like them, then it is harder to account for the partial anticipation of this exaggeration by the agents as predicted by projection equilibrium and established by our findings. In other contexts with biased beliefs about one’s own choices, such as naiveté about self-control problems (O’Donoghue and Rabin, 1999), or a person’s failure to appreciate the full extent to which she will be subject to the endowment effect (Kahneman et al., 1990; Loewenstein and Adler, 1995) as her reference point shifts, people might be better at anticipating the mistakes of others (Van Boven et al., 2003). This would then affect implications for trade, delegation, and optimal contracting more generally.
References


Appendix

5.3 Proofs

Before turning to the proofs, a few remarks are in order. First, the incentives for solving the basic task can be assumed to be independent of the incentives for the prediction and investment tasks. This is true because the payoff from the prediction and the investment tasks are based on the success rate of the population of reference agents and a randomly chosen principal’s estimate thereof. Second, the game is effectively an $N$-player game where two different agents need not infer the same information from the video. For any two agents, $j \neq k$, $P_{A_j}$ need not equal $P_{A_k}$. Hence when deriving the implications, we need to account for the fact that an agent also project onto the reference agents. This extension to the $N$-player game follows the same logic as the two-player definition presented. Here each player $i$ assigns probability $\rho_i$ to all of her opponents being a projected version who has the same information player $i$ has. As before, player $i$ assigns probability $(1-\rho_i)$ to the true distribution of his opponents’ strategies. Furthermore, player $i$ believes that the projected version of any of his opponent $k$, has the same belief about the strategies of the other players as player $i$ does. Specifically, a $\rho_A$ biased agent believes that with probability $\rho_A$ all other players have the same information he does. Furthermore, such projected versions have the same beliefs about the strategy of others as he does. This implies that a given biased agent believes that the projected version of the principal has the same belief about the strategy of the other agents as he does. Claim 1 uses the fact that a real principal projects on all agents. Claim 2 below uses the fact in addition that a real agent believes that the projected version of the principal has the same belief about the other agents as he does.

Let us then denote the success probability of agent $j$ given information $P_{A_j}$ by
\[
\max_{x \in D} E[f(\omega, x) | P_{A_j}(\omega)].
\]
As mentioned before, the ex-ante expected value of this problem is the same for any randomly chosen two players having access to the video only. Formally, $E_\omega[\max_{x \in D} E[f(\omega, x) | P_{A_j}(\omega)]] = E_\omega[\max_{x \in D} E[f(\omega, x) | P_{A_k}(\omega)]]$.

Finally, because the agents solving the task have no direct strategic interaction with each other, we can thus introduce a representative agent and denote it by $\overline{A}$. This is short-hand for the expected average performance of the population of agents. With a slight abuse of notation, we can represent the success rate of the agent population in a realized state $\omega$ by $\max_{x \in D} E[f(\omega, x) | P_{\overline{A}}(\omega)]$. This is the average of the individual success
probabilities given the individually realized informations of the reference agents. We will then equate the ex-ante expectation of this with the population success rate: \( E_\omega [\max_{x \in D} E[f(\omega, x) \mid P_\pi(\omega)]] = \pi \).

**Proof of Claim 1.** Let \( E^{\rho_P} \) denote the expectation of a \( \rho_P \)-biased principal. The ex-ante expected mean estimate of \( \pi \) by a principal is

\[
E_\omega \left[ E^{\rho_P} \max_{x \in D} E[f(\omega, d) \mid P_\pi(\omega)] \mid P_P(\omega) \right].
\]

(10)

Using the definition of projection equilibrium for a principal with information \( P_P(\omega) \), we obtain that

\[
E_\omega \left[ \rho_P \max_{x \in D} E[f(\omega, x) \mid P_P(\omega)] + (1 - \rho_P)E[\max_{x \in D} E[f(\omega, x) \mid P_\pi(\omega)] \mid P_P(\omega)] \right].
\]

Given the law of iterated expectations and the linearity of expectations this then becomes

\[
\rho_P E_\omega \left[ \max_{x \in D} E[f(\omega, x) \mid P_P(\omega)] \right] + (1 - \rho_P)E_\omega \left[ \max_{x \in D} E[f(\omega, x) \mid P_\pi(\omega)] \right],
\]

(11)

which equals \( \rho_P (d + \pi) + (1 - \rho_P)\pi = \pi + \rho_P d \).

**Proof of Claim 2.** Let \( E^{\rho_A} \) denote the expectations of a \( \rho_A \)-biased agent. The ex-ante expected mean estimate of \( \pi \) by agent \( A \) is

\[
E_\omega \left[ E^{\rho_A} \left[ \max_{x \in D} E[f(\omega, x) \mid P_\pi(\omega)] \mid P_A(\omega) \right] \right],
\]

(12)

where \( P_A(\omega) \) is the information of the particular agent \( A \) in a given state. Using the definition, this equals

\[
E_\omega \left[ \rho_A \max_{x \in D} E[f(\omega, x) \mid P_A(\omega)] + (1 - \rho_A)E[\max_{x \in D} E[f(\omega, x) \mid P_\pi(\omega)] \mid P_A(\omega)] \right].
\]

Given the law of iterated expectations as before, this then becomes \( \rho_A \pi + (1 - \rho_A)\pi = \pi \).

Consider now the expected estimate of the agent of the mean estimate of principal of \( \pi \). This corresponds to

\[
E_\omega \left[ E^{\rho_A} \left[ E^{\rho_P} \left[ \max_{x \in D} E[f(\omega, x) \mid P_\pi(\omega)] \mid P_P(\omega) \right] \mid P_A(\omega) \right] \right]
\]

(13)
Using the definition and the linearity of expectations, we can re-write this as

\[
E_\omega \left[ \rho_A E^{\rho_A} \left[ \max_{x \in D} E[f(\omega, x) \mid P_A(\omega)] \right] \mid P_A(\omega) \right] + (1 - \rho_A) E^{\rho_A} \left[ E\left[ \max_{x \in D} E[f(\omega, x) \mid P_A(\omega)] \mid P_P(\omega) \right] \mid P_A(\omega) \right].
\]

(14)

The first part of the above expression is based on the fact that agent A projects onto all other agents as well as the principal. Furthermore, he believes that the projected version of the principal has the same beliefs about the strategies of the other agents as he does. Using the linearity of expectations and the above result, we can re-arrange the above expression to obtain

\[
\rho_A \pi + (1 - \rho_A) E_\omega \left[ E^{\rho_P} \left[ \max_{x \in D} E[f(\omega, x) \mid P_A(\omega)] \mid P_P(\omega) \right] \mid P_A(\omega) \right].
\]

(15)

Given Claim 1, the second term equals \((1 - \rho_A)((\rho_P(d + \pi) + (1 - \rho_P)\pi))\). Hence the above expression equals \(\pi + (1 - \rho_A)\rho_Pd\). \(\square\)
5.4 Supplementary analysis

5.4.1 Stated beliefs of the principals

![Graph showing average performance estimates of principals and actual success rate of the reference agents over time.](image)

Figure 7: Average performance estimates of principals and actual success rate of the reference agents over time.
5.4.2 Investment decisions of the agents

![Graph showing investment rates per session over time with two lines representing agents matched to informed principals (solid line) and agents matched to uninformed principals (dashed line). The x-axis represents the period, and the y-axis represents the investment rate.]

Figure 8: Investment rates per session over time.
Table 4: Propensity to invest conditional on treatment and successful task completion.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Investment decision (1-investment, 0-no investment)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Treatment (1-informed)</td>
<td>−0.727***</td>
</tr>
<tr>
<td>(0.205)</td>
<td>(0.211)</td>
</tr>
<tr>
<td>Success (1-task solved)</td>
<td>0.429***</td>
</tr>
<tr>
<td>(0.096)</td>
<td>(0.126)</td>
</tr>
<tr>
<td>Treatment×Success</td>
<td></td>
</tr>
<tr>
<td>(0.190)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.467***</td>
</tr>
<tr>
<td>(0.146)</td>
<td>(0.149)</td>
</tr>
</tbody>
</table>

| N | 1410 | 1410 | 1410 |
| R² | −916.436 | −897.552 | −897.512 |
| F | 12.575 | 29.023 | 29.581 |

Note: Values in parentheses represent standard errors corrected for clusters on the individual level. Asterisks represent p-values: *p < 0.1, **p < 0.05, ***p < 0.01.
5.4.3 Stated beliefs of the agents

Figure 9: Agents’ first-order beliefs (estimates of the success rates of the reference agents) and second-order beliefs (estimates of the principals’ estimate) over time, conditional on being matched with informed or uninformed principals.
Average difference between second-order and first-order belief per agent

Figure 10: Empirical cumulative distribution functions of each agent’s average difference between her second-order belief (estimate of the principal’s estimate of the success rate) and her first-order belief (own estimate of the success rates), conditional on being matched with informed or uninformed principals.

Table 5 reports the results of regressions of average individual differences between the agents’ second- and first-order beliefs on a constant, a treatment dummy, a gender dummy, and individual risk attitudes as measured by DOSE. The treatment effect remains significant and becomes—if anything—somewhat larger when we add these controls. There is virtually no effect of risk attitude on the difference between agents’ first- and second-order beliefs (columns 4 and 5), and female agents in the uninformed treatment show a somewhat (but not significantly) larger gap between their own estimates of the success rate and their belief about the corresponding estimates of the principals (columns 3 and 5).
Table 5: Mean individual differences in second-order beliefs (estimate of the principal’s estimate) and first-order beliefs $b_{1,i}$ (estimate of success rate) by treatment and further controls.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>$\left( b_{2,i} - b_{1,i} \right) = T^{-1} \sum_t (b_{2,i,t} - b_{1,i,t})$</th>
<th>(OLS)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1-informed)</td>
<td>0.068***</td>
<td>0.019</td>
<td>0.019</td>
<td>0.024</td>
<td>0.020</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.013</td>
<td>0.020</td>
<td>0.029</td>
<td>0.030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1-female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment × Gender</td>
<td>-0.062</td>
<td>-0.062</td>
<td>-0.047</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coef. risk aversion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(DOSE)</td>
<td>-0.006</td>
<td>-0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.044***</td>
<td>0.014</td>
<td>0.015</td>
<td>0.016</td>
<td>0.014</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.040**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.030*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.048***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.034*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.220</td>
<td>0.228</td>
<td>0.270</td>
<td>0.236</td>
<td>0.278</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Values in parentheses represent standard errors. Asterisk represent $p$-values: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. 

Table 6: Individual differences between second-order beliefs and first-order beliefs conditional on treatment and successful task completion.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(OLS)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>0.068***</td>
<td>0.068***</td>
<td>0.067***</td>
<td></td>
</tr>
<tr>
<td>(1-informed)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>Success</td>
<td>-0.039***</td>
<td>-0.041**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1-task solved)</td>
<td>(0.011)</td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment×Success</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.044***</td>
<td>0.058***</td>
<td>0.058***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.017)</td>
<td>(0.019)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>470</th>
<th>470</th>
<th>470</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.090</td>
<td>0.117</td>
<td>0.117</td>
</tr>
<tr>
<td>$F$</td>
<td>12.828</td>
<td>17.767</td>
<td>13.296</td>
</tr>
</tbody>
</table>

Note: Values in parentheses represent standard errors corrected for clusters on the individual level. Asterisks represent $p$-values: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. 

48
5.5 Instructions

5.5.1 Instructions for principals (translated from German)

Welcome to the experiment!

The experiment that you will be participating in is part of a project funded by the German Research Foundation (DFG). It aims to analyze economic decision-making.

You are not allowed to use any electronic devices or to communicate with other participants during the experiment. You may only use programs and features that are allocated to the experiment. **Do not talk to any other participant.** Please raise your hand if you have any questions. We will then approach you and answer your question in private. Please do not under any circumstances raise your voice while asking a question. Should the question be relevant for everyone, we will repeat it aloud and answer it. If you break these rules, we may have to exclude you from the experiment and from receiving payment.

You will receive a show-up fee of 5 Euros for your attendance. You can earn additional money through the experiment. The level of your earnings depends on your decisions, the decisions of participants in former experiments, and on chance. The instructions are the same for everyone. A detailed plan of procedures and the conditions of payment will be explained below.

Tasks

You will face 20 tasks in the course of the experiment. For each task, you will be shown a short video. Each video consists of two images that are shown alternately. Your task is to spot the difference between the images.

The duration of each video is 14 seconds. After each video, you will have 40 seconds to submit an answer. The interface will show a numbered grid together with the image containing the difference. To solve the task, enter one of the numbers corresponding to a field containing the difference. If the difference is covered by more than one field, each field containing the difference will be evaluated as a correct answer.

You will receive [**Informed treatment:** 0.30 Euros] [**Uninformed treatment:** 0.50 Euros] for each task you solve correctly.

Estimates
Other participants in previous experiments performed all tasks you will face in this experiment. Like you, these previous performers also had to spot the difference between the two images of each video.

After performing each task, you will have the opportunity to estimate the percentage of previous performers that spotted the difference. Therefore, you will watch exactly the same videos as the previous performers. The previous performers also had 40 seconds to submit an answer, just like you, and were also paid according to their performance. [INFORMED TREATMENT ONLY: Before each video, you will receive a guide to the solution of the task. Please note that the previous performers did not receive solution guides.]

At the end of the experiment, the computer will randomly select two videos. Your estimates for these videos will be relevant for your payoff. For each of the payoff-relevant videos, the following holds: If your estimate is within the interval +/− 5 percentage points around the true percentage of previous performers that correctly identified the difference, you receive 12 Euros.

Consider the following example. Assume that for one of the two payoff-relevant videos, 50% of the previous performers correctly identified the difference in this video. If you estimated that 53% of the previous performers spotted the difference, then you will receive 12 Euros. However, if you estimated that 57% of the previous performers spotted the difference, then you will receive 0 Euros.

You will start with three practice videos to familiarize yourself with the procedure. The practice rounds are not payoff relevant. Afterward, the 20 payoff-relevant videos will follow.

**Further procedure**

After you submit your estimates for the 20 videos, you can earn money with additional decision-making problems. Further details will be given during the experiment.

At the end of the experiment, we will ask you to fill in a questionnaire. Even though your answers will not affect your payoff, we kindly ask you to answer the questions carefully.

After you completed the questionnaire, you will be informed about your payoff from performing the tasks, your payoff from your estimates, your payoff from the additional decision making problems, as well as your total
payoff in this experiment. Please remain seated until the experimenter lets you know that you may collect your payment.

Do you have questions? If yes, please raise your hand. We will answer your questions in private.

Thank you for participating in this experiment!
Comprehension questions

1. How many (potentially payoff-relevant) videos will you evaluate in total?

2. How many of these videos will be selected for payment regarding your estimate of the previous performers?

3. What are the components of your total payoff?

4. Assume the computer randomly selected the following exemplary videos for payment regarding your estimates

   Video a): 62% of the previous performers found the solution.
   Video b): 35% of the previous performers found the solution.

   What is your payoff from your estimates if you estimated that in video a) 57% and in video b) 36% of the previous performers spotted the difference?

5. What is your payoff from your estimates if you estimated that in video a) 87% and in video b) 29% of the previous performers spotted the difference?
5.5.2 Instructions for agents (translated from German)

Welcome to the experiment!

The experiment that you will be participating in is part of a project funded by the German Research Foundation (DFG). It aims to analyze economic decision-making.

You are not allowed to use any electronic devices or to communicate with other participants during the experiment. You may only use programs and features that are allocated to the experiment. **Do not talk to any other participant.** Please raise your hand if you have any questions. We will then approach you and answer your question in private. Please do not under any circumstances raise your voice while asking a question. Should the question be relevant for everyone, we will repeat it aloud and answer it. If you break these rules, we may have to exclude you from the experiment and from receiving payment.

You will receive a show-up fee of 8 Euros for your attendance. You can earn additional money through the experiment. The level of your earnings depends on your decisions, the decisions of participants in former experiments, and on chance. The instructions are the same for everyone. A detailed plan of procedures and the conditions of payment will be explained below.

Tasks

You will face 20 tasks in the course of the experiment. For each task, you will be shown a short video. Each video consists of *two images* that are shown *alternately*. Your task is to spot the difference between the images.

Figure 1 shows an example. Image A of the example shows a kayaker on the left side. In image B, the kayaker is not present.

The duration of each video is 14 seconds. After each video, you will have 40 seconds to submit an answer. The interface will show a numbered grid together with the image containing the difference (see Figure 2).

In each video, the difference between the images covers at least two fields. To solve the task, enter *one* of the grid numbers corresponding to a field containing the difference. That is, the number of any field containing the difference will be evaluated as a correct answer.
The experiment consists of 20 such tasks. The order of the tasks was randomly determined by the computer and is the same for all participants. You will receive 0.50 Euros for each task you solve correctly.

**Previous participants**

**Performers**

Participants in previous experiments performed all tasks you will face in this experiment. Like you, these *previous performers* had to spot the difference between the two images in each video. They were also paid according to their performance.

**Evaluators**
The degree of difficulty of the tasks that have been performed by the previous performers (and will be performed by you in this experiment) has been evaluated by further participants of previous experiments.

These evaluators were shown the tasks in the same way as the previous performers (and you), including the 40-seconds response time. After watching a video, the evaluators estimated the fraction of previous performers that solved this task correctly. The evaluators were paid according to the accuracy of their estimates.

[Informed treatment only: In contrast to the previous performers (and you), the evaluators received guides to the solution before each task. The evaluators were informed that the previous performers did not receive solution guides.]

At the beginning of the experiment, one of the evaluators will be randomly matched to you.

**Insurance decision**

During the experiment, you can earn additional money through insurance decisions. You will make one insurance decision after each task. At the end of the experiment, one of your insurance decisions will be randomly selected for payment.

Your endowment for each insurance decision is 10 Euros.

**Not buying insurance**

If you do not buy the insurance, your payoff depends on the following factors:

1. the number of previous performers (percent) who solved the current task + 10 (percent),
2. the number of previous performers (percent) who solved the current task in the evaluator’s estimation.

If (1) is at least as high as (2), you will keep your endowment of 10 Euros. If (1) is smaller than (2), you will lose your endowment; that is, you will receive 0 Euros.
In other words, if you do not buy the insurance, your payoff will be determined as follows: You will keep your endowment of 10 Euros if the evaluator’s estimate of the performance of the previous performers is correct, an underestimation, or an overestimation by not more than 10 percentage points. Otherwise, you will lose your endowment.

Buying insurance

You have the opportunity to insure against this risk. The insurance costs 6 Euros. If you buy the insurance, you will receive your endowment minus the cost of insurance, that is, 4 Euros.

Example 1

Assume that 50% of the previous performers actually solved the task. In the evaluator’s estimation, 60% of the previous performers solved the task. Of course, you will not learn these values during the experiment.

If you did not buy the insurance, you will keep your endowment, because 50% + 10% ≥ 60%. The payoff from your insurance decision will therefore be 10 Euros.

If you bought the insurance, you will receive your endowment minus the cost of insurance, that is, 4 Euros.

Example 2

Assume again that 50% of the previous performers actually solved the task. In the evaluator’s estimation, 20% of the previous performers solved the task. As in the previous example, you will receive 10 Euros if you did not buy the insurance and 4 Euros if you bought the insurance.

Example 3

Assume again that 50% of the previous performers actually solved the task. In the evaluator’s estimation, 70% of the previous performers solved the task. In this example, you will receive 0 Euros if you did not buy the insurance and 4 Euros if you bought the insurance.

Summary and further procedure

The experiment consists of 20 rounds in total. At the beginning of each round, you will work on the task; that is, you will receive 0.50 Euros if you spot the difference between the two images in the video.
After each task, you can earn additional money through an insurance decision. At the end of the experiment, one of your insurance decisions will be randomly selected for payment. Because you do not know which insurance decision is selected for payment, you should treat each insurance decision as if it were payoff relevant.

After you make your insurance decision, you will receive feedback with a guide regarding the solution to the current task. [INFORMED TREATMENT: The evaluators received the same guide before watching the task.] [UNINFORMED TREATMENT: The evaluators did not receive solution guides.]

You will start with five practice rounds to familiarize yourself with the procedure. The practice rounds are not payoff relevant. Afterward, the 20 payoff-relevant rounds will follow.

After the 20 rounds, you can earn money through additional decision-making problems. Further details will be given during the experiment.

At the end of the experiment, we will ask you to fill in a questionnaire. Even though your answers in this part will not affect your payoff, we kindly ask you to answer the questions carefully.

Do you have questions? If yes, please raise your hand. We will answer your questions in private.

Thank you for participating in this experiment!

**Comprehension questions**

1. When do you make your insurance decision—before or after you watched the task?

2. Are the following statements true or false?

   (a) The evaluators received a guide to the solution to each task.

   (b) The evaluators did not receive guides to the solution to the tasks.

   (c) The evaluators received a guide to the solution before watching and evaluating the task.

   (d) At the end of the experiment, one of my insurance decisions will be randomly selected by the computer for payment.
(e) My payment at the end of the experiment consists of the show-up fee (8 Euros), the payoff from performing the tasks (0.50 Euros per task solved), the payoff from one insurance decision, and the payoff from further decision-making problems.

3. Assume the computer randomly selected videos for payment with the following exemplary characteristics. Provide the payoff from your insurance decision for each example.

Example a)
- 62% of the previous performers found the solution.
- The evaluator randomly matched to you estimated that 72% of the previous performers spotted the difference.
- You did not buy the insurance.

Example b)
- 62% of the previous performers found the solution.
- The evaluator randomly matched to you estimated that 75% of the previous performers spotted the difference.
- You did not buy the insurance.

Example c)
- 34% of the previous performers found the solution.
- The evaluator randomly matched to you estimated that 41% of the previous performers spotted the difference.
- You did not buy the insurance.

Example d)
- 34% of the previous performers found the solution.
- The evaluator randomly matched to you estimated that 30% of the previous performers spotted the difference.
- You did not buy the insurance.

4. Consider again the examples in 3. For each example, assume you did not buy the insurance. What is your payoff from your insurance decisions in each of the examples?