Preferences for observable information in a strategic setting: An experiment

Adam Zylbersztejn, Zakaria Babutsidze, Nobuyuki Hanaki

GATE LYON SAINT-ÉTIENNE
UMR 5824

93, chemin des Mouilles
69130 Ecully - France

Maison de l'Université, Bâtiment B
10, rue Tréfilerie
42023 Saint-Etienne cedex 02 - France

http://www.gate.cnrs.fr
gate@gate.cnrs.fr

WP 1936 – December 2019

Preferences for observable information in a strategic setting: An experiment

Adam Zylbersztejn, Zakaria Babutsidze, Nobuyuki Hanaki

Abstract:

We experimentally investigate how much value people put in observable information about others in strategic interactions. The incentivized experimental task is to predict an unknown target player's trustworthiness in an earlier hidden action game. In Experiment 1, we vary the source of information about the target player (neutral picture, neutral video, video containing strategic content). The observed prediction accuracy rates then serve as an empirical measure of the objective value of information. In Experiment 2, we elicit the subjective value of information using the standard stated preferences method ("willingness to accept"). While the elicited subjective values are ranked in the same manner as the objective ones, subjects attach value to information which does not help predict target behavior, and exaggerate the value of helpful information.

Keywords:
Prediction, observable information, individual characteristics, stated preferences, willingness to accept, experiment

JEL codes:
C72, D83

Working Paper

Preferences for observable information in a strategic setting: An experiment*

Adam Zylbersztejn[†] Zakaria Babutsidze[‡] Nobuyuki Hanaki[§]

December 18, 2019

Abstract

We experimentally investigate how much value people put in observable information about others in strategic interactions. The incentivized experimental task is to predict an unknown target player's trustworthiness in an earlier hidden action game. In Experiment 1, we vary the source of information about the target player (neutral picture, neutral video, video containing strategic content). The observed prediction accuracy rates then serve as an empirical measure of the objective value of information. In Experiment 2, we elicit the subjective value of information using the standard stated preferences method ("willingness to accept"). While the elicited subjective values are ranked in the same manner as the objective ones, subjects attach value to information which does not help predict target behavior, and exaggerate the value of helpful information.

Keywords: Prediction, observable information, individual characteristics, stated preferences, willingness to accept, experiment
JEL Code: C72, D83

*We acknowledge the support from the following programs operated by the French National Research Agency (Agence Nationale de Recherche): UCA^{JEDI} (ANR-15-IDEX-01) and LABEX CORTEX (ANR-11-LABX-0042) being part of Université de Lyon (ANR-11-IDEX-007). We are grateful to the editor and three anonymous referees, Astrid Hopfensitz, Carsten de Dreu, Charles Noussair, Sigrid Suetens, Chia-Jung Tsay, Boris van Leeuwen, the participants to the 2019 ASFEE conference in Toulouse, the 7th Workshop on Non-market Valuation in Marseille and the 2019 ESA European Meeting in Dijon for helping us improve the paper. Ynès Bouamoud, Ann Thampi and Vincent Théroude provided quality research assistance. Quentin Thévenet provided valuable assistance with software programming.

[†]Univ Lyon 2, Université Lumière Lyon 2, GATE L-SE UMR 5824, 69130 Ecully, France

[‡]SKEMA Business School, Université Côte d'Azur (GREDEG) and OFCE, Sciences Po Paris

[§]Institute of Social and Economic Research, Osaka University

Your face, my thane, is as a book where men

May read strange matters. To beguile the time,

Look like the time. Bear welcome in your eye,

Your hand, your tongue. Look like the innocent flower,

But be the serpent under't.

Lady Macbeth in William Shakespeare's *Macbeth* (act 1, scene 5)

1 Introduction

Notwithstanding standard theoretical predictions, economic cooperation often emerges among strangers facing adverse individual incentives (Camerer, 2003). One of the central behavioral patterns documented in the experimental literature is conditional cooperation: humans are willing to act cooperatively as long as others do the same (Fischbacher et al., 2001; Kocher et al., 2008). Thus, the ability to recognize other people's willingness to cooperate facilitates efficient social relationships and helps avoid exploitation by others. Recent literature shows that the information gathered through observing certain physical characteristics – one's face, body language, way of expression (sometimes referred to as "thin slices" of observable information) – may improve cooperation detection in various economic games (Bonnefon et al., 2017). In some of these studies, the stimulus is *loaded* with strategic content – for instance, when individuals are observed while communicating with other players prior to deciding whether to cooperate with them or not (Belot et al., 2012), or at the very moment of decision making in the game (Verplaetse et al., 2007). In other studies, the stimulus is *neutral*, i.e. completely unrelated to the strategic interaction itself, such as video recordings that show prospect players making improvised statements prior to learning about the game they are about to play (Fetchenhauer et al., 2010; Vogt et al., 2013), or images of neutral facial experessions (De Neys et al., 2015), and even if the time of exposure to such facial cues is measured in split seconds (De Neys et al., 2017).

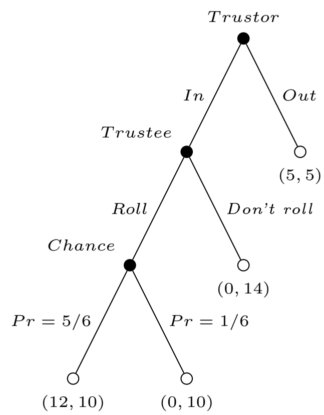
Importantly, although observable information about people's physical characteristics may constitute a strategic asset in social interactions, not all sources of information are equally useful.

On the one hand, the presence of strategic content may improve prediction accuracy. Verplaetse et al. (2007) report that loaded pictures (grasping the facial expression at the very moment of the decision making in the game) are more helpful in predicting cooperation in the prisoner's dilemma game than the ones taken in a neutral setting. In a related experiment, Sylwester et al. (2012) find that such loaded pictures may nonetheless underperform compared to loaded videos showing a short extract of pre-play communication between players in the game. On the other hand, providing a richer set of cues based on physical appearance does not per se improve the ability to detect cooperativeness, and may even be inhibitory with that respect. For instance, Bonnefon et al. (2013) report that standard passport-like pictures including non-facial features (e.g., one's hair and clothing) do not improve trustworthiness detection beyond chance, in contrast to cropped pictures only showing faces. In a similar vein, Van Leeuwen et al. (2018) report that pictures showing neutral facial expressions help predict the behavior of the second mover in the ultimatum game, yet muted neutral videos showing the same faces do not.

If observable information has the capacity to foster efficiency of human interactions, it also bears an objective economic value. In the context of a trust game experiment, Eckel and Petrie (2011) show that many people exhibit a preference (which generates economic demand) for acquiring this kind of information about interaction partners. The central question we are asking in this paper is: To what extent do these preferences reflect the actual value of information? Existing evidence, albeit scarce, suggests that people do not efficiently distinguish between more and less valuable sources of observable information. In the aforementioned paper, Bonnefon et al. (2013) show that once given a direct choice, people predominantly favor the source of information with a lower objective value (i.e., a passport-style picture instead of a cropped one). Similar evidence is provided by Tsay (2013) in a non-strategic setting. She investigates people's (novices as well as experts) ability to predict how well piano players performed in a professional contest. Comparing three different sources of information – either purely auditory, or purely visual, or a combination of both, she reports that looking at the musicians' performance overrides listening to their music, as well as having both sources of information combined. However, despite its relatively low objective value, the latter source happens to be chosen by the vast majority of subjects. Note, however, that in these two studies the preferred sources of information also happen to be the richest ones, so that the presence of confounding non-strategic motives (like mere curiosity, or more broadly the hedonic value of observable information) cannot be excluded.

In the present study, we use the classic “thin slice” paradigm in a novel design to investigate whether people rationally value observable information, i.e. whether the subjective value of information is aligned with its objective value. We experimentally implement a prediction task in which subjects determine whether a target person was trustworthy or not as a trustee in an earlier hidden action game with one-way, pre-play communication. Our treatments vary the sources of observable information about the target person. Our empirical strategy requires a variation in the prediction accuracy (and thus, the objective value of information) across treatments. To achieve this, Experiment 1 implements a set of distinct sources of information that differ along the two dimensions outlined in the previous paragraph: the richness of cues (neutral picture, henceforth PHOTO vs. neutral video, henceforth VIDNE), and the presence of strategic content (loaded video, henceforth VIDLO, showing a pre-play statement the target person made in front of other players in the hidden action game). We demonstrate that the resulting prediction accuracy systematically varies across the three conditions: PHOTO and VIDNE generate significantly lower prediction accuracy than VIDLO. We then assess the extent to which the subjective value of information (measured in Experiment 2) is aligned with its objective value. Importantly, our design rules out the presence of any non-strategic confounds such that the subjective value of information (elicited in an incentivized way through a standard willingness to accept procedure) stems solely from its perceived usefulness in the prediction task. Notwithstanding the previous findings, our data point to a consistent valuation of observable information in a strategic setting: The elicited subjective values are ranked in the same manner as the objective values. However, we also document important misperceptions of the objective value of information. First, subjects attach value even to those sources of information that are not helpful in predicting target behavior. Second, the perceived value of helpful information is substantially overstated.

Figure 1: Experimental hidden action game



2 Experimental design, procedures, and hypotheses

We run two sets of experiments (henceforth Experiment 1 and Experiment 2) with a total of $N = 287$ subjects. Twelve sessions (six per experiment) were run between October and December 2018 in the GATE-LAB, the experimental laboratory of the GATE Lyon-Saint-Etienne Research Institute in France. The experiment was entirely computerized: subjects were recruited from the GATE-Lab subject pool using ORSEE (Greiner, 2015), and all the experimental tasks were programmed in z-Tree (Fischbacher, 2007).

Background data for the prediction task in Experiments 1 and 2. For implementing the prediction task, we use the dataset previously reported in Babutsidze et al. (2019). That study is based on the hidden action game by Charness and Dufwenberg (2006) presented in Figure 1. All payoffs are in Euros. The game is played between two parties: the trustor and the trustee. The trustor may either choose an outside option *Out* which yields 5 to both players and ends the interaction, or go *In*. Then, the trustee may either choose to *Roll* a die (which yields 12 to the trustor and 10 to the trustee with the probability of 5/6, and 0 to the trustor and 10 to the trustee with the probability of 1/6), or not to *Roll* (yielding 0 to the trustor and 14 to the trustee with certainty). This game provides a simple setting for studying cooperation in a principal-agent relationship with moral hazard: incentives are not aligned between the two parties, and earning 0 is not perfectly informative for the trustor about the trustee's action.

Like Charness and Dufwenberg (2006), we simultaneously elicit both players' decisions. Namely, the trustee makes a decision without knowing the trustor's move, and that decision is only implemented had the trustor gone *In*. The game is preceded by a pre-play stage with face-to-face communication and is implemented as follows. In every experimental session, six trustors are seated in one room (in separate cubicles and without a possibility to communicate) where they make all their decisions in the game. Each of the six trustees, in turn, makes an individual decision in a separate room. Prior to the decision-making stage of the game, each trustee is given approximately two minutes to prepare a short statement for the trustors. At this point, we provide an additional set of instructions emphasizing the fact that the statement may affect the trustors' decisions and, consequently, the trustee's gain from the experiment.[1] Then, the trustee enters the trustors' room and delivers the statement in front of them. The trustors can clearly see and hear the trustee, and the trustee can also observe the trustors while delivering the statement. After that, the trustee leaves to a separate room to make a decision. Simultaneously, the six trustors privately make their decisions. At the end of the experiment, the trustees and the trustors are randomly and anonymously matched into six pairs for payoffs. Further implementation details, including the instructions used in that experiment, are provided in Appendix A.

In addition to the trustees' decisions in the experimental game (and, if relevant, the outcomes of die rolls), our dataset contains several recordings. Following Van Leeuwen et al. (2018), upon arrival to the laboratory and before learning about the rules of the hidden action game, each subject in the role of a trustee is invited to a separate room for a mugshot picture and a standardized video recording (reading a short extract from a printer instruction manual, while keeping a neutral facial expression). These two sources of information are used, respectively, in our PHOTO and VIDNE ("neutral video") treatments. Finally, the trustees are also video recorded while making a statement in the pre-play communication stage of the hidden action game. We use this information in our VIDLO ("loaded video") treatment.

The database has been collected at Laboratoire d'Economie Expérimentale de Nice (LEEN) of

[1] This information is part of the summary of the hidden action game experiment provided in the instructions for Experiments 1 and 2. See Appendix B for details.

the University of Nice, France. It contains information about 41 trustees (20 females; average age 22.51, SD 4.70) which are the target agents in the prediction tasks implemented in Experiments 1 and 2.[2] 61% of them decided to *Roll*, while 51.2% included a promise to *Roll* in their statements.[3]

Design of Experiment 1. Participants (N = 132; 68 females; average age of 20.64, SD 1.87) make a series of twenty predictions of trustees' behavior in an earlier hidden action game (i.e., whether the target person rolled a die or not).[4] Each time, a trustee is randomly drawn without replacement from the main sample of 41 observations. A correct (an incorrect) prediction is worth 10 (2) euros, no feedback is provided from one prediction to the other, and two rounds out of twenty are randomly drawn for payoff at the end of each experimental session. Note that unlike some previous studies using the "better than chance" paradigm, we do not constrain the base rate of "success" at the chance level of 50%.[5] Our experimental treatments are implemented in a between-subject manner and progressively enrich the set of information about the trustee that is provided to the subject prior to making a prediction: either a mugshot picture (PHOTO; N = 44), or a neutral video recording (with sound) showing that person making a non-strategic statement that has been recorded before (and independently of) the experimental hidden action game (VIDNE;

[2]There were seven experimental sessions in total, each having six trustors and six trustees. However, as explained in the Appendix A, the data for one trustee are incomplete and had to withdrawn from the analysis.

[3]Like Charness and Dufwenberg (2006), we define a promise as a statement of intent to *Roll*. Our classification method echoes the recent study by Schwartz et al. (2018). All statements were classified as promises or no-promises by two independent coders (research assistants). The first coder classified the content of messages while preparing the transcripts of player Bs' statements. Then, another coder received the full list of 41 transcripts and independently classified each of them. Ties were broken by one of the authors. The resulting conditional likelihoods of *Roll* following a promise vs. without a promise to *Roll* equal 71.4% vs. 50%.

[4]Due to minor technical glitches in the laboratory – visual content not being displayed properly, or sound being muted – we lost the data from 4 individual predictions (involving 3 subjects) in VIDLO and 8 individual predictions (involving 6 subjects) in VIDNE. Since for those few individuals the ultimate number of predictions is below 20, there are marginal differences between the aggregate accuracy rates reported in Figure 2a and the average individual accuracy rates outlined in Figure 2b and Table 1.

[5]Under the "better than chance" paradigm, subjects typically receive randomly generated pairs of stimuli – one coming from a person that exhibited certain behavior, and one from another person that did not (which is common knowledge; see, e.g., Bonnefon et al., 2013; Van Leeuwen et al., 2018). Another method is to show a series of individual stimuli and inform the subjects about the underlying base rate (50%) of a given behavioral outcome, but not about the length of the series (Vogt et al., 2013). Although the "better than chance" paradigm provides a clean and simple benchmark for measuring the extent to which observable information affects prediction accuracy, it has been criticized for the lack of external validity. As pointed out by Todorov et al. (2015), this criterion seems weak when it comes to evaluating prediction performance in many real-world environments in which the different types of behavior are unequally prevalent. Following this argument, in our experiment the lack of information about the underlying base rate adds to the overall complexity of the prediction task. See Fetchenhauer et al. (2010) for a similar approach.

$N = 43$), or a loaded video recording (with sound) in which the trustee makes a strategic pre-play statement in front of the trustors (VIDLO; $N = 45$).[6]

Based on the previous prediction experiments varying the sources of observable information in ways similar to ours (Verplaetse et al., 2007; Sylwester et al., 2012; Bonnefon et al., 2013; Van Leeuwen et al., 2018, as reviewed in the introduction), we make the following set of hypotheses regarding the expected prediction accuracy across treatments:

Hypothesis 1

a. Having strategic content as part of observable information improves prediction accuracy (VIDLO > VIDNE).

b. Increasing the richness of cues in neutral settings does not improve prediction accuracy (VIDNE ≈ PHOTO).

This hypothesis implies the following ordering of the prediction accuracy rates across the experimental treatments: VIDLO > VIDNE ≈ PHOTO.

On a final note, our design makes the acquaintance between the experimental subjects in Lyon and the trustees recorded in Nice unlikely, so that one may plausibly assume that performance in the prediction task actually measures the individual capacity to detect cooperativeness in strangers.[7]

Design of Experiment 2. The second experiment ($N = 145$; 84 females; average age of 20.87, SD 1.57) is based on the one-shot version of the prediction task used in the first experiment. We rely on the stated preferences approach to eliciting participants' subjective valuation of the different sources of information. We use a within-subject design to elicit individual "willingness to accept" (WTA) via the standard Becker-DeGroot-Marschak method (Becker et al., 1964). This

[6]The average duration of a recording in VIDNE (VIDLO) is 33.29 (26.39) seconds with SD 4.90 (2.09). Given that PHOTO only involves static content, in this treatment we adopted the following procedure. Each time, the picture of the target person is displayed on the computer screen. After 15 seconds, a button appears underneath the picture allowing the subject to move on to the prediction-making stage. This choice came about as the outcome of the pilot test of our experimental setup, and appears to be a remedy against the risk of "under-exposing" – the exposure to the displayed content being insufficient to fully grasp all the available information, as well as "over-exposing" – participants eventually getting inattentive due to factors such as boredom, impatience, or a decay in their interest in the displayed static content. For similar reasons, we chose to use a random subsample of 20 out of 41 items rather than the whole set of available stimuli.

[7]See Centorrino et al. (2015) and Van Leeuwen et al. (2018) for a similar empirical strategy.

method is incentive-compatible (i.e., it provides incentives for a truthful revelation of certainty equivalent, henceforth CE) and runs as follows.

In the first step, we ask each subject to state a certainty equivalent (between 21 and 100 points, with a conversion rate of 1 point=0.10 euro) they would be willing to accept for not having to make a prediction in each of our three main environments of interest, i.e. PHOTO, VIDNE, VIDLO, as well as an additional environment where no specific information is available about the trustee from the previous hidden action game (NOINF).[8] The NOINF condition grasps the subjective value of the prediction task *per se* (i.e., with no additional information about the trustee) which we use as a baseline for estimating the subjective value of the additional information provided in PHOTO, VIDNE, and VIDLO conditions. Note that the participants are only provided a short description of the kind of information provided in each of the four environments, but do not inspect any specific content at this point.[9] We see this feature of our experiment as an additional layer of complexity which, if anything, makes it harder to the subjects to correctly assess the actual value of information. Thus, this part of our design keeps our experimental testbed for rational valuation of observable information on the conservative side.

In the second step, one of the four environments is chosen at random. The computer also generates a uniform random number between 20 and 99. If the CE corresponding to the randomly chosen environment exceeds the number randomly generated by the computer, then the subject is asked to inspect the informational content (if any) of a given environment, and then to make a prediction.[10] Otherwise, the subject earns an amount of points (later converted into euros) equal to the random number, and asked to inspect the informational content without making any prediction. Hence, in stark contrast to the previous experiments (Bonnefon et al., 2013; Tsay, 2013), in our experiment subjects do not directly choose the source of information (which is exogenously determined by the computer), and are always given the opportunity to inspect its content regardless of whether it is part of the incentivized prediction task or not. Thus,

[8]The range of CE is defined by the payoff scheme in the prediction task – 20 points for an incorrect prediction, and 100 points for a correct one.

[9]This element of our design differs from Bonnefon et al. (2013) who provide exemplary samples prior to letting subjects choose the source of information.

[10]For PHOTO, VIDNE and VIDLO, the prediction-making procedure is analogous to a single round in Experiment 1.

our design guarantees that the stated preferences reflect the strategic value of information in a prediction exercise rather than any other confounding motives (such as the prediction-maker's curiosity about the target person). We can make the following hypothesis:

Hypothesis 2 *The stated preferences for observable information are aligned with its empirical value in the prediction task, so that the elicited subjective values run in the following order: VIDLO> VIDNE≈PHOTO.*

Between these two steps (and without prior notice), we also elicit subject's beliefs about the prediction accuracy rates previously observed in Experiment 1. This procedure is based on the Most Likely Interval (MLI) elicitation rule due to Schlag and van der Weele (2015). For each of the three conditions, the participant is asked how many predictions out of 100 were accurate, and to choose one of the following intervals: ± 5, ± 10, or ± 15. The task is incentivized as follows. An answer is considered as correct if the actual prediction accuracy rate from Experiment 1 lies within the chosen band from the stated belief. A participant is only rewarded for providing three correct answers (i.e., a correct answer for each of three conditions). Giving three correct answers with interval ± 5 is worth 10 euros, and whenever interval ± 10 (± 15) is chosen instead of ± 5, this amount decreases by 1.50 (3) euros.[11]

Importantly, the rich information on individual heterogeneity (beliefs about the prediction accuracy rates in Experiment 1, the degree of confidence in those beliefs, as well as individual risk preferences – see the experimental procedures below) allows us to advance our investigation of stated preferences for information beyond the plain ordinal scope of Hypothesis 2. We focus on (*i*) the extent to which beliefs in Experiment 2 are consistent with the actual outcomes of Experiment 1, and (*ii*) whether and how the elicited subjective values for the three sources of information deviate from the belief-based expected values (controlling for confidence and risk preferences).[12]

[11]As an example, suppose a person provided three correct answers: two with interval ± 5 and one with interval ± 15. Then, the corresponding gain equals $10 - 3 = 7$ euros. The same gain would occur if a person provided three correct answers with a different set of intervals – one ± 5 and two ± 10 (since $10 - 1.50 - 1.50 = 7$ euros).

[12] We elicit risk preferences using the lottery task by Gneezy and Potters (1997). A decision-maker holds an initial endowed of 100 ECU (equivalent of 2.5 EUR), some (or all) of which he can invest in the following lottery: 50% chance of multiplying the investment by the factor of 2.5 times, and 50% chance of losing the invested amount. Investing an amount below 100 indicates risk aversion and the lower the investment, the stronger the risk aversion. Note that the Gneezy-Potters method does not allow to assess if the subjects investing their whole endowment are risk neutral or risk seeking.

This allows us not only to test Hypothesis 2 (which is solely about the ordering among three elicited subjective values), but also look at the relationship between the elicited subjective values and the reference expected values implied by individual beliefs:

Hypothesis 3

a. Participants form accurate beliefs about the objective value of information.

b. Controlling for individual confidence and risk preferences, the subjective values of information match the values implied by the beliefs about the empirical value of information.

We note that the underlying assumption behind Hypothesis 3b is that the empirical value of information is indeed *objective*, so that one's beliefs about the empirical value of information coincide with one's beliefs about the value of that information for *oneself*. This may not be the case if, for instance, someone considers him or herself better than average in "reading" human faces. Since our dataset does not contain information about one's relative confidence in one's own abilities in the experimental task, our experimental design does not suffice to test the validity of this assumption. We acknowledge this is a caveat for the empirical analysis that follows.[13]

Experimental procedures. Upon arriving to the lab, subjects are seated in individual cubicles and informed about the general rules of a lab experiment.[14] The preliminary part of the session consists of a basic socio-demographic questionnaire (age, gender, education, major, current occupation, score at the *baccalauréat* exam at the end of high school), as well as a set of (moderately) incentivized and non-incentivized computerized tasks designed to measure specific individual characteristics.[15] After that, subjects receive paper instructions describing the details of the previous hidden action game experiment, as well as their own experimental task. Those instructions are read aloud by the experimenter, any remaining questions are immediately answered, and the experiment moves to its main stage, as described above. In addition to earnings in the experimental

[13]We thank an anonymous referee for pointing this out.

[14]A complete set of experimental instructions (translated from French to English) can be found in Appendix B.

[15]This procedure follows Babutsidze et al. (2019), and its details may be found therein. The set of tasks includes standard measures of other-regarding preferences (Social Value Orientation, SVO, task by Murphy et al., 2011), cognitive skills (3-item Cognitive Reflection Test, CRT, Frederick, 2005), the theory of mind (The Reading the Mind in the Eyes Test, RMET, Baron-Cohen et al., 2001), risk preferences (Gneezy and Potters, 1997) described in footnote 12, and trust attitudes (based on the German Socio-Economic Panel Study, SOEP).

Table 1: Predicted vs. actual behavior: regression analysis

Treatment:	VIDLO		VIDNE		PHOTO	
	coeff. (SE)	p	coeff. (SE)	p	coeff. (SE)	p
Intercept (α_0)	0.491	<0.000	0.476	<0.000	0.471	<0.000
	(0.039)		(0.038)		(0.037)	
$1[ActualRoll]$ (α_1)	0.088	0.005	-0.019	0.532	-0.019	0.529
	(0.030)		(0.031)		(0.030)	
N of obs./clusters	896/45		852/43		880/44	

Note. Results of OLS regression models of the individual prediction (indicator variable $1[PredictionRoll] = 1$ if one predicts that the second-mover rolled a die in the previous experiment; 0 otherwise) on the indicator variable $1[ActualRoll]$ (set to 1 if the second-mover actually rolled a die in the previous experiment, and to 0 otherwise). Observations are clustered for each individual, standard errors are cluster-robust.

tasks, each subject earns a show-up fee of 5 euros. The duration of a session is approximately 1h30 and the average total payoff is 22.52 (16.45) euros in Experiment 1 (2).

3 Results

The goal of the first experiment is to estimate the objective value of the different sources of information in the experimental prediction task (testing Hypothesis 1). In the second experiment, we elicit the subjective value of those sources of information using the standard stated preferences approach ("willingness to accept"). The treatment effects on the experimental variables are evaluated using standard parametric and nonparametric methods. As a robustness check, in Appendix D we report Bayes factors associated with the main comparisons of interest. Our main finding is that the elicited values and beliefs are aligned with the objective value of information (Hypotheses 2 and 3 are thus confirmed).

3.1 Experiment 1: the objective value of information

We first draw a statistical link between the predicted behavior and the actual behavior depending on the source observable information by testing Hypothesis 1. Let p_R (p_{DR}) be the likelihood of making a prediction *Roll* conditional on the target person actually choosing to *Roll* (*Don't roll*). $p_R = p_{DR}$ implies that subjects are unable to discriminate between trustworthy and untrustworthy target players, and make a prediction *Roll* at a constant rate (freely ranging between 0 and 1) irrespective of the trustee's underlying type. $p_R > p_{DR}$, in turn, implies that subjects do distinguish between target players' types and are more likely to make a prediction *Roll* for those who actually rolled a die in the hidden action game.[16] Following this intuition, for each treatment we regress an indicator variable $1[PredictionRoll]$ (set to 1 if one predicts that the target person rolled a die in the previous experiment, and to 0 otherwise) on another indicator variable $1[ActualRoll]$ (set to 1 if the target person actually rolled a die in the previous experiment, and to 0 otherwise). The intercept (denoted α_0) captures the aggregate likelihood of predicting *Roll* for those trustees that did not actually roll a die (such that $\alpha_0 = p_{DR}$). Our key measure of interest is the coefficient of the explanatory variable (denoted α_1) which captures the extent to which subjects are able to distinguish between those who rolled and those who did not based on a given source of information (so that $\alpha_1 = p_R - p_{DR}$). The resulting estimates are provided in Table 1. Coefficient α_1 is found to be close to zero and insignificant for PHOTO and VIDNE, suggesting that these sources of information do not suffice to distinguish between the two types of trustees. In VIDLO, α_1 is positive and significant, meaning that the predictions become adjusted to the trustee's actual type. Figure 2a summarizes the resulting aggregate rates of accurate predictions which equal: 55.36% in VIDLO, 48.36% in VIDNE, and 48.18% in PHOTO. Figure 2b disaggregates those data and shows the distributions of individual accuracy rates.[17] For those data, the Epps-Singleton test (Epps and Singleton, 1986) rejects the null hypothesis of identical distributions when comparing VIDLO with either VIDNE ($p = 0.007$) or PHOTO ($p = 0.014$), but does not reject it when comparing VIDNE with PHOTO ($p = 0.255$).[18] The parametric analysis reported in Table 2 corroborates these findings. In conclusion, we formulate the following result:

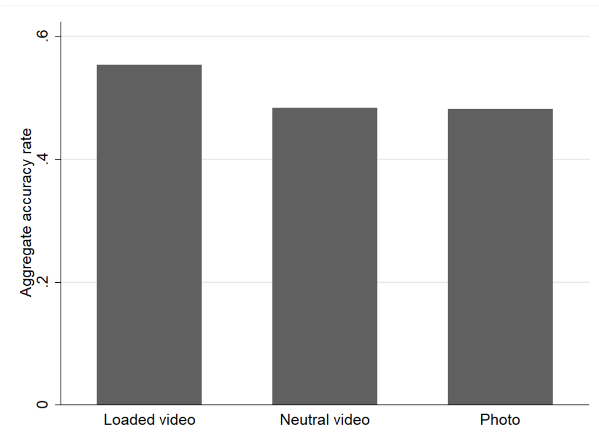
Result 1 (Objective value of information) *Prediction accuracy does not vary between differ-*

[16]For a perfect ability to discriminate between the two types of trustees, we would have $p_R = 1$ and $p_{DR} = 0$.

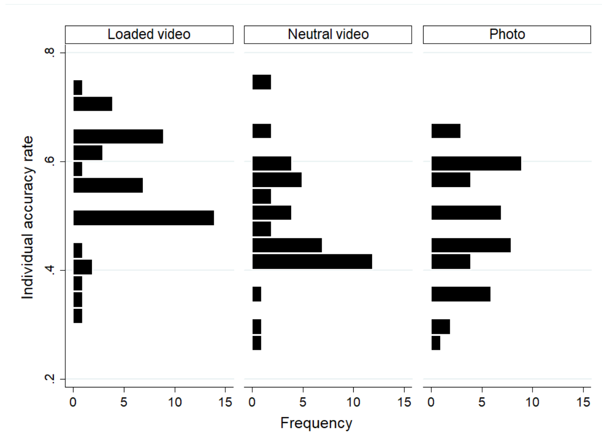
[17]In line with the previous findings by Sylwester et al. (2012), there is a systematic lack of correlation between the individual prediction accuracy rate and the score in The Reading the Mind in the Eyes Test of Baron-Cohen et al. (2001).

[18]The Epps-Singleton test has two main advantages compared to other nonparametric tests that became commonplace in the experimental literature, such as the Kolmogorov-Smirnov test and the Wilcoxon-Mann-Whitney ranksum test. First, the Epps-Singleton does not require the data to be drawn from a continuous distribution. Second, it attains a greater statistical power. See Forsythe et al. (1994) and Goerg and Kaiser (2009) for details.

Figure 2: Information and prediction accuracy in Experiment 1



(a) Aggregate accuracy rates



(b) Histograms of individual accuracy rates

ent forms of non-strategic information provided in PHOTO and VIDNE. Strategic information delivered in VIDLO improves prediction accuracy as compared to any non-strategic content.

This result yields full support to Hypothesis 1.

To shed more light on the possible reasons for which loaded information is more helpful in the prediction task than the neutral one, we now turn to investigating how the content of the target players' statements affects prediction accuracy in the VIDLO treatment. Following a growing body of experimental literature in economics (see, for instance, Charness and Dufwenberg, 2006; Ismayilov and Potters, 2016; Schwartz et al., 2018), our analysis focuses on promise-making as a potential signal of trustworthiness. In the sample of 896 stimuli delivered in VIDLO, 452 contain a promise to *Roll* (see footnote 3 for details of the coding procedure). Such a promise acts as an important predictor of subsequent behavior in the hidden action game: the rate of *Roll* in the subsample with a promise equals 76.11%, compared to 52.25% observed in the subsample with no promise. The participants to our Experiment 1 seem to correctly read this signal of trustworthiness and use it in their predictions. The rate of prediction *Roll* increases from 45.5% for recordings not containing a promise to 63.72% for recording containing such a promise. Importantly, the presence of a promise to *Roll* improves prediction accuracy rate which increases to 58.84% from

Table 2: Individual accuracy across treatments in Experiment 1: regression analysis

	coeff. (SE)	p
Intercept (β_0)	0.553	<0.001
	(0.016)	
$1[VIDNE]$ (β_1)	-0.069	0.002
	(0.022)	
$1[PHOTO]$ (β_2)	-0.071	0.002
	(0.022)	
Additional test:		
$H_0 : \beta_1 = \beta_2$, $p = 0.919$		

Note. Results of an OLS regression of the individual accuracy rate on treatment indicator variables ($1[VIDNE] = 1$ for VIDNE, $= 0$ otherwise; $1[PHOTO] = 1$ for PHOTO, $= 0$ otherwise). The intercept provides the average individual accuracy rate in VIDLO, and the coefficient β_1 (β_2) provides the difference between the average individual accuracy rates in VIDLO and VIDNE (VIDLO and PHOTO). All p-values correspond to two-sided t-test. $N = 132$, $R^2 = 0.094$.

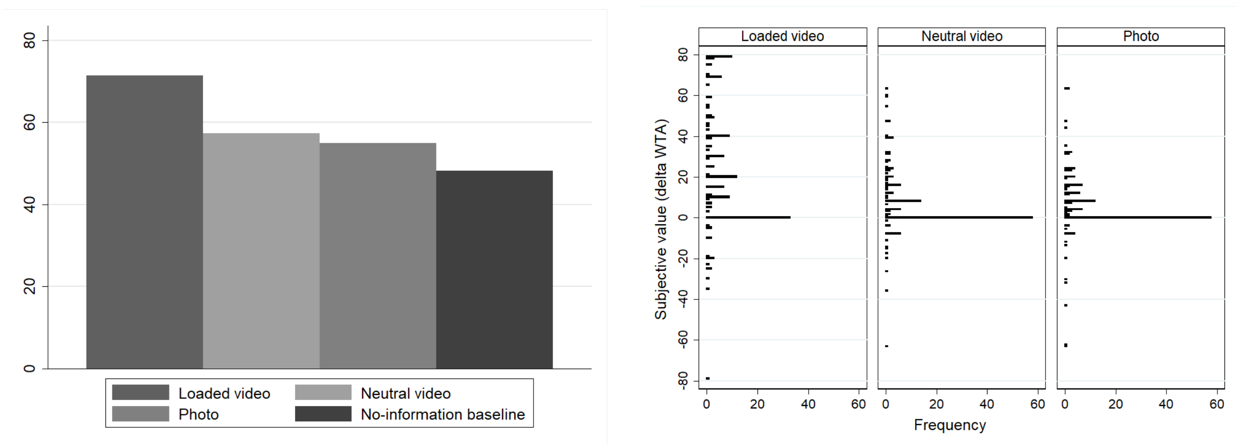
51.80% without a promise (which, in turn, is close to the rates observed in both conditions with neutral information – slightly above 48%). Logistic regression of the accurate prediction indicator variable on the promise-making indicator variable with individual-level standard error clustering shows that the positive effect of a promise on prediction accuracy is statistically significant. The average marginal effect is estimated at 7.05% with $p = 0.045$ (see Appendix C). Altogether, these data suggest that a promise to *Roll* stands out as a component of strategic communication that is capable of reducing the degree of uncertainty about the target player's trustworthiness.

3.2 Experiment 2: the subjective value of information

We now move on to testing Hypothesis 2 based on the result of Experiment 2. Figure 3a summarizes the mean aggregate WTA in the four informational environments of our Experiment 2: $WTA_{VIDLO} = 71.48$, $WTA_{VIDNE} = 57.41$, $WTA_{PHOTO} = 55.01$, and $WTA_{NOINF} = 48.24$.[19] In line with the previous evidence from Eckel and Petrie (2011), we first notice a general preference for receiving observable information about the target person (notwithstanding our previous result that only VIDLO actually allows to distinguish trustworthy players from the untrustworthy ones, as

[19]The most general pattern depicted in this figure, i.e. $WTA_{VIDLO} \geq WTA_{VIDNE} \geq WTA_{PHOTO} \geq WTA_{NOINF}$, is reflected by the stated preferences of a vast majority (65%) of subjects in Experiment 2. At the individual level, we fail to detect any correlation between the elicited WTA and the score in The Reading the Mind in the Eyes Test of Baron-Cohen et al. (2001).

Figure 3: Stated preferences in Experiment 2



(a) Average WTA

(b) Histograms of ΔWTA

shown in Table 1). $\Delta WTA_X \equiv WTA_X - WTA_{NOINF}$ is the difference between one's stated WTA in a given environment with observable information (X is either VIDLO, VIDNE, or PHOTO) and one's WTA in the no-information baseline environment. As noted above, ΔWTA captures the subjects value of receiving additional information provided in PHOTO, VIDNE, and VIDLO conditions. Figure 3b provides the distributions of ΔWTA across treatments. In aggregate, we observe the following mean values: $\Delta WTA_{VIDLO} = 23.14$, $\Delta WTA_{VIDNE} = 9.17$, $\Delta WTA_{PHOTO} = 6.77$ each of which is significantly different to zero ($p < 0.001$, t-test).[20] Paired t-test does not detect a significant difference between ΔWTA_{VIDNE} and ΔWTA_{PHOTO} ($p = 0.144$), but does so once we compare VIDLO to either PHOTO or VIDNE (both $p < 0.001$).[21] This leads us to the main result of the second experiment:

Result 2 (Subjective value of information) *The subjective values of the different sources of information follow the same pattern as their objective values: equivalence between PHOTO and VIDNE, both of which are dominated by VIDLO.*

This result yields full support to our Hypothesis 2.

[20] Analogous nonparametric sign-rank test also yields $p < 0.001$.

[21] Analogous sign-rank test confirms these findings, yielding $p = 0.109$ for the first comparison and $p < 0.001$ for the two remaining ones.

Table 3: Beliefs about the objective value of information and interval measures of confidence

Condition	Accuracy rate in Exp. 1	Mean belief	Interval choice			
			Mean interval	share of "±5%"	share of "±10%"	share of "±15%"
VIDLO	55.36%	69.68%	± 7.66%	51.03%	44.83%	4.14%
VIDNE	48.36%	53.40%	± 9.41%	24.83%	62.07%	13.10%
PHOTO	48.18%	47.14%	± 8.97%	30.34%	60.00%	9.66%

To gain more insight into the formation of the subjective value of information, we exploit the data on how individuals perceive the empirical efficiency of the three sources of information. Table 3 summarizes the individual beliefs about accuracy rates generated by the three sources of information in Experiment 1, as well as the related confidence measurement based on interval choice. Comparing the mean beliefs and the actual accuracy rates, we find that participants' beliefs are fairly aligned with the actual accuracy rates in PHOTO and VIDNE, and tend to be overstated in VIDLO. Moving to pairwise comparisons, we report that the mean stated beliefs are higher in VIDLO than in either VIDNE or PHOTO (both comparisons yield $p < 0.001$ using two-sided t-test). The difference between VIDNE and PHOTO is small yet significant ($p < 0.001$). Turning to the interval measure of confidence reveals that the participants are most confident in their beliefs in VIDLO ($p < 0.001$ compared to either VIDNE or PHOTO, two-sided t-test). This time, however, the average interval chosen in PHOTO is slightly lower than in VIDNE, although this difference falls short of attaining statistical significance at the 5% level ($p = 0.118$). Altogether, these data show that:

Result 3 (Individual beliefs and confidence) *The source of information that generates the highest accuracy rate in Experiment 1 and the highest WTA in Experiment 2 – VIDLO – also dominates the two other conditions in terms of subjects' beliefs about the accuracy rate in Experiment 1 and their confidence in those beliefs.*

This result yields partial support to our Hypothesis 3a. Although beliefs may be inaccurate in absolute terms (which is especially pronounced in VIDLO), their relative ordering meshes well with the main patterns documented in Experiment 1.

Finally, we look at the consistency between the stated preferences and the subjective beliefs about the prediction accuracy rates in Experiment 1. Let $RP_X^i \equiv (100 \times q_X^i + 20 \times (1 - q_X^i)) - WTA_X^i$. RP_X^i be the difference between subject i's WTA under condition X and the subjective expected payoff from the prediction task under condition X given i's belief about the accuracy rate in Experiment 1 (q_X^i). We interpret RP_X^i as the "risk premium" one would be willing to pay to avoid payoff uncertainty from making a prediction under condition X. We conjecture that the stonger the risk aversion and the weaker the confidence about q_X^i, the higher the risk premium.

Table 4 reports the result of a series of Seemingly Unrelated Regression (SUR) models estimated at the treatment level and taking RP_X^i as dependent variable.[22] In Model 1, the explanatory variables are related to risk preferences captured by the amount (between 0 and 100) invested in the Gneezy-Potters lottery task. First, we exploit the information on whether a subject is risk averse or not: an indicator variable $1[RA]$ is set to 1 for risk averse subjects (i.e., the ones not investing their whole endowment in the lottery).[23] Second, for the risk averse subjects ($1[RA] = 1$) we also exploit the information about the degree of risk aversion: the lower the investment in the lottery (variable GP_Invest), the stronger the risk aversion. Model 2, in turn, accounts for individual confidence in own beliefs ($1[Conf10]$ and $1[Conf15]$ are indicator variables for the invervals $\pm 10\%$ and $\pm 15\%$, respectively; $\pm 5\%$ is the omitted reference interval). Finally, Model 3 combines the two sets of explanatory variables.

The coefficients of Model 1 are interpreted as follows. γ_0 is the risk premium for non-risk-averse subjects. Across the three treatment-specific regressions, this coefficient is close to zero and insignificant; a joint test of insignificance yields $p = 0.678$. This means that non-risk-averse subjects do not ask for a risk premium and their WTA corresponds to the belief-based expected payoff from the prediction task. For the risk averse subjects ($1[RA] = 1$), in turn, we estimate the risk premium as a linear function of their degree of risk aversion (captured by GP_Invest):

[22] The SUR specification (see Cameron and Trivedi, 2005, pp. 209-210) is due to the within-subject design of Experiment 2: for a given model, the residuals coming from the same individual may be correlated across treatments.

[23] As mentioned in footnote 12, the Gneezy-Potters task does not allow us to check whether those who invested their whole endowment (42 subjects) are risk neutral or risk seeking. We categorize them as non-risk-averse ($1[RA] = 0$).

Table 4: Risk premium, risk aversion and confidence: SUR analysis

	Model 1			Model 2			Model 3		
	VIDLO	VIDNE	PHOTO	VIDLO	VIDNE	PHOTO	VIDLO	VIDNE	PHOTO
Intercept (γ_0)	0.876	-1.762	-3.114	6.707***	5.936	0.148	3.246	-0.454	-5.311
	(3.377)	(3.747)	(3.597)	(2.342)	(3.613)	(3.000)	(3.660)	(4.440)	(4.115)
$1[RA]$ (γ_1)	10.825*	26.302***	16.478**				11.064*	26.645***	16.324**
	(6.321)	(7.013)	(6.732)				(6.313)	(7.047)	(6.624)
$1[RA] \times GP\_Invest$ (γ_2)	-0.144	-0.389***	-0.197				-0.146	-0.390***	-0.198
	(0.116)	(0.129)	(0.124)				(0.116)	(0.129)	(0.122)
$1[Conf10]$ (γ_3)				-5.173*	-1.642	3.969	-5.254*	-2.788	3.683
				(3.056)	(4.000)	(3.360)	(3.056)	(3.949)	(3.345)
$1[Conf15]$ (γ_4)				-3.158	2.968	1.729	-3.400	1.606	1.284
				(7.607)	(5.740)	(5.550)	(7.639)	(5.635)	(5.531)
R^2	0.020	0.089	0.041	0.004	0.002	0.020	0.024	0.091	0.058
$Prob > \chi^2$	0.229	<0.001	0.045	0.238	0.648	0.489	0.203	0.004	0.103

Note. $N = 145$. ***/**/* indicate statistical significance at the 1%/5%/10% level. $1[RA]$ is an indicator for risk averse subjects (103 in our sample). $1[RA] \times GP\_Invest$ captures the investment made by those subjects. $1[Conf10]$ ($1[Conf15]$) is an indicator of the interval of confidence $\pm 10\%$ ($\pm 15\%$).

$E(RP_X^i | 1[RA] = 1, GP\_Invest) = \gamma_0 + \gamma_1 + \gamma_2 \times GP\_Invest$. The slope coefficient γ_2 captures the monotone pattern in which the risk premium in the prediction varies with the degree of risk aversion. Its negative sign suggests that the risk premium increases with the degree of risk aversion: the lower the investment in the Gneezy-Potters lottery task, the higher the estimated risk premium in the prediction task.[24] In particular, the risk premium for by the most risk averse agents (i.e., those who invest zero) is estimated as $\gamma_0 + \gamma_1$ and found to be positive and significant in all the treatments ($p_{VIDLO} = 0.029$; $p_{VIDNE} < 0.001$; $p_{PHOTO} = 0.019$). It takes the highest value in VIDNE, and the lowest value in VIDLO; a joint test rejects the null hypothesis of the equality of the three risk premia ($p = 0.049$). At the opposite extreme, the risk premium for the least risk averse subjects (i.e., the who came close to investing their whole endowment – 99 out of 100) is estimated as $\gamma_0 + \gamma_1 + \gamma_2 \times 99$ and insignificant across treatments ($p_{VIDLO} = 0.712$; $p_{VIDNE} = 0.071$; $p_{PHOTO} = 0.406$; joint nullity: $p = 0.312$). We do not reject the null hypothesis that the risk premium demanded by those subjects is the same as the one of non-risk-averse ones ($p_{VIDLO} = 0.656$; $p_{VIDNE} = 0.156$; $p_{PHOTO} = 0.712$; joint test: $p = 0.506$).[25]

The coefficients of Model 2, in turn, are found to be jointly insignificant across all the treatments, pointing to a lack of statistical relationship between individual confidence in beliefs and risk

[24] Although γ_2 is only statistically significant in VIDNE, we cannot reject the null hypothesis of equality of γ_2 across the three conditions ($p = 0.119$), but we do reject their joint nullity ($p = 0.027$).

[25] This exercise involves testing $H_0 : \gamma_1 + \gamma_2 \times 99 = 0$.

premium. Finally, Model 3 fully confirms the outcomes of Model 1 while accounting for individual confidence.

Result 4 (Subjective valuation and risk aversion) *High degree of risk aversion is associated with a positive risk premium: the elicited WTA is significantly lower than the belief-based expected payoff from making a prediction. The risk premium fades away as the degree of risk aversion falls.*

Altogether, Hypothesis 3b is supported by our data.

4 Conclusion

Eckel and Petrie (2011) show that many people exhibit a preference, and thus generate demand, for observable information (e.g., the counterpart's face) in strategic interactions. Following a growing body of experimental research in social sciences (see Bonnefon et al., 2017, for a survey), one way to rationalize this phenomenon is to claim that such a preference exists because observable information may be helpful in predicting human behavior (e.g., detecting the counterpart's cooperativeness). However, the existing empirical evidence seems to contradict this intuition. Recent work by Bonnefon et al. (2013) and Tsay (2013) suggests that the preferences for observable information in prediction experiments are far from being rational (or "appropriate" in the parlance of Bonnefon et al., 2013) in that the most useful sources of information happen to be the least preferred ones.

In this study, we challenge the view that people exhibit irrational preferences for observable information when it comes to detecting other people's trustworthiness. Both our methodology and main result differ from the previous experimental test of preferences for observable information. Unlike these studies, we use a methodology allowing us to highlight the importance of information (in both objective and subjective terms) for the ability to predict human behavior, while muting any factors that are unrelated to the prediction task (e.g., mere curiosity, or more generally the hedonic value of receiving such information about other people). In this refined setting, we document a general consistency between the main experimental outcomes: the actual value of information in the prediction task, the stated preferences for information, and the elicited beliefs about the actual value of information. Thus, the main part of our experimental findings suggests that people do not value observable information in strategic interactions as irrationally as previously claimed. However, we also acknowledge the existence of some biases in participants' valuation of observable information in our experiment. First, subjects place value (albeit small) on the two sources of information that do not help identify the target player's type (i.e., neutral photo and neutral video).[26] Second, their perception of the most valuable source of information – loaded video – is biased: subjects' beliefs about its value are substantially exaggerated relative to its objective value.

Previous experiments (Fetchenhauer et al., 2010; Bonnefon et al., 2013; Vogt et al., 2013; De Neys et al., 2015; Van Leeuwen et al., 2018) suggest that physical cues (neutral facial expressions, whether they are static or dynamic) may be helpful for predicting economic behavior when individual decisions are made in isolation. Notwithstanding those studies, the target behavior in our experimental prediction task comes from a two-stage game in which the predicted action (i.e., the trustees' decision in the hidden action game) is preceded by a pre-play social interaction phase (the trustees' face-to-face communication with the trustors). Our data suggest that once the social vacuum of the decision-making environment is lifted, the information about the physical attributes of the target player (such as a mugshot pictures of a neutral facial expressions, or a standardized video recording with non-strategic auditory content) are insufficient for distinguishing between trustworthy and untrustworthy trustees. Only observing the trustees' behavior (for instance, whether they made a promise to *Roll* or not) in front of the trustors during the pre-play communication stage improves prediction accuracy. Thus, the social cues override the physical ones. Importantly, understanding social cues requires the knowledge of a common cultural code for interpreting not only verbal, but also non-verbal content. In our setting, the cultural proximity of experimental samples (drawn from student populations in major French cities) naturally satisfies this requirement. However, recent evidence by Tognetti et al. (2018) suggests that increasing the cultural distance between the interaction parties may affect the predictive value of non-verbal

[26] As suggested by a referee, this behavior may be related to another form of curiosity: the participants may be willing to pay a moderate fee for the possibility to test their own ability to "read" human faces. We acknowledge that our design is not suitable for testing this explanation.

content, making it more useful within cultural proximity than across cultures. A natural question to ask is whether the rationality of preferences for observable information would persist in an inter-cultural setting. This robustness test is next on our agenda.

References

Babutsidze, Z., N. Hanaki, and A. Zylbersztejn (2019): "Digital Communication and Swift Trust," Working Papers 1909, Groupe d'Analyse et de Théorie Economique Lyon St-Étienne (GATE Lyon St-Étienne), Université de Lyon.

Baron-Cohen, S., S. Wheelwright, J. Hill, Y. Raste, and I. Plumb (2001): "The 'Reading the mind in the eyes' test revised version: A study with normal adults, and adults with Asperger Syndrome or High-Functioning autism," Journal of Child Psychology and Psychiatry, 42, 241–252.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964): "Measuring utility by a single-response sequential method," Behavioral Science, 9, 226–232.

Belot, M., V. Bhaskar, and J. Van De Ven (2012): "Can observers predict trustworthiness?" Review of Economics and Statistics, 94, 246–259.

Bonnefon, J.-F., A. Hopfensitz, and W. De Neys (2013): "The modular nature of trustworthiness detection." Journal of Experimental Psychology: General, 142, 143.

——— (2017): "Can we detect cooperators by looking at their face?" Current Directions in Psychological Science, 26, 276–281.

Camerer, C. F. (2003): Behavioral Game Theory: Experiments in Strategic Interaction, Princeton University Press.

Cameron, A. C. and P. K. Trivedi (2005): Microeconometrics: Methods and Applications, Cambridge University Press.

Centorrino, S., E. Djemai, A. Hopfensitz, M. Milinski, and P. Seabright (2015): "Honest signaling in trust interactions: Smiles rated as gesnuine induce trust and signal higher earning opportunities," Evolution and Human Behavior, 36, 8–16.

Charness, G. and M. Dufwenberg (2006): "Promises and Partnership," Econometrica, 74, 1579–1601.

De Neys, W., A. Hopfensitz, and J.-F. Bonnefon (2015): "Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults," Journal of Economic Psychology, 47, 17–22.

——— (2017): "Split-second trustworthiness detection from faces in an economic game," Experimental Psychology.

Eckel, C. C. and R. Petrie (2011): "Face value," American Economic Review, 101, 1497–1513.

Epps, T. and K. J. Singleton (1986): "An omnibus test for the two-sample problem using the empirical characteristic function," Journal of Statistical Computation and Simulation, 26, 177–203.

Fetchenhauer, D., T. Groothuis, and J. Pradel (2010): "Not only states but traits–Humans can identify permanent altruistic dispositions in 20 s," Evolution and Human Behavior, 31, 80–86.

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

Fischbacher, U., S. Gächter, and E. Fehr (2001): "Are people conditionally cooperative? Evidence from a public goods experiment," Economics Letters, 71, 397–404.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994): "Fairness in simple bargaining experiments," Games and Economic Behavior, 6, 347–369.

Frederick, S. (2005): "Cognitive reflection and decision making," Journal of Economic Perspectives, 19, 25–42.

Gneezy, U. and J. Potters (1997): "An Experiment on Risk Taking and Evaluation Periods," Quarterly Journal of Economics, 112, 631–645.

Goerg, S. J. and J. Kaiser (2009): "Nonparametric testing of distributions–The Epps-Singleton two-sample test using the empirical characteristic function," Stata Journal, 9, 454.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," Journal of the Economic Science Association, 1, 114–125.

Ismayilov, H. and J. Potters (2016): "Why do promises affect trustworthiness, or do they?" Experimental Economics, 19, 382–393.

Jacquemet, N. and O. L'Haridon (2019): Experimental Economics: Methods and Applications, Cambridge University Press.

Kocher, M. G., T. Cherry, S. Kroll, R. J. Netzer, and M. Sutter (2008): "Conditional cooperation on three continents," Economics Letters, 101, 175–178.

Murphy, R. O., K. A. Ackermann, and M. J. Handgraaf (2011): "Measuring social value orientation." Judgment and Decision Making, 6, 771–781.

Schlag, K. H. and J. J. van der Weele (2015): "A method to elicit beliefs as most likely intervals," Judgment and Decision Making, 10, 456–469.

Schwartz, S., E. Spires, and R. Young (2018): "Why do people keep their promises? A further investigation," Experimental Economics, 1–22.

Sylwester, K., M. Lyons, C. Buchanan, D. Nettle, and G. Roberts (2012): "The role of theory of mind in assessing cooperative intentions," Personality and Individual Differences, 52, 113–117.

Todorov, A., F. Funk, and C. Olivola (2015): "Response to Bonnefon et al.: Limited'kernels of truth'in facial inferences." Trends in Cognitive Sciences, 19, 422.

TOGNETTI, A., N. YAMAGATA-NAKASHIMA, C. FAURIE, AND R. ODA (2018): "Are non-verbal facial cues of altruism cross-culturally readable?" Personality and Individual Differences, 127, 139 – 143.

TSAY, C.-J. (2013): "Sight over sound in the judgment of music performance," Proceedings of the National Academy of Sciences, 110, 14580–14585.

VAN LEEUWEN, B., C. N. NOUSSAIR, T. OFFERMAN, S. SUETENS, M. VAN VEELEN, AND J. VAN DE VEN (2018): "Predictably angry – facial cues provide a credible signal of destructive behavior," Management Science, 64, 3352–3364.

VERPLAETSE, J., S. VANNESTE, AND J. BRAECKMAN (2007): "You can judge a book by its cover: the sequel.: A kernel of truth in predictive cheating detection," Evolution and Human Behavior, 28, 260–271.

VOGT, S., C. EFFERSON, AND E. FEHR (2013): "Can we see inside? Predicting strategic behavior given limited information," Evolution and Human Behavior, 34, 258–264.

A Implementation and instruction of the hidden action game experiment

This appendix provides details of the implementation as well as the instructions used in the hidden action game experiment of Babutsidze et al. (2019).

A.1 Implementation

Each experimental session involves 6 trustors and 6 trustees. All trustors remain in one room during the whole experiment. They are seated in a single row, isolated one from another by separators, and not allowed to talk. The space in front of them is left open and used by a trustee to make a brief statement. Trustees enter the room one by one, so that trustors play six rounds of the game (which is common knowledge). Each time, trustee faces the center of trustors' row, and all trustors have a clear view on the speaker. Trustee also has a clear, unobstructed view on all six trustors. After making a statement, trustee is invited to a separate room where s/he privately decides whether to *Roll* a die or not. Then, s/he is asked to leave the laboratory and wait outside until the end of the experiment. At the same time, each trustor makes a decision whether to go *In* or stay *Out*. All decisions are made on a sheet of paper, which is then put in an envelope, sealed, and collected by the laboratory staff after each round. In addition, once trustee has made a decision and left the separate room, a laboratory staff member rolls a die in private and marks the outcome on trustee's sealed envelope. At the end of the experiment, trustors and Bs are randomly and anonymously matched in pairs. The outcome of the game for each pair is based on the payoff structure described in Figure 1 and defined by the decision made by trustor after trustee's statement, as well as the decision made by trustee in a private room had the trustor chosen to go *In*. For the trustee's decision to *Roll*, the outcome of the die roll is also taken into account.

For the sake of logistics and efficient time management, trustees arrive 30 minutes prior to trustors. First, they are asked to take up several computerized tasks that measure their preferences and characteristics. Then, they are all led to a waiting room. To avoid any communication or subjects overhearing what others are saying or doing, each participant is seated in a separate cubicle, puts on a headphone and listens to a classical music until further notice. Then, they are taken one by one to a separate room for a mugshot picture and a short, standardized video recording.[27]

Then, each subject is seated back in his cubicle with headphones on. He now listens to an audio file containing the experimental instructions (paper version is also provided). There is a brief comprehension quiz assisted by a laboratory staff member. Finally, he receives additional paper instructions about the upcoming statement in front of trustors, as well as a pen and an empty sheet of paper, and is given approximately two minutes to prepare his message.[28] After that, a trustee is invited to trustors' room where he delivers a statement, leaves for another room, and the game proceeds as explain above in Section 2. The average duration of a message is 26.39 seconds (SD 2.09). Trustees' statements are recorded using a small, non-intrusive video camera set up in the middle of trustors' row, right in front or trustees' zone, so that the perspective in the video camera recording resembles the one of a trustor. The camera is always adjusted to the height of trustee (so as to capture head, shoulders, and thorax), and to the luminosity in the room. The sake of the quality of the video recordings, the background in trustees' zone is covered with light canvas. While making a statement, each trustee also has a portable microphone attached below their face. The distance between trustors and a trustee is set to 2.50 meters.

Upon their arrival to the laboratory, trustors also take up the set of preliminary questionnaires. Then, they receive and read paper instructions for the experimental game, and finally they fill in a short comprehension quiz. A laboratory staff member then reads aloud all the questions from the quiz along with the corrects answers, and answers any remaining questions. Finally, trustors are asked to wait for the arrival of the first trustee.

We have conducted 7 sessions. However, one trustee in session 6 decided to quit after the

[27]Like in Van Leeuwen et al. (2018), subjects are asked to read neutral content (a short extract from a printer instruction manual) and keep a neutral face expression. The recording takes about 30 seconds. This information is not part of the present investigation and is not reported in the paper.

[28]Those additional instructions remind the subject about his role in the game; emphasize the fact that the message may affect trustors' decisions and, consequently, the subject's gain from the experiment; instruct the subject to avoid making a visual or verbal contact with the experimenter, to aim at communicating with all trustors, and not to introduce oneself or give any details about one's own identity.

preliminary measurements and before receiving the instruction of the hidden action game, and was replaced by a research assistant unknown to trustors. To avoid any contamination of trustors' behavior, that research assistant acted as trustee in the final round of the experimental game. The data from that round were dismissed and our dataset from that session only covers 5 trustees, and thus 41 trustees in total.

A.2 Instructions

A.2.1 Preliminary instructions given to all subjects

You are about to take part in an experiment in which you can earn money. The amount of your gains will depend on your decisions, as well as on the decisions made by other participants. In addition, you will receive a fixed fee of [*5 for player As, 10 for player Bs*] EUR for completing the experiment. Your total earnings will be paid privately in cash at the end of the experiment.

The experiment consists of several parts. Each part will involve tasks the rules of which will be explained to you in due time. It is crucial that you understand and obey the rules of this experiment. Violation of these rules might result in an exclusion from the experiment and all payments. Please raise your hand whenever you have questions or need assistance.

All the information you provide, as well as the amount of your gains from this experiment, will remain strictly confidential and anonymous.

We would now like to ask you to answer a series of preliminary questions. You will answer these questions using the interface on your computer screen. Some of these questions will generate monetary gains. These gains will be determined and added to your overall earnings at the end of the experiment.

Note: the following instructions were given only to player Bs in F2F for the preliminary recordings.

Now, we would like to take a picture and video recoding of you.

First, you will be asked to stand by the wall and look into the camera. Please, try to keep a neutral facial expression.

Second, you will be asked to read aloud the content display on the screen in front of you. While reading, you will be video recorded.

All pictures and video recordings produced during this experiment will only serve strictly scientific purposes of this research project. They may be used in other experimental sessions related to this research project.

A.2.2 Instructions for the hidden action game

Rules of the game

You will now play a game with monetary stakes. The rules of the game are as follows.

The game is played by two players: player A and player B. Each player must choose between two possible actions. Player A chooses between actions "Left" and "Right". Player B chooses whether she want a six-sided die to be rolled (action "Roll") or not (action "Don't roll").

You will play the role of player [*A for player As, B for player Bs*]

Each players' payoff depends on the actions chosen by herself as well as the other player:

- if player A chooses "Left", then regardless of player Bs' choice:
  - player A's payoff is 5 EUR and player B's payoff is 5 EUR;
- if player A chooses "Right" and player B chooses "Don't roll":

- player A's payoff is 0 EUR and player B's payoff is 14 EUR;

- if player A chooses "Right" and player B chooses "Roll":
  - if the number of on the die is between 1 and 5, then player A's payoff is 12 EUR and player B's payoff is 10 EUR;
  - if the number of on the die is 6, then player A's payoff is 0 EUR and player B's payoff is 10 EUR;

How the game proceeds

The game will consist of six identical rounds.

At the beginning of a round, one player B is asked to enter the room in which there are six players As. Player As are separated one from another and are not allowed to talk.

Player B is then placed in front of player As and remains silent. Then, player B is allowed to talk for no longer than 20 seconds, and then asked to leave player As' room. While talking, player B is video recorded and should look straight into the camera.

Once player B leaves player As' room:

- player B makes a decision in a separate room. Player B privately and individually indicates her decision (either "Roll" or "Don't roll") on a separate answer sheet, puts it in an envelope and seals the envelope. The experimenter collects the envelope and player B leaves the room. Then, the experimenter privately rolls a six-sided die and marks the result on the envelope (without opening it). The outcome of the die roll will only be taken into account if player A's decision is "Right" and player B's decision is "Roll".

- each player A privately and individually indicates her decision (either "Left" or "Right") on a separate answer sheet, puts it in an envelope and seals the envelope. Then, all the envelopes are collected by the experimenter. Player As are either asked to remain silent and await the next player B, or informed that the experiment is over and given further instructions about their payment.

No envelope will be opened before the end of the experiment.

At the end of the experiment, each player A is anonymously and randomly matched with one player B. The outcome of the game for each pair of players is determined by the decisions made by both playes (and also by the outcome of the die roll if the decisions in a pair are "Right" and "Roll") in the round in which the player B was in player As' room. Players are only informed about their personal payoffs, and not about the payoffs of or the decisions made by other players, or about the outcome of the die roll.

Additional information

Note that this set of instructions is provided to and read by each player A and each player B. Furthermore, player Bs cannot communicate between themselves at any point of the experiment. The same applies to the communication between player As.

You will play the role of player [*A for player As, B for player Bs*]

B Experimental instructions

This appendix provides instruction used in Experiment 1 and 2. Part B.1 provides the preliminary instructions that are common for all the treatments and experiments. Parts B.2 and refApp23 provide instruction specific to Experiment 1 and 2, respectively.

B.1 Preliminary instructions given to all subjects in all the treatments

You are about to take part in an experiment in which you can earn money. The amount of your gains will depend on your decisions, as well as on the decisions made by other participants. In addition, you will receive a fixed fee of 5 EUR for completing the experiment. Your total earnings will be paid privately in cash at the end of the experiment.

The experiment consists of several parts. Each part will involve tasks the rules of which will be explained to you in due time. It is crucial that you understand and obey the rules of this experiment. Violation of these rules might result in an exclusion from the experiment and all payments. Please raise your hand whenever you have questions or need assistance.

All the information you provide, as well as the amount of your gains from this experiment, will remain strictly confidential and anonymous.

We would now like to ask you to answer a series of preliminary questions. You will answer these questions using the interface on your computer screen. Some of these questions will generate monetary gains. These gains will be determined and added to your overall earnings at the end of the experiment.

B.2 Specific instructions for Experiment 1

Note: Below, the parts of instructions that are distinct for each treatment are marked with "(treatment's name:)". Other parts are common to all three treatments.

(PHOTO:) In this part of the experiment, you will see a series of pictures of people.

(VIDLO and VIDNE:) In this part of the experiment, you will watch a series of video recordings. In each recording, you will see a person making a short statement.

You will be asked to predict the decisions those people previously made in another experiment (the details of which are described below). Your final gain will depend on the accuracy of your predictions.

The previous experiment

In each session, two groups of participants (six players A and six players B) were installed in two different rooms. Participants in each room could not communicate with each other. They all received instructions explaining the rules of the experiment they were about to participate in. Players were informed that their decisions and earnings would remain private and anonymous, and would never be disclosed to other participants.

Each session was organized as follows:

1. One by one, player Bs entered the room in which players A were sitting. Then, each player B made a short speech in front of player As. Before entering the room, each player B was give a couple of minutes to prepare the statement. Each player B was also informed that his statement could affect player As' decisions and, consequently, his own gain in the experiment. (VIDLO:) **All the statements have been recorded, and you will be watching some of them**.

2. After his speech, player B left player As' room, and entered an empty room.

3. After player B's departure, each player A made a decision ("Left" or"Right") in private and individually. At the same time, player B made a decision ("Roll" or "Don't roll" a die) in privately and individually.

4. Thereafter, player B left the room and waited outside the laboratory until the end of the experiment. Meanwhile, a new player B was entering the players' room A to make a speech. The experiment ended when all the players had completed their task.

At the end of the experiment, each player A was anonymously and randomly matched with a player B. The outcome of the game for each pair of players was determined by the decisions made by both players following player B's speech:

- if player A chose "Left", then regardless of player Bs' choice:
  - player A's payoff was 5 EUR and player B's payoff was 5 EUR;
- if player A chose "Right" and player B chose "Don't roll":
  - player A's payoff was 0 EUR and player B's payoff was 14 EUR;
- if player A chose "Right" and player B chose "Roll":
  - if the number of on the die was between 1 and 5, then player A's payoff was 12 EUR and player B's payoff was 10 EUR;
  - if the number of on the die was 6, then player A's payoff was 0 EUR and player B's payoff was 10 EUR;

Your role

(PHOTO:)

This experiment consists of **20 rounds**. At the beginning of each round, you **will see a picture.**

Each picture presents a person in the role of player B from the previous experiment, as described above. The picture was taken privately and independently of the previous experiment.

Then, you will be asked to predict if the player B from the picture decided to roll a die in the previous experiment. Your gain will depend on the accuracy of your prediction: you will earn 10 euros for correct prediction and 2 euros for an incorrect one.

(VIDNE:)

This experiment consists of **20 rounds**. At the beginning of each round, you **will watch a short video recording (with the sound on).**

Each recording presents a person in the role of player B from the previous experiment, as described above. The recording was made privately and independently of the previous experiment.

Then, you will be asked to predict if the player B from the picture decided to roll a die in the previous experiment. Your gain will depend on the accuracy of your prediction: you will earn 10 euros for correct prediction and 2 euros for an incorrect one.

(VIDLO:)

This experiment consists of **20 rounds**. At the beginning of each round, you **will watch a short video recording (with the sound on).**

Each recording presents the statement made by a player B in front of player As during the previous experiment, as described above.

Then, you will be asked to predict if the player B from the recording decided to roll a die in the previous experiment. Your gain will depend on the accuracy of your prediction: you will earn 10 euros for a correct prediction and 2 euros for an incorrect one.

At the end of the experiment, two rounds will be drawn at random. Your final gain will correspond to the predictions you have made in those two rounds.

B.3 Specific instructions for Experiment 2

In this part of the experiment, you will be asked to predict the decisions those people previously made in another experiment (the details of which are described below). Your final gain will depend on the accuracy of your predictions.

The previous experiment

In each session, two groups of participants (six players A and six players B) were installed in two different rooms. Participants in each room could not communicate with each other. They all received instructions explaining the rules of the experiment they were about to participate in. Players were informed that their decisions and earnings would remain private and anonymous, and would never be disclosed to other participants.

Each session was organized as follows:

1. One by one, player Bs entered the room in which players A were sitting. Then, each player B made a short speech in front of player As. Before entering the room, each player B was give a couple of minutes to prepare the statement. Each player B was also informed that his statement could affect player As' decisions and, consequently, his own gain in the experiment.

2. After his speech, player B left player As' room, and entered an empty room.

3. After player B's departure, each player A made a decision ("Left" or"Right") in private and individually. At the same time, player B made a decision ("Roll" or "Don't roll" a die) in privately and individually.

4. Thereafter, player B left the room and waited outside the laboratory until the end of the experiment. Meanwhile, a new player B was entering the players' room A to make a speech. The experiment ended when all the players had completed their task.

At the end of the experiment, each player A was anonymously and randomly matched with a player B. The outcome of the game for each pair of players was determined by the decisions made by both players following player B's speech:

- if player A chose "Left", then regardless of player Bs' choice:
  - player A's payoff was 5 EUR and player B's payoff was 5 EUR;
- if player A chose "Right" and player B chose "Don't roll":
  - player A's payoff was 0 EUR and player B's payoff was 14 EUR;
- if player A chose "Right" and player B chose "Roll":
  - if the number of on the die was between 1 and 5, then player A's payoff was 12 EUR and player B's payoff was 10 EUR;
  - if the number of on the die was 6, then player A's payoff was 0 EUR and player B's payoff was 10 EUR;

Your role

This stage of the experiment consists of several parts. Below are the instructions for the first part. Instructions for the remaining parts will be provided later.

Part 1

First, computer software will **randomly choose a player B who participated in the previous experiment**, as described above. Your task is to predict if that player decided to roll a die or not in the previous experiment. Your gain will depend on the accuracy of your prediction:

- you will earn 100 points for a correct prediction
- you will earn 20 points for an incorrect prediction.

1 point is worth 0.10 euro.

Your prediction may be made in **one of the four contexts**:

- prior to making a prediction, you will be given the opportunity to see a picture showing the player B. The picture was taken privately and independently of the previous experiment;
- prior to making a prediction, you will be given the opportunity to watch a short video recording (with the sound on) showing the player B. The recording was made privately and independently of the previous experiment. It presentes the player reading a paragraph of a printer instruction manual;
- prior to making a prediction, you will be given the opportunity to watch a short video recording (with the sound on) showing the player B while making a statement in front of player As in the previous experiment, as described above;
- before making the prediction, you will not receive any information about the player B from the previous experiment.

For each of these contexts, you will be first asked to indicate an amount (denoted M) between 21 and 100 points. This amount should make you indifferent between receiving M for sure (and without the need to predict player B's choice) and making a prediction about player B's choice which, depending on the accuracy of your prediction, may earn you either 100 points (for a correct prediction) or 20 points (for an incorrect prediction).

Then, **one of the contexts will be chosen at random** by the computer software. For this particular context, the computer will then randomly generate a number between 20 and 99 (denoted N), all values within this range being equally probable.

If $M \leq N$ (that is, if the amount you have chosen for the randomly selected context is smaller than or equal to the random number N generated by the computer), then you will receive a **certain payoff of** $N \times 0.10$ **euro and will not make a prediction** in the context randomly selected by the computer. **If this context contains a video or a photo of the player B, you will still have the opportunity to inspect its content.**

If $M > N$ (that is, if the amount you have chosen for the randomly selected context is greater than the random number N generated by the computer), then you will make a prediction in the context randomly selected by the computer and your gain will depend on the accuracy of your prediction: **you will earn 100 points (10 euros) for a correct prediction** or **20 points (2 euros) for an incorrect prediction**.

Here are some examples illustrating this procedure:

Example 1. If the amount M you chose for the context randomly selected by the computer equals 32, and the randomly generated number N equals 26, then you will make a prediction in the context selected by the computer.

Example 2. If the amount M you chose for the context randomly selected by the computer equals 37, and the randomly generated number N equals 76, then your gain will be equal to 7.60 euros (76×0.10 euro) and you will not be asked to make a prediction in the context selected by the computer.

Example 3. If the amount M you chose for the context randomly selected by the computer equals 81, and the randomly generated number N equals 81, then your gain will be equal to 8.10 euros (81×0.10 euro) and you will not be asked to make a prediction in the context selected by the computer.

Part 2

Before the random draw, we would like to ask you to perform an extra task allowing you to earn money. Its rules are outlined below.

Previously, we have conducted **additional experimental sessions** in which participants were asked to guess the decisions made by player Bs in the experimental game. Each **participant in those previous experiments made multiple predictions in one (and only one) of the three contexts** you have seen before:

- either the following context: *prior to making a prediction, they were given the opportunity to watch a short video recording (with the sound on) showing the player B while making a statement in front of player As in the previous experiment, as described above*;
- or the following context: *prior to making a prediction, they were given the opportunity to watch a short video recording (with the sound on) showing the player B. The recording was made privately and independently of the previous experiment. It presentes the player reading a paragraph of a printer instruction manual*;
- or the following context: *prior to making a prediction, they were given the opportunity to see a picture showing the player B. The picture was taken privately and independently of the previous experiment*;

For each of those **contexts, we have calculated the average rate of correct predictions**, i.e. **how many predictions in 100 were correct on average**.

For each of the three contexts, you will be asked to guess that average rate of correct predictions. You will choose a number between 0 and 100, as well as ONE of the following intervals: ± 5, ± 10, ± 15.

Your answer is correct whenever the difference between the average rate of correct predictions observed in the previous experiment and the number you have stated lies within the interval you have chosen. For interval ± 5, this difference cannot be greater than 5 in either direction. For intervals ± 10 or ± 15, this difference cannot exceed 10 or 15 (respectively) in either direction.

In order to earn money, **you need to provide corrects answers in all three contexts**. The amount you will earn **may also vary depending on the chosen intervals**.

You earn 10 euros by providing three corrects answers with intervals ± 5. Each time you choose an interval ± 10, this amount will decrease by 1.50 euros. If you choose ± 15, instead, it will decrease by 3 euros.

Example 1. Suppose a person provided three correct answers: two with interval ± 5 and one with interval ± 15. The corresponding gain equals $10 - 3 = 7$ euros.

Example 2. Suppose a person provided three correct answers: one with interval ± 5 and two with interval ± 10. The corresponding gain equals $10 - 1.50 - 1.50 = 7$ euros.

Example 3. Suppose a person provided three correct answers: two with interval ± 15 and one with interval ± 5. The corresponding gain equals $10 - 3 - 3 - 1.50 = 2.50$ euros.

C Prediction accuracy and promise-making in VIDLO: logistic regression analysis

Table 5: Predictions and promises: logistic regression

	coeff. (SE)	p	AME (SE)	p
Intercept	0.072 (0.087)	0.409	—	
$1[Promise]$	0.286 (0.071)	0.046	0.070 (0.035)	0.045

Note. Results of a logistic regression of the accurate prediction indicator variable on the promise-making indicator variable with individual-level standard error clustering. The left-hand side (right-hand side) columns provide coefficients (average marginal effect) along with p-values corresponding to z-test of statistical significance. $N = 896$ in 43 clusters, $pseudo - R^2 = 0.004$.

D Robustness check: Bayes factor analysis

Tables 6 and 7 provide an alternative way of performing two-sample and paired t-tests related to our Hypotheses 1-3 and reported in Section 3. Instead of standard hypothesis testing, we now turn to Bayesian hypothesis testing based on Bayes factor (BF_{10}) that gives information about the relative likelihood of H_0 compared to H_1. BF_{10} can thus be interpreted in terms of evidence against H_0: higher BF_{10} means stronger the evidence against H_0. The estimated Bayes factors are based on the default Cauchy prior with scale parameter $r = 0.707$ and obtained using the JASP software (version 0.11.1, available at *https://jasp-stats.org/*). The interpretation of the results provided in both tables is commonplace in the experimental literature and can be found in Jacquemet and L'Haridon (2019), *inter alia*. Altogether, this analysis complements, and fully stand in line with, the conclusions drawn from the significance testing based on the respective p-values reported in Section 3.

Table 6: Experiment 1: Bayes factors from two-sample t-test

Comparison	BF_{10}	Interpretation
Accuracy rate: VIDLO vs. VIDNE	13.826	Substantial evidence against H_0
Accuracy rate: VIDLO vs. PHOTO	18.942	Substantial evidence against H_0
Accuracy rate: VIDNE vs. PHOTO	0.225	Negative evidence (H_0 supported)

Table 7: Experiment 2: Bayes factors from paired t-test

Comparison	BF_{10}	Interpretation
WTA_{VIDLO} vs. WTA_{NOINF}	1.329e+10	Decisive evidence against H_0
WTA_{VIDNE} vs. WTA_{NOINF}	7.482e+19	Decisive evidence against H_0
WTA_{PHOTO} vs. WTA_{NOINF}	8879.751	Decisive evidence against H_0
ΔWTA_{VIDLO} vs. ΔWTA_{VIDNE}	1.829e+8	Decisive evidence against H_0
ΔWTA_{VIDLO} vs. ΔWTA_{PHOTO}	3.482e+10	Decisive evidence against H_0
ΔWTA_{VIDNE} vs. ΔWTA_{PHOTO}	0.265	Negative evidence (H_0 supported)
Belief: VIDLO vs. VIDNE	676.852	Decisive evidence against H_0
Belief: VIDLO vs. PHOTO	1.188e+7	Decisive evidence against H_0
Belief: VIDNE vs. PHOTO	2.395	Evidence against H_0 barely worth mentioning
Interval: VIDLO vs. VIDNE	1.154e+6	Decisive evidence against H_0
Interval: VIDLO vs. PHOTO	693.584	Decisive evidence against H_0
Interval: VIDNE vs. PHOTO	0.308	Negative evidence (H_0 supported)