EXPERIMENTS ON EMOTIONS, MATCHING, AND DECISION-MAKING

Elisabetta Leni

A thesis submitted for the degree of Doctor of Philosophy

Department Economics

University of Essex

December 2020

Abstract

This Doctoral thesis includes three articles. In *Heterogeneity of Incentive Effects on Elicited Guilt Aversion*, I study guilt intended as the feeling that arises when we think we disappoint others' expectations. I investigate how the anticipated cost of guilt is traded-off with financial incentives in the investment game. I also explore the characteristics of guilt averse participants with respect to personality traits. In *Matching and Confidence*, I study how matching affects confidence. In collaboration with Friederike Mengel (University of Essex and Lund University) and Mehmet Yigit Gurdal (Bogazici University), we use an experiment that allows us to identify the effect of being matched with others of either similar or dissimilar performance (assortative or disassortative matching) on people's confidence in their own ability. In *Incidental Affects, Belief Management, and Decision-Making*, I use a lab experiment to study how incidental affects - immediate emotions unrelated to the judgement at hand - influence decisions and beliefs in two classical choice situations: a dictator game and an effort task.

Acknowledgements

I would like to thank my supervisor, Friederike, for providing guidance and encouragement throughout the doctoral degree. I express sincere gratitude to the EssexLab for the incredible support given to my experiments. I thank my colleagues, Ludovica, Andrea, and Xiao: our meetings and conversation have been vital to keep up the good work... and the moral! Finally, a special thanks to my partner, Vanni, for the encouragement and patience in putting up with my stresses in this final year...

Contents

Heterogeneity of Incentive Effects on Elicited Guilt Aversion

Elisabetta Leni*
University of Essex

Abstract

An increasing amount of empirical studies suggests that in many strategic interactions, people's preferences depend on guilt aversion. This paper explores the presence of heterogeneity in the way people experience guilt under different incentive schemes and studies the influence of personality traits on elicited guilt aversion. We use an experiment where participants play three rounds of the investment game by Berg, Dickhaut and McCabe (1995) with different rates of return. At the end, participants complete a Big-Five questionnaire. The results reveal significant heterogeneity in the way people trade-off guilt and monetary incentives with a prevalence of players for which guilt aversion - measured as the correlation between second order beliefs and choices - decreases as the stakes in the game become higher. The analysis of the personality data suggests that guilt aversion is negatively correlated to openness to experience and positively correlated with neuroticism. All the results are robust upon controlling for standard measures of guilt proneness and compliance used in psychology.

*Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK; *E-mail address*: eleni@essex.ac.uk

1 Introduction

In many strategic interactions, people's preferences are belief-dependent. Specifically, it has been shown that preferences depend on the beliefs people hold about others' expectations, and not only on their material payoff. A possible explanation for the importance of expectations on people's preferences involves the concept of guilt. People can feel guilty when they fail to live up to others' expectations (Baumeister et al., 1994). Classic guilt aversion models (Battigalli and Dufwenberg, 2007) assume that a feeling of guilt arises when people believe their behavior is responsible for the disappointment of others. When individuals are guilt averse, their choices should be positively correlated to what they think others expect from them (their second order beliefs).

A fast-growing empirical literature has been testing the predictions of guilt aversion models by investigating the correlation between choices, and either the decision makers' self-reported second order beliefs or the counterparts' expectations (first order beliefs) revealed to the decision makers before the choices are made. Many studies provide evidence that supports the guilt aversion hypothesis (Dufwenberg and Gneezy, 2000; Guerra and Zizzo, 2004; Charness and Dufwenberg, 2006; Bacharach et al., 2007; Reuben et al., 2009; Bellemare et al., 2011; Dufwenberg et al., 2011; Hauge, 2016; Khalmetski, 2016), while some authors challenge this conclusion (Ellingsen et al., 2010; Vanberg, 2008; Kawagoe and Narita, 2014). A number of attempts have been made to explain those seemingly conflicting findings (Bellemare et al., 2017a; Balafoutas and Fornwagner, 2017; Khalmetski et al., 2015) and some authors suggest that understanding the contextual and individual factors that interact with guilt aversion is crucial (Khalmetski et al., 2015). Still, the topic is poorly explored.

There are reasons to believe that the incentive scheme in a game might be a source of heterogeneity. Indeed, people might respond differently to the beliefs of others when the incentive scheme in a game varies, ultimately producing a significant interaction between guilt aversion and monetary incentives. For instance, some individuals that are not guilt averse when the stakes in a game are low might feel guilty when the monetary incentives in the game are high enough. Other individuals that are guilt averse when the stakes in a game are low might decide to consider only their material payoff when the monetary incentives become significantly large. Finally, some individuals might be guilt averse, regardless of the level of monetary incentives. Analyzing individual heterogeneity across different incentive schemes is crucial for understanding under which circumstances guilt aversion is likely to play a role.

This paper investigates how the rate of return in the investment game (Berg et al., 1995) interacts with guilt aversion, measured as the correlation between second order beliefs and choices. We use an experiment where participants play three rounds of the investment game with different rates of return and explore the presence of heterogeneity in the way people experience guilt under the different incentive schemes both at the aggregate and within-subject level. The results indicate that, on average, the coefficient of the interaction term between second order beliefs and choices becomes smaller as the rate of return increases. The within-subject analysis provides evidence of individual heterogeneity in the way people trade-off guilt and monetary incentives. For 60% of the participants considered in the analysis, the interaction term between second order beliefs and choices is not significantly affected by the rate of return, suggesting that the incentive scheme does not influence guilt aversion; among the rest there is a strong prevalence of participants for which the interaction term decreases with the rate of return.

To single out the characteristics of guilt averse participants, we analyze the influence of personality traits on guilt aversion. We collect data on personality traits using a Big-Five questionnaire (Costa Jr. and McCrae, 1992) and study the relationship between each Big-Five and the guilt aversion elicited in the investment game. We use two measures of guilt aversion. One captures 'unconditionally' guilt averse subjects (whose correlation coefficient between second order beliefs and choices is not affected by the incentive scheme) while the other considers the average correlation across the three rounds of the game. In the sample, we observe a significant negative correlation between guilt aversion and openness to experience, regardless of the definition of guilt aversion used. Moreover, neurotic players are more likely to be 'unconditionally' guilt averse. Those effects are robust upon controlling for standard measures of guilt proneness and compliance used in psychology.

This study is linked to the emerging literature studying heterogeneity in guilt aversion (Khalmetski et al., 2015; Attanasi et al., 2016, 2013; Bellemare et al., 2017b). The work most closely related to ours is Bellemare et al. (2017b). The authors analyze the trade-off between people's regards for their own payoff and letting down others across stakes in a mini-dictator game. They found that 60% of the participants in their experiment display a stake-dependent guilt aversion. Among those subjects, 10% have a sensitivity to guilt that is increasing with the stake while for 24% it is decreasing. Those results are not directly comparable with ours since Bellemare et al. (2017b) rely on a different theoretical specification and experimental design. Nonetheless, the outcomes of our experiment are in agreement with Bellemare et al. (2017b) as our study reports a significant proportion of participants with stake-dependent guilt aversion and, among them, there is a prevalence of participants for which guilt aversion decreases with the stakes in the game.

This paper recognizes the importance of personality traits in economic decisions and is the first study to explore their significance in a belief-dependent preferences framework. Several studies have shown that personality traits predict economic outcomes in the fields of education, health, labor market or crime (for a review of the literature see Almlund et al., 2011). Personality variables have been found to be more predictive than standard measures of cognition (Heckman et al., 2013) and more flexible over the life cycle (Cunha et al., 2010). Knowing how personality traits relate to guilt aversion is crucial to understanding the multiple influences that intervention targeting personality might generate and to assess how guilt aversion can be moderated via institutional design. Our research provides evidence of a robust link between personality traits and guilt aversion, as measured by the correlation between second order beliefs and choices. The result is strong, especially considering that other studies tried to relate experimental measures of trust (e.g., amount sent by the first player in the investment game) and reciprocity (e.g., amount returned by the second player) to personality traits (Evans and Revelle, 2008; Fahr and Irlenbusch, 2008; Ben-Ner and Halldorsson, 2010; Becker et al., 2012; Müller and Schwieren, 2012) finding at most weak relationships.

The rest of the paper is organized as follows. Section 2 describes the experimental design. In section 3, we present the theoretical background and the empirical strategy. In section 4, we show the results and discuss them. Section 5 concludes. The appendices report details of the experimental instructions.

2 Experimental design and procedures

2.1 Design

The experiment is structured in 3 phases: investment game, survey phase, and risk game. At the end of the experiment, participants are asked to complete a post-experimental questionnaire with demographics.

Phase 1: Investment game. We employ a variation of the investment game (Berg et al., 1995). In the game, two participants are paired, and one of them (A-player or A) is endowed with an amount of money (4 pounds). The A-player is asked to decide how much of her initial endowment she wants to transfer to the other player (B-player or B). The amount chosen by A is multiplied by the rate of return and then given to B. After receiving the amount, B decides how much to keep and how much to send back to A.

Before playing the game, participants are randomly assigned one role - either A or B - and must answer a question that tests their understanding of how payoffs are calculated to be able to proceed.[1] Once they provide the correct answer, they enter the belief elicitation stage. Each B-player is asked to guess the most common transfer that will be chosen by As in the room. Similarly, each A-player is requested to guess the most common amount that Bs will give back. Since the decision of the B-player depends on the amount received by A, the A-players have to make four different guesses, one for each possible level of transfer received by Bs (i.e., 1, 2, 3 or 4 multiplied by the rate of return). Figure 1 reproduces the belief elicitation stage when the rate of return is 3. Analogous procedures are in place for rates of return 1 and 5. To avoid any source of confusion, the reward scheme is simple and correct guesses are compensated with 1 pound (B-player) or 50 Cents (A-player).[2]

[1]If the answer provided is wrong, a warning window appears on the screen indicating the inappropriateness of the response and suggesting the range of value for the correct answer. During the experiment, two participants made several unsuccessful attempts and were not able to proceed, so they asked for assistance. The experimenter gave them a brief explanation of the game and payoff calculation. After the additional explanation, everybody gave the right answer and was able to proceed.

[2]The belief elicitation protocol invites rational participants to report the mode of their prior probability distribution. We choose this incentive scheme instead of other proper scoring rules because it is simple and easy to describe in instructions. As the A-players have four chances of getting a reward and Bs just one, it seems reasonable to offer different payments.

A-player
When Players B receive 3 pounds, the amount most frequently returned is: ...
When Players B receive 6 pounds, the amount most frequently returned is: ...
When Players B receive 9 pounds, the amount most frequently returned is: ...
When Players B receive 12 pounds, the amount most frequently returned is: ...

B-player
What do you think is the most common amount chosen by Players A when the rate of return is 3? ...

Figure 1: Belief elicitation stage when the rate of return is 3.

After the belief elicitation stage, each A-player is randomly matched with a B-player, and the game starts. Participants make their choices using a variant of the strategy method (Selten, 1967) akin to Khalmetski et al. (2015). First, the A-player is asked to make her choice of transfer for every possible guess of her matched B. The transfer corresponding to the true guess of the matched B-player is then selected, multiplied by the rate of return, and assigned to B. Afterwards, the B-player is asked to determine the amount to give back for every possible guess of his matched A.[3] The back transfer corresponding to the actual guess of the matched A-player is calculated, and the game ends. Figure 2 reproduces the decision stage when the rate of return is 3, assuming that the A-player transfers 1 pound to her matched B. Analogous procedures are in place for different levels of transfer and rates of return. Note that the choice set of B depends on the amount received by his matched A-player. As a consequence, Bs receiving different amounts face a different array of possible choices. Appendix A displays screen-shots of the decision stage of B receiving 6 and 9 pounds.

A-Player
If the guess of my matched Player B is 0, I transfer him/her the following amount: ...
If the guess of my matched Player B is 1, I transfer him/her the following amount: ...
If the guess of my matched Player B is 2, I transfer him/her the following amount: ...
If the guess of my matched Player B is 3, I transfer him/her the following amount: ...
If the guess of my matched Player B is 4, I transfer him/her the following amount: ...

B-Player
You received 3 pounds from A.
If the guess of my matched Player A is 0, I give him/her back the following amount: ...
If the guess of my matched Player A is 1, I give him/her back the following amount: ...
If the guess of my matched Player A is 2, I give him/her back the following amount: ...
If the guess of my matched Player A is 3, I give him/her back the following amount: ...

Figure 2: Decision stage when the rate of return is 3. We are assuming that A transfers 1 pound.

All participants play the game three times under different treatment conditions corresponding to the

[3]If the A-player transfers 0, the matched B-player skips this stage. He is simply asked to wait, and he is not informed that his counterpart transferred nothing. The possibility of long waiting time in the game was anticipated before the start of the experiment with a generic announcement.

rates of return 1, 3 and 5 (treatments R1, R3, and R5 respectively).[4] They are informed that they will keep the same role - either A or B - in all the rounds of the game. To minimize learning, participants always play with a different counterpart and receive feedback on the payoffs from the game only after the last round has been played. To avoid strategic reporting of beliefs, the beliefs for all the rates of return are elicited before the start of the game, and in the belief elicitation stage, participants are not told that their belief will be relevant to the payoff calculation.[5] The order of treatment is randomized.[6]

Phase 2: Survey phase. We use psychological surveys to measure personality traits, proneness to guilt and tendency to comply with group standards. The Big Five personality traits - openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism - are measured with the NEO Five Factory Revised (NEO FFI-R). This survey is the short version of the Revised NEO Personality Inventory (NEO PI-R) by Costa Jr. and McCrae (1992). The NEO FFI-R contains 60 items, 12 per trait. Each item is a statement about oneself (e.g., 'I am not a worrier') for which participants report their level of agreement on a 5-point Likert scale. Each Big Five is measured as the sum of the response to the 12 items referring to the trait. Big Five scores represent degrees of the personality traits; more extreme scorers have a higher probability of showing the distinctive features of the trait as reported in table 1.

We measure guilt proneness with the guilt scale of the Guilt and Shame Proneness Scale (GASP) (Cohen et al., 2011).[7] The GASP allows separating guilt from shame. Guilt is considered the result of some evaluation focused on one's behavior ('I did a bad thing') in a situation where the bad behavior has not been publicly exposed. In contrast, shame is defined as an evaluation on one's self ('I am a bad person') after failures are publicly exposed. The GASP is the only scale that considers both the self-behavior and private-public distinction. Considering that in the experiment decisions are anonymous, the non-public exposure of the behavior seems an important dimension, and this is the reason why we chose GASP over

[4]In the investment game, R>1 as both players must be better off when trust occurs (Alós-Ferrer and Farolfi, 2019). In the experiment, we consider the case R=1 as a baseline treatment. Nonetheless, to simplify the exposition, we refer to the baseline treatment as a round of the investment game.

[5]There is no reason to believe that not disclosing the information could mislead or deceive participants since players make choices using the strategy method, without getting to know the actual belief of the counterpart in the decision stage. Khalmetski et al. (2015) use a similar design.

[6]The beliefs elicitation stage and the investment game follow the same order of treatment.

[7]We slightly changed the choice of words in one of the items to make it more appropriate for an English student-based subject pool. Details are provided in Appendix B.

other more widely used guilt assessment scale such as TOSCA-3 (Tangney et al., 2000) that does not make the distinction.[8]

Trait	Distinctive features of high scorer in the trait
Openness to experience	Curious, willing to entertain novel ideas and unconventional values. Experiencing both positive and negative emotions more keenly than do closed individuals.
Conscientiousness	Purposeful, strong-willed, and determined. Scrupulous, punctual, and reliable.
Extraversion	Sociable, assertive, active, and talkative. Upbeat, energetic and optimistic.
Agreeableness	Altruistic, sympathetic to others and eager to help them. Believing that others will be equally helpful in return.
Neuroticism	Tendency to experience negative affects, inability to control impulses, poor coping with stress.

Table 1: Big Five traits, distinctive features.

The GASP measures emotional traits (guilt proneness) rather than emotional states (transitory feelings of guilt). The guilt scale of GASP is formed of 8 statements describing situations that people could face in day-to-day life followed by typical ways of responding to those situations (e.g. 'You lie to people but they never find out about it. What is the likelihood that you would feel terrible about the lies you told?'). Participants are asked to report the likelihood that they would react as it is described on a 7-point Likert scale. Guilt proneness corresponds to the average of the scores of the 8 items. Appendix B reports the complete questionnaire.

Finally, we measure the tendency to be susceptible to social influence and group pressures with the Cooperativeness scale of the Jackson Personality Inventory Revised (JPI-R)(Jackson, 2004). The survey is formed of 20 items, each item is a statement about oneself (e.g., 'I am very sensitive to what other people think of me') for which participants should indicate if they think the statement is true or false. The cooperativeness variable is the sum of the scores that indicate a propensity to cooperate. A person scoring high on the cooperativeness scale is susceptible to social influence and group pressure and tends to modify behavior to be consistent with standards set by others. This last measure further helps distinguishing guilt aversion from a generic motivation of social compliance possibly experienced by players.

[8]Bracht and Regner (2013) used the GASP scale in the context of guilt aversion.

Phase 3: Risk elicitation. In the last phase, participants are asked to take part in a risk game which uses the classical multiple price list method of Holt and Laury (2002). Each participant has to compare a list of 10 decisions between paired lotteries A and B. Along with this list of choices, the transition from the less risky gamble A to the more risky gamble B is rewarded by an increased risk premium. The number of safe choices returns a measure of individual risk attitudes.

Participants are paid one randomly chosen payoff from the investment game, the rewards for the correct beliefs in all the treatments, the payment for the survey, the outcome of the risk game and the show-up fee. The order in which participants face the investment game and the survey is balanced across sessions. The risk game is always played at the end, before the post-experimental questionnaire.

2.2 Procedure

The experiment took place in April and May 2017 at the EssexLab of the University of Essex. The participants were recruited through the ESSEXLab Online Recruitment System and invited to take part in a computerized session at the lab on campus. The research received Ethical Approval from the Social Sciences Faculty Ethics Sub-Committee of the University of Essex. Conforming to departmental ethics guidelines and current practice in experimental research with human subjects, participants read and signed informed consent before taking part in the session. The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007).

The research involved 144 participants; more than 90% were students. Experimental sessions consisted of 24 or 30 participants.[9] Written instructions were distributed at the beginning of each session. All participants received the same instructions. For the complete set of instructions, see Appendix C.

The sessions lasted on average 85 minutes, including random assignment to cubicles, reading the instructions, answering the post-experimental questionnaire and receiving payments. The participants earned between 8 and 31 pounds (including the show-up fee), with an average of 16.93 pounds.

[9]A pilot session with 12 participants was run two weeks before the experimental sessions. The pilot session ran smoothly, and the data are included in the analysis.

3 Theoretical background

The player of interest for our analysis is B. The reason is that the conditional choices of the A-players cannot have a clear interpretation in terms of belief-dependent preferences as they might be driven by strategic considerations. For instance, the A-players might think that the B-players with higher first order beliefs are also willing to give back higher amounts. In this case, a positive correlation between conditional transfers and beliefs would be driven by As' willingness to maximize payoff and not guilt aversion. In this section, we briefly introduce a guilt aversion model applied to the B-player of our experiment and outline the empirical strategy.

3.1 Guilt aversion model

We assume that B has belief-dependent preferences according to a model of guilt aversion adapted from Khalmetski et al. (2015).[10] In the game, B decides an amount $t_i \in [0, x_i R]$ to send back to A, where x_i is the transfer received from A and R the rate of return. A holds a probability distribution over B's actions and beliefs.[11] We denote the cumulative distribution function (cdf) of A's first order beliefs as F_A and the corresponding probability density function (pdf) as f_A. The cdf of B's second order beliefs conditional on the amount transferred by A is defined as $F_{BA}(z) = \mathbb{E}_B[F_A \mid x_i]$, and the pdf is $f_{BA}(z)$, where z is a random variable taking values on all his possible conditional second order beliefs.

B faces a trade-off between his material payoff (m_i) and a psychological cost stemming from anticipated guilt. The utility from material payoff is assumed to be increasing and strictly concave ($m_i' > 0$, $m_i'' < 0$). Given that B's monetary outcome is $x_i R - t_i$, his utility function is:

$$U_i(t_i, f_{BA}) = m_i(x_i R - t_i) - \beta_i \max\{0, \mathbb{E}_B[z - t_i \mid z \geq t_i]\}$$

where $\max\{0, \mathbb{E}_B[z - t_i \mid z \geq t_i]\}$ is B's level of guilt from falling below A's expectations and β_i measures

[10] Khalmetski et al. (2015) assume that people not only care about not disappointing others but may also get utility from positively surprising. Since the focus of this paper is on guilt aversion, we simplify their model by assuming that the propensity to make positive surprises is zero.

[11] As Khalmetski et al. (2015) note, 'the distribution-wise representation of beliefs provides a natural way to make the optimal transfer monotonically changing with the second order belief.'

guilt sensitivity.

B is unaware of A's first order belief, but he observes an informative signal θ_j which is equal to the median of $F_{BA}(z)$ and corresponds to A's elicited belief in the experiment. Assuming that a higher value of the signal θ_j, induces first order stochastic dominance in the conditional distribution of beliefs F_{BA} conditional on a lower signal, the strategy method returns a mapping of signals θ_j (and hence beliefs) to back transfers $t_i^*(\theta_j)$. Under a few additional assumptions,[12] Khalmetski et al. (2015) show that if a player is guilt averse ($\beta_i > 0$), $t_i^*(\theta_j)$ is systematically positively correlated with i's induced second order beliefs.[13]

3.2 Empirical strategy

Building on the implications of the model introduced in the previous section, we aim to study three aspects of guilt aversion. Before introducing them and the statistical strategy, we care to clarify the terminology adopted throughout the rest of the paper. We use the term 'second order beliefs' to refer to B's second order beliefs induced by the strategy method after observing the transfer received by A. Hence, we are not referring to B's *initial* beliefs but rather his *conditional* second order beliefs (Attanasi et al., 2016).

3.2.1 Guilt aversion

We study the correlation between conditional choices and second order beliefs to compare our results to other studies on guilt aversion. First, we use a non-parametric correlation test both at the aggregate and within-subject level, separately for each treatment. Then, we estimate the coefficients of the linear regression model:

$$\text{Conditional back transfer}_{ir} = \alpha + \beta_1 \text{Second order belief}_{ir} + \beta_2 \text{Treatment}_r + \beta_3 X_i + \epsilon_{ir} \tag{1}$$

[12]The assumptions are: B has some preferences regarding guilt ($\beta_i > 0$); B's marginal monetary cost of giving back is larger than half of his sensitivity to guilt $m_i^{`}(1 - t_i) \geq \frac{\beta_i}{2}$; $F_{BA}(z \mid \theta_j')$ is continuously differentiable on $[0, x_i R]$x$[0, x_i R]$. We refer to Khalmetski et al. (2015) for a discussion on the assumptions.

[13]As the stated results can be directly derived from the model of Khalmetski et al. (2015) by assuming that the propensity to make positive surprises is zero, we refer to their paper for the formal proof of the results.

where Treatment contains indicator variables for each treatment r with $r \in R = \{1, 3, 5\}$, and X_i is a set of individual-level controls. We use a panel approach to perform the analysis at the aggregate level and obtain average guilt aversion for each treatment. Then, we perform the same regression but within-subject to analyze the presence of heterogeneity in the sample.

3.2.2 Interaction between guilt aversion and the incentive scheme

We study the interaction between guilt aversion and the rate of return. This is an original contribution of the paper. As mentioned in the introduction, there are reasons to believe that guilt aversion might interact with the incentive scheme in the game as people might trade-off their feeling of guilt with material payoff. We estimate the regression:

$$\begin{aligned}\text{Conditional back transfer}_{ir} = \alpha &+ \beta_1 \text{Second order belief}_{ir} + \beta_2 \text{Treatment}_r \\ &+ \beta_3 \text{Treatment x Second order belief}_{ir} + \beta_4 X_i + \epsilon_{ir} \qquad (2)\end{aligned}$$

Also in this case, we perform the analysis both at the aggregate and within-subject level to comment on individual heterogeneity in the sample.

3.2.3 Guilt aversion and personality traits

To analyze the influence of personality traits on guilt aversion, we estimate the parameters of the linear regression model:

$$\begin{aligned}\text{Guilt aversion}_i = \alpha &+ \beta_1 \text{Openness to experience}_i + \beta_2 \text{Conscientiousness}_i + \beta_3 \text{Extraversion}_i \\ &+ \beta_4 \text{Agreeableness}_i + \beta_5 \text{Neuroticism}_i + \beta_6 Z_i + \epsilon_i \qquad (3)\end{aligned}$$

where Z_i is a vector of control variables. We employ two definitions of guilt aversion - one discrete and one continuous. First, we calculate the within-subject correlation coefficients between second order beliefs and choices for every treatment. Then, for the discrete definition, we construct a dummy variable which assumes value 1 for participants whose correlation coefficient is positive and significant (at least at 5% level) in all treatments. For the continuous definition of guilt aversion, we take the average of the correlation coefficients across treatments. It follows that the discrete definition captures ‘unconditionally’ guilt averse participants.[14]

4 Results

In the following sections, we report the results of the experiment. First, we present descriptive statistics concerning the investment game and the psychological surveys. Then, we introduce the main findings.

4.1 Descriptives

4.1.1 Investment game

We present descriptive statistics concerning the realized back transfer - intended as the back transfer corresponding to the actual guess of A in each pair of players - and the conditional back transfer.

Realized back transfer

Figure 3 displays the frequency of the realized transfers and back transfers in each treatment. The graphs on the left hand side of the figure show the number of A-players choosing the amount reported on the x-axis. The bars are coloured in different shades of grey with lighter shades indicating lower amounts. The majority of the A-players (23 players) transfers 0 in R1. As the rate of return increases, the A-players more frequently transfer positive amounts (56 players in R3 and 63 in R5). The graphs on the right of figure 3 show the number of B-players choosing the amount reported on the x-axis as a back transfer. The bars coloured in different shades of grey correspond to the amount received from the A-players. Since the game is played sequentially, the B-players receiving 0 from their matched A-players do not make any decision. Hence, the lightest shade of grey is missing from the right side of figure 3.[15] Among the B-players receiving positive transfers, the majority transfers back 0 in R1 (71% of the players). As the

[14]With ‘unconditional’ we mean that the rate of return does not change the degree of guilt aversion of a participant.

[15]Note that the number of A-players transferring 0 under each rate of return corresponds to the number of missing observation for B-players in each treatment. The sample size of B-players is indicated in the caption of figure 3.

rate of return increases a higher proportion of B-players back transfer positive amounts (60% in R3 and 78% in R5).

Table 2 displays summary statistics for the B-players' first order beliefs and realized back transfers. As expected, B-players expect to receive more as the rate of return increases - 1.43 in R1, 1.92 in R3 and 2.37 in R5. Moreover, within each treatment the B-players back transfer more the more they received from the As. Also, given the same amount received from the As, the mean amount back transferred increases with the rate of return.

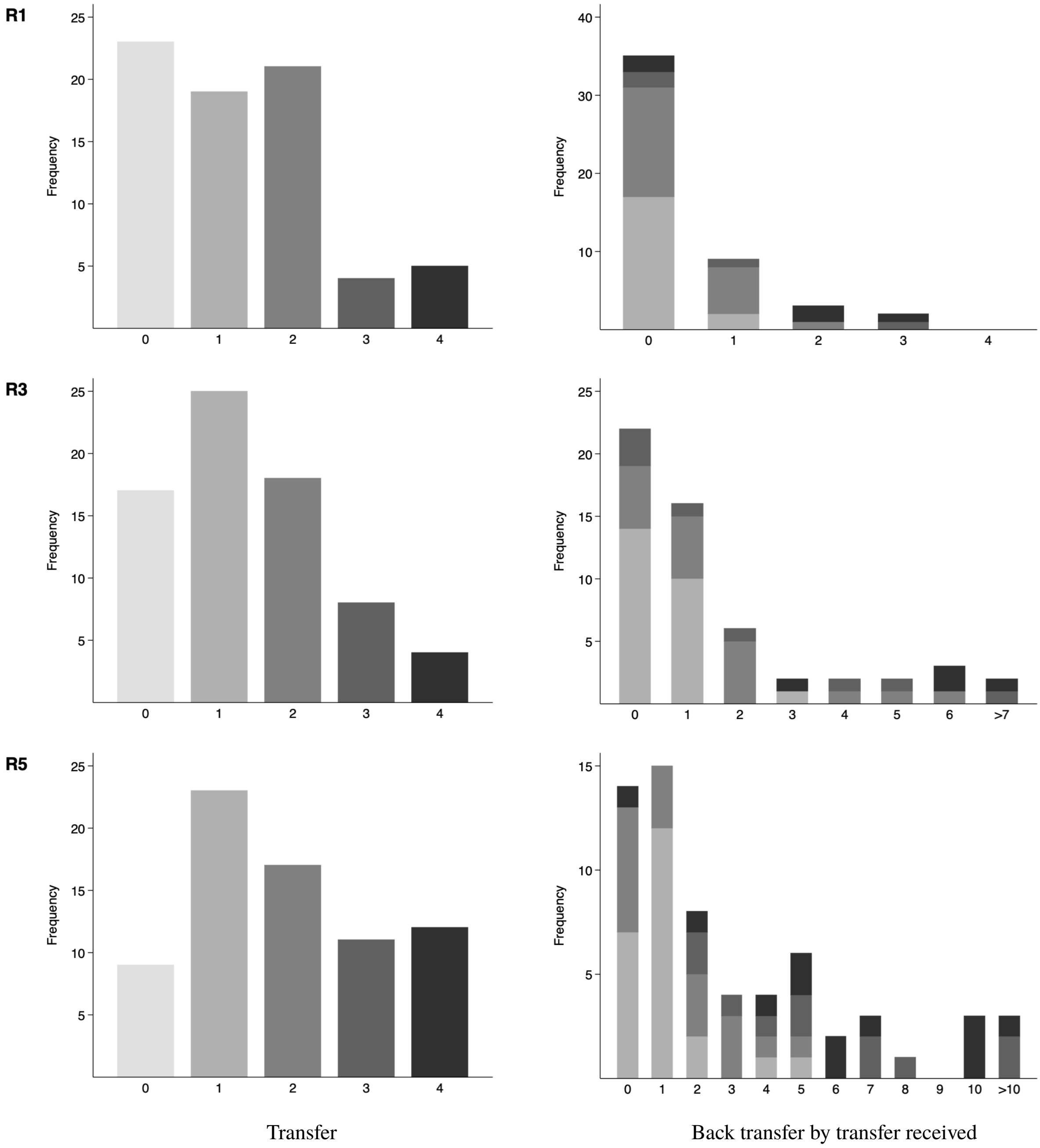


Figure 3: Frequency of realized transfer (left) and back transfer (right) by treatment. Different shades of grey correspond to different amounts of realized transfers (lighter shades indicate lower amounts). The sample size of A-players is 72 in every treatment. The sample size of B-players is 49 in R1, 55 in R3, and 63 in R5.

It is worth noting that the mean back transfer and belief in the three treatment conditions do not vary significantly according to whether the participants first play the game and then complete the survey or vice versa. To test the order effect in the experimental phases, we performed a two-sample t-tests with equal variances for each treatment, and none of the differences is significant. Also, the order of treatments does not significantly affect choices and beliefs. To test the order effect in treatments, we run a Bonferroni multiple-comparison test, and none of the pairwise differences between treatment orders is significant.

			Mean (SD)	Observations
R1	Belief on transfer		1.43 (1.18)	72
	Back transfer	if transfer=1	0.11 (0.32)	19
		if transfer=2	0.38 (0.59)	21
		if transfer=3	1.00 (1.41)	4
		if transfer=4	1.40 (1.34)	5
R3	Belief on transfer		1.92 (0.95)	72
	Back transfer	if transfer=1	0.52 (0.71)	25
		if transfer=2	1.67 (1.75)	18
		if transfer=3	2.62 (3.20)	8
		if transfer=4	5.75 (2.06)	4
R5	Belief on transfer		2.37 (1.26)	72
	Back transfer	if transfer=1	1.09 (1.24)	23
		if transfer=2	1.59 (1.58)	17
		if transfer=3	6.27 (3.98)	11
		if transfer=4	7.08 (5.12)	12

Table 2: Mean beliefs and realized back transfers of B-players split by level of transfer received from A. The mean values are split by treatment.

The outcomes described for the realized back transfer is similar to that observed in other experiments using the investment game with multiple rates of return (for a meta-analysis see Johnson and Mislin, 2011). This finding supports the validity of the experimental design. In fact, the strategy method could potentially bias choices by inducing a demand effect. Considering the outcome of the game, the demand effect does not seem to be significant.

Conditional back transfer

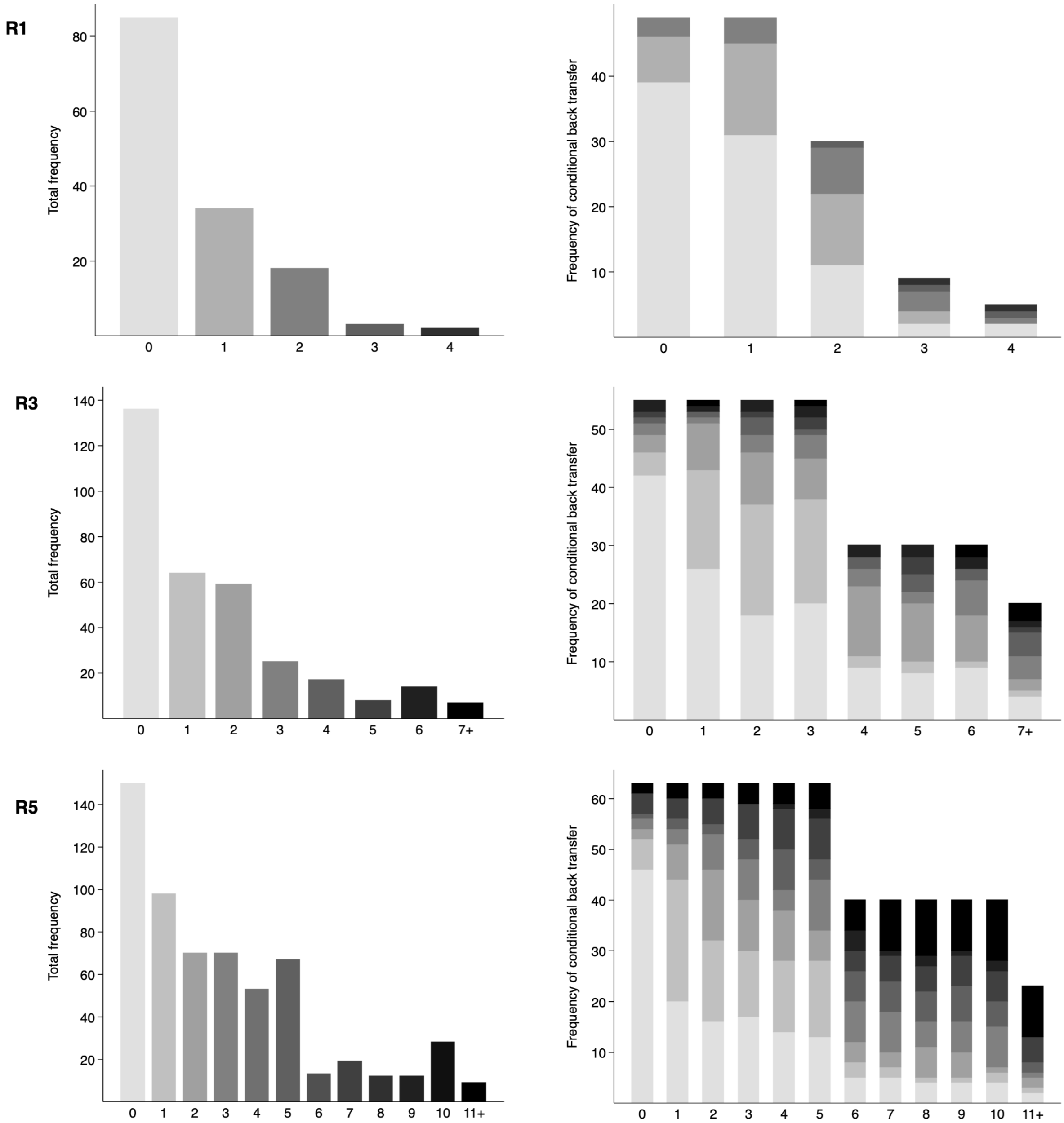
In this section, we consider all the choices of back transfer conditional on As' beliefs. We collected data on 1,073 conditional back transfers across 72 B-players.

Purely outcome-based utility theories (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) predict that a rational player should choose the same back transfer for any level of A's beliefs as beliefs should not affect choices. Hence, those theories predict no variation in the conditional choices made by B-players with the strategy method. However, this is not the case. In R1, 22 B-players (45%) change their conditional choice at least once. Excluding the choices conditional on a transfer equal to 1 - which grants only two options in B's choice set -, the percentage rises to 67% (20 players over 30). In R3, 39 players (71%) change their choice at least once, and 50 (79%) do so in R5.[16] Hence, the results of the experiment strongly suggest that preferences in the game are belief-dependent and not only driven by purely outcome-based motives.

Figure 4 displays the frequency of conditional back transfers in the three treatments. The graphs on the left show the frequency of conditional choices equal to the amount reported on the x-axis. As before, the bars are coloured in different shades of grey with lighter shades indicating lower amounts. The graphs on the right show how the conditional back transfers are distributed across the various levels of As' beliefs reported on the x-axis. For instance, consider the bar in the lightest shade of grey on the left side of the graph. In R1, the B-players choose 85 times to back transfer 0, conditional on any level of beliefs. The bars in the lightest shade of grey on the right indicate how back transfers of 0 are distributed across the various level of first order beliefs of A. In R1, the B-players choose 0 80% of the times when A believes to receive back 0. The proportion of 0 back transfers decreases as the level of As' beliefs increases (63% if As' believe to receive back 1, 37% if 2, 22% if 3), except when A expects to receive 4 (40%). Overall, figure 4 shows that in all the treatments, the proportion of 0 choices almost monotonically decreases as A's beliefs increases. Also, the B-players tend to choose higher back transfer conditional on higher levels of As' beliefs as guilt aversion models would predict.

Table 3 displays summary statistics of the B-players' conditional back transfers. The mean conditional back transfer is monotonically increasing in As' beliefs in all treatments. Given the same level of As' belief, the mean amount back transferred is increasing with the rate of return with few exceptions.

[16]The sample size is 55 in R3, and 63 in R5, which corresponds to the number of B-players receiving a positive transfer in those treatments.



Conditional back transfers

A's first order beliefs

Figure 4: Frequency of conditional back transfers by treatment. On the right side, conditional back transfers are split by the corresponding level of A's belief. Different shades of grey correspond to different amounts of conditional back transfers (lighter shades indicate lower amounts). The sample size is 142 in R1, 330 in R3, and 601 in R5.

			Mean (SD)	Observations
R1	Conditional back transfer	if A's first order belief = 0	0.26 (0.57)	49
		if A's first order belief = 1	0.45 (0.65)	49
		if A's first order belief = 2	0.93 (0.87)	30
		if A's first order belief = 3	1.67 (1.32)	9
		if A's first order belief = 4	1.80 (1.79)	5
R3	Conditional back transfer	if A's first order belief = 0	0.67 (1.50)	55
		if A's first order belief = 1	0.98 (1.50)	55
		if A's first order belief = 2	1.36 (1.52)	55
		if A's first order belief = 3	1.44 (1.86)	55
		if A's first order belief = 4-6	2.13 (1.93)	30
		if A's first order belief > 6	3.45 (2.76)	12
R5	Conditional back transfer	if A's first order belief = 0	0.90 (1.99)	63
		if A's first order belief = 1	1.62 (2.18)	63
		if A's first order belief = 2	1.98 (2.17)	63
		if A's first order belief = 3	2.29 (2.43)	63
		if A's first order belief = 4	2.52 (2.43)	63
		if A's first order belief = 5	2.60 (2.37)	63
		if A's first order belief = 5-10	4.59 (3.32)	40
		if A's first order belief > 10	6.61 (4.91)	23

Table 3: Mean conditional back transfers of the B-players split by level of As' first order beliefs. The mean values are split by treatment.

4.1.2 Surveys

Table 4 reports summary statistics on the psychological surveys.[17]

Variable name	Value Range	Mean	(SD)	Min Value	Max Value
Openness to experience	0-60	31.40	(6.60)	13	48
Conscientiousness	0-60	30.61	(8.22)	10	47
Extraversion	0-60	28.89	(7.97)	12	45
Agreeableness	0-60	31.07	(6.86)	14	44
Neuroticism	0-60	24.28	(8.19)	7	46
GASP	0-7	5.46	(0.89)	2.87	7
JPI-cooperativeness	0-20	7.01	(3.71)	0	14

Table 4: Surveys: summary statistics. Sample size: 72 B-players.

The mean values obtained in all the surveys are fairly similar to the normative values obtained for the general population (for details on the normative values see Costa Jr. and McCrae 1992, Cohen et al. 2011, and Jackson 2004).

[17]Even if all the 144 participants to the experiment took part in the survey phase, the table presents data referring to the sample of 72 B-players only. Note that there are no statistically significant differences between the B-players and the A-players for all the variables under analysis.

4.2 Results

4.2.1 Result 1 - Guilt aversion

In this section, we analyze data on conditional back transfer to check whether B-players' choices and second order beliefs are systematically correlated. We perform correlation and regression analysis both at the aggregate and within-subject level.

A simple Spearman correlation test between conditional back transfers and second order beliefs, reveals a positive and strongly significant correlation between the two variables in each treatment. The correlation coefficient in R1 is 0.4765 (p-value<.001), 0.3646 (p-value<.001) in R3, and 0.4689 (p-value<.001) in R5.

Figure 5 shows the distribution of coefficients resulting from the within-subject correlation analysis between conditional back transfers and second order beliefs performed separately for each treatment. The vast majority of the coefficients are positive, suggesting that many subjects might be motivated by guilt aversion.

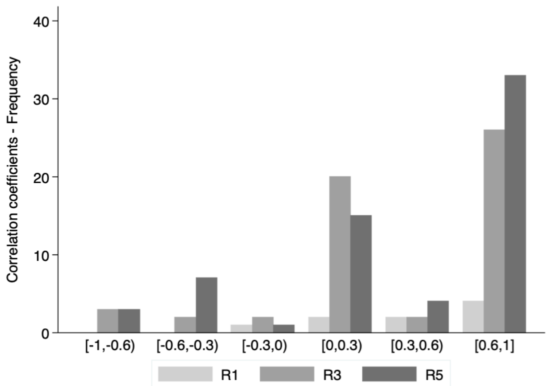


Figure 5: Distribution of the coefficients resulting from the within-subject correlation analysis between conditional back transfers and second order beliefs performed by treatment.

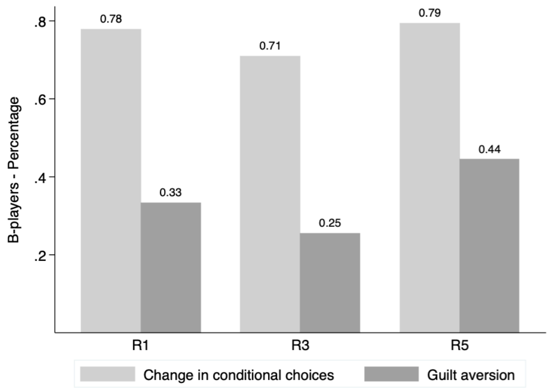


Figure 6: Proportion of B-players changing their choice of conditional back-transfer at least once - light grey bars - and having a positive and significant (at least at 5% level) within-subject correlation - dark grey bars.

Figure 6 displays the percentage of B-players changing their conditional choice at least once - light grey bars - and having a correlation coefficient positive and significant at least at 5% level according to a Spearman correlation test - dark grey bars - in each treatment.[18] The B-players represented with dark grey bars display a pattern of choices consistent with the guilt aversion model introduced in section 3.1. We observe 33% of such players in R1, 25% in R3, and 44% in R5.

To check whether the pattern described can be generated by a random process, we perform a Monte-Carlo simulation with 5,000 replications of random samples of conditional back transfers obtained by bootstrapping the original sample. In R1, on average, the share of participants with a positive and significant correlation between conditional back transfers and beliefs in random samples is 0.77% with a standard deviation of 0.26. None of the replications produce a sample with more than 22% of the players with significant correlation. In R3, we observe an average share of 1.57% (sd=0.92), and a maximum share of 11%. In R5, the average share is 4.45% (sd=1.64) and the maximum is 16%. The Monte-Carlo simulation provides support to the hypothesis the pattern observed in the experiment is the result of a systematic choice.

The regression analysis confirms the results of the correlation tests. Columns 1 to 3 of table 5 show the result of regression (1) using different sets of control variables. In column 1, we control for transfer received by the matched A-player, sex, age, number of siblings, risk aversion,[19] and order of treatment. In column 2, we additionally control for GASP and JPI-cooperativeness scales. In column 3, we also include random effects at the participant level. Since we normalize the conditional back transfer dividing it by the transfer received by As, the dependent variable is bounded between 0 and 1, we apply a two-limit Tobit model.

In all the specifications, the coefficient of second order beliefs is positive and significant. On average, a unit increase in second order belief generates a 3.7% increase in the dependent variable (i.e., proportion of transfer given back). Adding the GASP and JPI-cooperativeness scales and the random effects to the regression does not change the significance of the coefficient and has little effect on its dimension. Also, the B-players tends to return proportionally less as the rate of return increases. The difference is

[18]The minimum sample size to perform a statistically significant correlation is 4. The B-players receiving less than 3 pounds in R1 are thus excluded from the analysis. Sample size is 9 in R1, 55 in R3 and 63 in R5.

[19]We excluded 7 players from the sample due to inconsistencies in their answers to the Holt and Laury task. To calculate risk aversion, we consider the first switching point and remove from the sample players that switched their answer more than 3 times.

statistically significant between R1 and R3 and between R3 and R5 (Wald test on equality of coefficients returns p-value<0.01 in the specification with random effects). The result is consistent with previous findings suggesting that the B-players take into account the 'size of the pie' by adjusting downwards their back transfer (Johnson and Mislin, 2011).

Dependent var.:	Tobit Regression					
Conditional back transfer	(1)	(2)	(3)	(4)	(5)	(6)
Second order beliefs	0.0369***	0.0365***	0.0389***	0.2004***	0.1975***	0.2035***
	(0.0051)	(0.0050)	(0.0034)	(0.0298)	(0.0277)	(0.0267)
R3	-0.0795*	-0.0711	-0.0799**	0.2463***	0.2506**	0.2359***
	(0.0448)	(0.0510)	(0.0356)	(0.0935)	(0.1079)	(0.0742)
R5	-0.1472***	-0.1549***	-0.1572***	0.2349***	0.2217**	0.2384***
	(0.0428)	(0.0599)	(0.0447)	(0.0870)	(0.0991)	(0.0763)
Second order beliefs x R3				-0.1567***	-0.1546***	-0.1543***
				(0.0368)	(0.0358)	(0.0277)
Second order beliefs x R5				-0.1683***	-0.1658***	-0.1700***
				(0.0312)	(0.0297)	(0.0269)
RE at participant level			Yes			Yes
Psychological scales		Yes	Yes		Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Observations	954	954	954	954	954	954
Number of groups	65	65	65	65	65	65

Table 5: The excluded category for the indicator variable are R1 and second order beliefs x R1. Controls include: transfer received by A, sex, age, number of siblings, risk aversion, and order of treatment. Psychological scales: GASP and JPI-cooperativeness. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Figure 7 refers to the within-subject regression analysis and shows the distribution of statistically significant (at least at 5% level) coefficients from regressing conditional back transfers on second order beliefs for each B-player in the three treatments. Each bar in the graph represents a participants. Across the various treatments, more than 80% of the significant coefficients are positive suggesting that many players behave according to the predictions of the guilt aversion model.

Result 1 - There is a systematic positive correlation between second order beliefs and conditional back transfers, both at the aggregate and within-subject level. This result holds across the different incentive schemes considered.

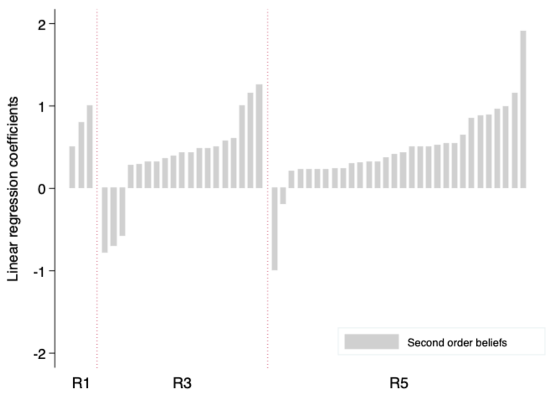


Figure 7: Distribution of coefficients, significant at least at 5% level, estimated in within-subject regressions of conditional back transfers on second order beliefs by treatment. Each bar represents one participant.

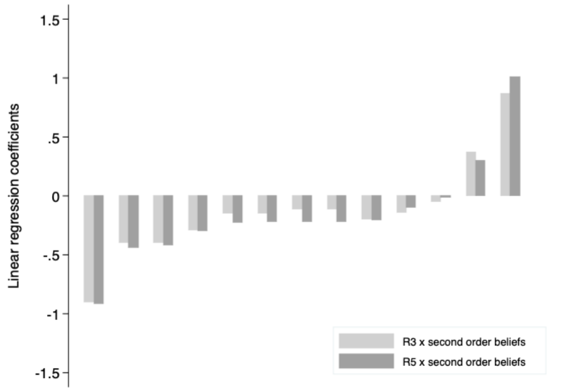


Figure 8: Distribution of coefficients estimated in within-subject regressions of conditional back transfers on second order beliefs fully interacted with treatment dummies. Each couple of stacked bars represents one participant. Only participants with at least one significant coefficient are represented.

4.2.2 Result 2 - Interaction between guilt aversion and the incentive scheme

In this section, we study how guilt aversion interacts with the incentive scheme. Table 5 - columns 4 to 6 - shows the result of regression (2). As in the previous section, the dependent variable is normalized over the transfer received. We use a Tobit specification and adopt the same sets of control variables.

The coefficient of second order beliefs is positive and significant which implies that, on average, the behavior of B-players is consistent with the prediction of the guilt aversion model in R1, after controlling for covariates. Both the interaction terms of treatment with beliefs are negative and strongly significant. The Wald test indicates that the coefficient for R3 interacted with beliefs is different from the interaction with R5 (p-value=0.0547) in the specification including random effect), but the effect is only marginally significant using the other specifications. The coefficients of the dummy variables R3 and R5 are positive and significant indicating that the B-players return proportionally more as the rate of return increases, once controlling for guilt aversion.

We performed a within-subject analysis where we regress conditional back transfer (normalized as in the aggregate analysis) on second order beliefs, indicator variables for treatment, and interaction between the two terms controlling for the transfer received. The analysis considers only the B-players that received a positive transfer in all treatments (32 participants). For 60% of them no interaction term is significantly different from 0). For the remaining participants, there is at least one interaction term which is significantly different from 0.[20] Figure 8 displays the coefficients of the interaction term resulting from the within-subject regression for those participants. The vast majority of the coefficients - 85% - are negative.

Result 2 - On average the coefficient of the interaction term between second order beliefs and treatment decreases with the rate of return. The effect is strong when the rate of return increases from 1 to 3 and much weaker when the increase is from 3 to 5. There is individual heterogeneity in the sample. For 60% of the players, no interaction term is significantly different from 0; among the other 40%, there is a strong prevalence of players for which both interaction terms are significantly smaller than 0.

4.2.3 Result 3 - Guilt aversion and personality traits

Table 6 shows the result of regression (3). We use two definitions of guilt aversion, as anticipated in section 4.2.3. In columns 1 and 2, the dependent variable is an individual level dummy variable which assumes value 1 whenever a player is guilt averse across all the incentive schemes, 0 otherwise.[21] We use a probit specification and the table reports the marginal effects at mean values of the independent variables. In columns 3 and 4, we use the average of the correlation coefficients between conditional back transfers and second order beliefs across the three treatments. The small set of control includes risk aversion, age, sex, number of siblings, average back-transfer, and average of first order beliefs. The large set of controls additionally includes GASP and JPI-cooperativeness scales. In all the regressions standard errors are clustered at the session level.

[20]70% of such regressions have both coefficients statistically significant.

[21]Following the notation introduced in the previous section, guilt aversion corresponds to a Spearman correlation coefficient that is positive and significant at least at 5% level.

Dependent variable: Guilt aversion	Probit (1)	(2)	OLS (3)	(4)
Openness to experience	-0.0311***	-0.0345***	-0.0298**	-0.0281**
	(0.0105)	(0.0128)	(0.0087)	(0.0100)
Conscientiousness	0.0083	0.0049	0.0121	0.0121
	(0.0055)	(0.0030)	(0.0120)	(0.0122)
Extraversion	-0.0076	-0.0062	0.0004	-0.0004
	(0.0065)	(0.0053)	(0.0057)	(0.0058)
Agreeableness	0.0078	0.0011	-0.0097	-0.0122
	(0.0061)	(0.0056)	(0.0098)	(0.0094)
Neuroticism	0.0189**	0.0201***	0.0147	0.0122
	(0.0074)	(0.0067)	(0.0135)	(0.0164)
Psychological scales		Yes		Yes
Controls	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes
Observations	65	65	65	65
(Pseudo) R-squared	0.3028	0.3939	0.2051	0.2164
Log-likelihood	-23.6103	-20.5263		

Table 6: Columns 1 and 2 report marginal effects at the mean of the independent variables. Controls include: risk aversion, age, number of siblings, average back-transfer, and average of first order beliefs. Psychological scales: GASP and JPI-cooperativeness. Standard errors clustered at session level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Only one personality factor is robustly significant throughout all the specifications. The higher a participant scores in openness to experience the less likely he is to display a positive correlation between second order beliefs and conditional back transfers. Using the discrete definition of guilt aversion an increase of 1 point in the openness to experience scale generates 3 to 4% decrease in the likelihood of being guilt averse at mean values of the independent variables. Using the continuous definition (i.e., average correlation) the effect is slightly smaller but still significant. Baumeister et al. (1994) say that 'combining empathic distress and exclusion anxiety furnishes a potentially powerful basis for analyzing guilt and predicting its patterns of occurrence.' Openness to experience positively correlates with good relationship satisfaction (McCrae and Sutin, 2009), and the sign of the coefficient seems to support the hypothesis that less socially included individuals are more prone to experience guilt.

Consistently with the psychological literature, the participants who score high on neuroticism are more likely than average to experience guilt (Boyle et al., 2008), and indeed, they are more likely to be display a behavior consistent with the predication of the guilt aversion model. The coefficient, however, is statistically significant only using the discrete definition of guilt aversion suggesting that more neurotic participants are more likely to be 'unconditionally' guilt averse, but there is no significant effect on the average degree of guilt aversion. The coefficients of the other personality traits are not significant.

The results of the regressions in table 5 do not vary significantly after the inclusion of GASP and JPI-cooperativeness scales as controls (columns 2 and 4). This outcome, together with the results of the regressions in the previous sections, seems to suggest that the definition of guilt aversion adopted in the economic literature is capturing something different from the emotional trait of guilt proneness captured by surveys typically used in psychology.[22]

Result 3 - Participants scoring high in openness to experience are significantly less likely to display a positive correlation between second order beliefs and conditional back transfers. Neurotic participants are more likely to display such positive correlation, but only when it does not depend on the incentive scheme.

5 Conclusion

An increasing amount of empirical studies suggests that in many strategic interactions, people's preferences depend on what others expect from them. Aversion to guilt seems to drive some individuals to reciprocate trust when they are expected to do so. The data presented in this paper are consistent with the guilt aversion hypothesis and suggest that purely outcome-based motives, such as altruism or fairness, are insufficient to explain players' behaviors in the investment game.

My study offers an analysis of two potential sources of heterogeneity that have received little or no attention in the literature - incentive schemes and personality traits. The data reveal significant heterogeneity in the way people trade-off guilt and incentives. Guilt aversion - as measured by the correlation between second order beliefs and choices - depends on the rate of return for a significant proportion of participants, and, on average, guilt aversion decreases as the stakes in the game become higher. The results of the experiment also suggest that guilt aversion is negatively correlated to openness to experience and positively to neuroticism.

[22]Bellemare et al. (2019) find a positive correlation between guilt aversion measured within a framework of psychological games and guilt aversion measured using a slightly modified version of TOSCA-3 (Tangney et al., 2000). In the present experiment, such a correlation is not present. However, the possibility to compare the present results with Bellemare et al. (2019) is limited by the fact that the surveys and experimental design used in the two studies are different.

This paper contributes to unveiling the contextual and individual factors that influence guilt aversion. The analysis suggests that guilt aversion is highly heterogeneous and sensitive to both the experimental context and individual characteristics such as personality traits. How these results generalize to other strategic interactions is left for future research.

References

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality psychology and economics. In *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Elsevier.

Alós-Ferrer, C. and F. Farolfi (2019). Trust games and beyond. *Frontiers in Neuroscience 13*, 887.

Attanasi, G., P. Battigalli, and E. Manzoni (2016). Incomplete-Information Models of Guilt Aversion in the Trust Game. *Management Science 62*(3), 648–667.

Attanasi, G., P. Battigalli, and R. Nagel (2013). Disclosure of belief-dependent preferences in a trust game. Technical report, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Bacharach, M., G. Guerra, and D. J. Zizzo (2007). The Self-Fulfilling Property of Trust: An Experimental Study. *Theory and Decision 63*(4), 349–388.

Balafoutas, L. and H. Fornwagner (2017). The Limits of Guilt. *Journal of the Economic Science Association 3*(2), 137–148.

Battigalli, P. and M. Dufwenberg (2007). Guilt in Games. *American Economic Review 97*(2), 170–176.

Battigalli, P. and M. Dufwenberg (2009). Dynamic Psychological Games. *Journal of Economic Theory 144*(1), 1–35.

Baumeister, R. F., A. M. Stillwell, and T. F. Heatherton (1994). Guilt: An Interpersonal Approach. *Psychological Bulletin 115*(2), 243–267.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). The Relationship Between Economic Preferences and Psychological Personality Measures. *Annual Review of Economics 4*(1), 453–478.

Bellemare, C., A. Sebald, and M. Strobel (2011, apr). Measuring the Willingness to Pay to Avoid Guilt: Estimation Using Equilibrium and Stated Belief Models. *Journal of Applied Econometrics 26*(3), 437–453.

Bellemare, C., A. Sebald, and S. Suetens (2017a). A Note on Testing Guilt Aversion. *Games and Economic Behavior 102*, 233–239.

Bellemare, C., A. Sebald, and S. Suetens (2017b). Heterogeneous Guilt Sensitivities and Incentive Effects. *Experimental Economics 1*(21).

Bellemare, C., A. Sebald, and S. Suetens (2019). Guilt aversion in economics and psychology. *Journal of Economic Psychology 73*(452), 52–59.

Ben-Ner, A. and F. Halldorsson (2010). Trusting and Trustworthiness: What Are They, How to Measure Them, and What Affects Them. *Journal of Economic Psychology 31*(1), 64–79.

Berg, J., J. Dickhaut, and K. A. McCabe (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior 10*, 122–142.

Bolton, G. E. and A. Ockenfels (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review 90*(1), 166–193.

Boyle, G. J., G. Matthews, and D. H. Saklofske (2008). *The SAGE Handbook of Personality Theory and Assessment*. SAGE Publications.

Bracht, J. and T. Regner (2013). Moral emotions and partnership. *Journal of Economic Psychology 39*, 313–326.

Charness, G. and M. Dufwenberg (2006). Promises and Partnership. *Econometrica 74*(6), 1579–1601.

Cohen, T. R., S. T. Wolf, A. T. Panter, and C. A. Insko (2011). Introducing the GASP Scale: A New Measure of Guilt and Shame Proneness. *Journal of Personality and Social Psychology 100*(5), 947–966.

Costa Jr., P. T. and R. R. McCrae (1992). *Neo Personality Inventory–Revised (NEO-PI-R) and Neo Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Cunha, F., J. Heckman, and S. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Dufwenberg, M., S. Gächter, and H. Hennig-Schmidt (2011, nov). The Framing of Games and the Psychology of Play. *Games and Economic Behavior 73*(2), 459–478.

Dufwenberg, M. and U. Gneezy (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior 30*, 163–182.

Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010, jan). Testing Guilt Aversion. *Games and Economic Behavior 68*(1), 95–107.

Evans, A. M. and W. Revelle (2008). Survey and Behavioral Measurements of Interpersonal Trust. *Journal of Research in Personality 42*(6), 1589–1597.

Fahr, R. and B. Irlenbusch (2008). Identifying Personality Traits to Enhance Trust between Organisations: An Experimental Approach. *Managerial and Decision Economics 29*(6), 469–487.

Fehr, E. and K. M. Schmidt (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics 114*, 817–868.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics 10*(2), 171–178.

Guerra, G. and D. J. Zizzo (2004). Trust Responsiveness and Beliefs. *Journal of Economic Behavior and Organization 55*(1), 25–30.

Hanushek, E. A. and L. Woessmann (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal 116*(510), C63–C76.

Hauge, K. E. (2016). Generosity and Guilt: The Role of Beliefs and Moral Standards of Others. *Journal of Economic Psychology 54*, 35–43.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–86.

Holt, B. C. A. and S. K. Laury (2002). Risk Aversion and Incentive Effects. *American Economic Review 92*(5), 1644–1655.

Jackson, D. N. (2004). *JPI-R: Jackson Personality Inventory, Revised*. Sigma Assesment Systems, Incorporated.

Johnson, N. D. and A. A. Mislin (2011). Trust Games: A Meta-Analysis. *Journal of Economic Psychology 32*(5), 865–889.

Kawagoe, T. and Y. Narita (2014). Guilt Aversion Revisited: An Experimental Test of a New Model. *Journal of Economic Behavior and Organization 102*, 1–9.

Khalmetski, K. (2016). Testing Guilt Aversion with an Exogenous Shift in Beliefs. *Games and Economic Behavior 97*, 110–119.

Khalmetski, K., A. Ockenfels, and P. Werner (2015, sep). Surprising Gifts: Theory and Laboratory Evidence. *Journal of Economic Theory 159*(61), 163–208.

McCrae, R. R. and A. R. Sutin (2009). Openness to Experience. *Handbook of Individual Differences in Social Behavior 15*, 257–273.

Müller, J. and C. Schwieren (2012). What Can the Big Five Personality Factors Contribute to Explain Small-Scale Economic Behavior ? *Tinbergen Institute Discussion Paper*.

Reuben, E., P. Sapienza, and L. Zingales (2009, aug). Is Mistrust Self-Fulfilling? *Economics Letters 104*(2), 89–91.

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In *Beiträge zur experimentellen Wirtschaftsforschung*, pp. 136–168. Tübingen, Germany: Sauermann, Heinz. Mohr.

Tangney, J. P., R. L. Dearing, P. E. Wagner, and R. H. Gramzow (2000). *The Test of Self-Conscious Affect-3 (TOSCA-3)*. Fairfax, VA: George Mason University.

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica 76*(6), 1467–1480.

Online Appendix

A Appendix: Decision stage of B-player

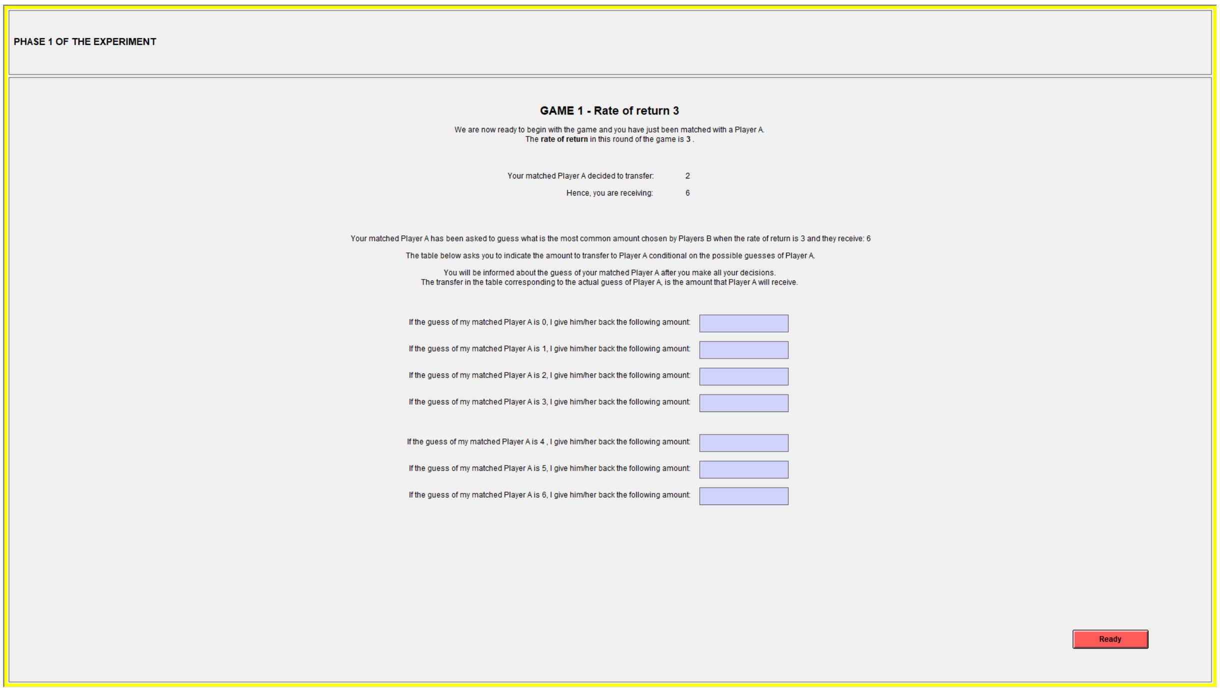


Figure A1: Player B (receiving 6 pounds), treatment R3

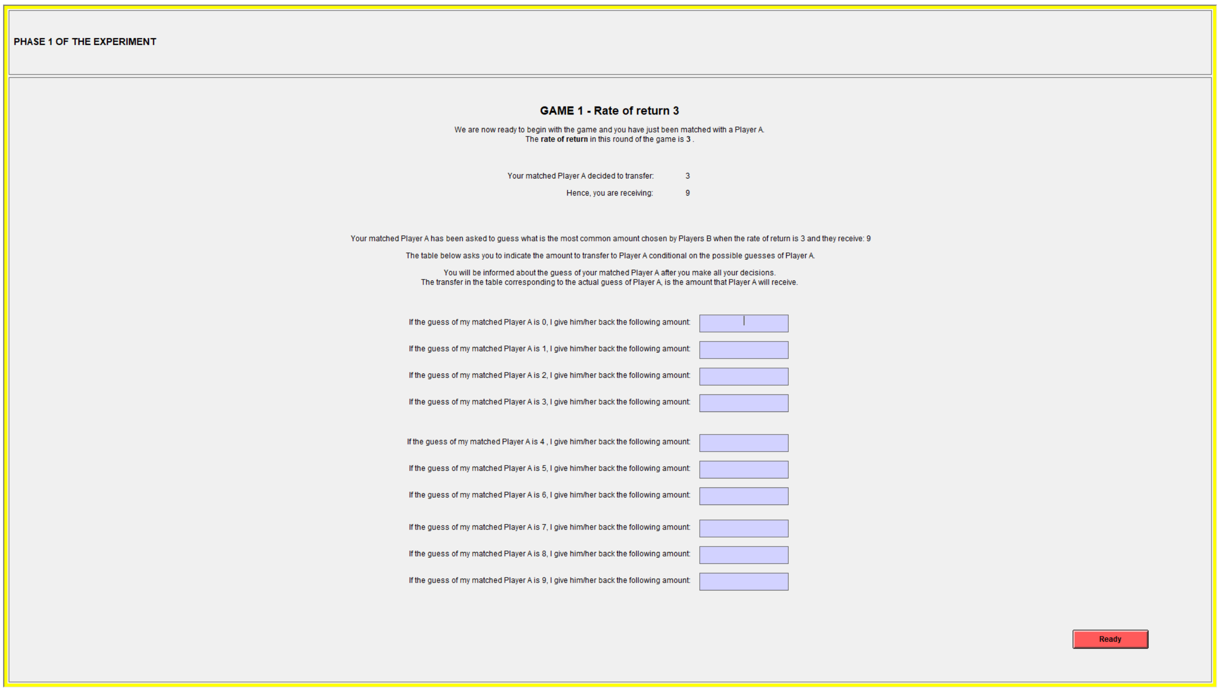


Figure A2: Player B (receiving 9 pounds), treatment R3

B Appendix: Guilt proneness scale

Instructions: In this questionnaire you will read about situations that people are likely to encounter in day-to-day life, followed by common reactions to those situations. As you read each scenario, try to imagine yourself in that situation. Then indicate the likelihood that you would react in the way described.

1. After realizing you have received too much change at a store, you decide to keep it because the salesclerk doesn't notice. What is the likelihood that you would feel uncomfortable about keeping the money?
2. You are privately informed that you are the only one in your group that did not receive the mention of honor because you skipped too many days of school.[1] What is the likelihood that this would lead you to become more responsible about attending school?
3. You reveal a friend's secret, though your friend never finds out. What is the likelihood that your failure to keep the secret would lead you to exert extra effort to keep secrets in the future?
4. You secretly commit a felony. What is the likelihood that you would feel remorse about breaking the law?
5. You strongly defend a point of view in a discussion, and though nobody was aware of it, you realize that you were wrong. What is the likelihood that this would make you think more carefully before you speak?
6. At a coworker's housewarming party, you spill red wine on their new cream-colored carpet. You cover the stain with a chair so that nobody notices your mess. What is the likelihood that you would feel that the way you acted was pathetic?
7. While discussing a heated subject with friends, you suddenly realize you are shouting though nobody seems to notice. What is the likelihood that you would try to act more considerately toward your friends?
8. You lie to people but they never find out about it. What is the likelihood that you would feel terrible about the lies you told?

[1] The original sentence was 'You are the only one in your group that did not make the honor society'. We slightly changed the choice of words in the scenario to make the situation more appropriate for an English student-based subject pool.

The answers are given on a 7-point Likert scale (1 Very Unlikely - 7 Very Likely). The GASP is scored by averaging the 4 items in each subscale (Guilt-NBE: 1, 4, 6, 8; Guilt repair: 2, 3, 5, 7).

C Appendix: Experimental instructions

GENERAL INFORMATION

Welcome! You are about to participate in a decision making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions. This set of instructions is for your private use only. During the experiment you are not allowed to communicate with anybody. In case of questions, please raise your hand. Then, we will come to your seat and answer your questions. Any violation of this rule excludes you immediately from the experiment and all payments. For your participation you will receive a show-up fee of 4 pounds. You can earn additional amounts of money. Below we describe how.

EXPERIMENTAL INSTRUCTIONS

The experiment is composed of 3 phases: 2 involve the participation in games and the other one consists in surveys. First, you will complete the surveys. Then, you will take part in GAME 1. Lastly, you will participate in GAME 2.

PHASE 1 - SURVEY

You will take part in 3 surveys. On the screen you will find the instructions on how to answer the questions in each of the survey. If you complete all the questionnaires you will receive a payment of 4 pounds.

PHASE 2 - GAME 1

There are two players: Player A and Player B. You will be randomly selected to play as Player A or B and you will play in the same role during the entire experiment.

Player A decision. Player A is endowed with 4 pounds and has to decide an amount she/he wants to transfer to Player B. Player A can choose any integer between 0 and 4 pounds. Player B receives this amount multiplied by a rate of return R. The game is played three times, each time with a different rate of return: R = 1, 3 or 5. At each new round of the game you will face a different Player B. For example, if Player A decides to give 2 pounds to Player B:

- When the rate of return is 1, Player B receives: 2 x 1 = 2 pounds
- When the rate of return is 3, Player B receives: 2 x 3 = 6 pounds
- When the rate of return is 5, Player B receives: 2 x 5 = 10 pounds

Note: if Player A decides to transfer nothing, Player B gets nothing with any rate of return.

Player B decision. After receiving the money, Player B has to decide an amount to give back to Player A. Player B can choose any integer amount between 0 and the amount received. Recall that the game is played three times, each time with a different rate of return: R = 1, 3 or 5. At each new round of the game you will face a different Player A.

Payoffs GAME 1. The payoffs are calculated as follows:

- Payoff of Player A = 4 pounds-Amount transferred to Player B + Amount Player B gives back
- Payoff of Player B = Amount received from Player A x Rate of return - Amount given back to A

Your guesses. Before starting the game, you are also asked to report your beliefs regarding the other players' decision. If your guess is accurate you will receive rewards (details about the amounts and how the accuracy is computed will appear on your screen). Please note that ALL THE GUESSES will be rewarded, regardless of the rate of return.

PHASE 3 - GAME 2

A new instruction sheet will be provided upon termination of the previous two phases.

CONFIDENTIALITY

All your decisions will be treated confidentially both during the experiment and after the experiment. This means that none of the other participants nor the researcher will ever know the decisions you made.

PAYMENT

At the end of the experiment, one of the payoffs from the three rounds of GAME 1 will be randomly selected and paid together with the rewards for the correct guesses, the payoff from GAME 2, the payment for completing the surveys and the show-up fee. Please remain seated after you finish the experiment. We will come to your desk with an envelope containing the money you earned. Don't forget to check the amount in the envelope, sign the receipt and leave it on your desk.

Matching and Confidence*

Mehmet Yigit Gurdal
Bogazici University

Elisabetta Leni
University of Essex

Friederike Mengel †
University of Essex
and Lund University

Abstract

We study how matching affects confidence. Our lab experiment allows us to identify the effect of being matched with others of either similar or dissimilar performance (assortative or disassortative matching) on people's confidence in their own ability. Across a variety of tasks we find that assortative matching does not have a substantial nor statistically significant effect on confidence compared to a control group with random matching. By contrast, disassortative matching has a negative effect on confidence on average that is driven by the bottom half of performers. This group becomes substantially less confident compared to random matching. We discuss potential mechanisms and implications of this result.

*We thank the British Academy (Newton Mobility Grant NMG2R2-100082) for valuable financial support.
†Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ. Department of Economics, Lund University, SE-220 07 Lund (SE); *e-mail*: fr.mengel@gmail.com

1 Introduction

A crucial aspect of institutional design, relevant for workplace organization, schools and the higher educations sector alike, is how to match people to achieve the best outcomes. Outcomes that have received a lot of attention in the literature include the academic performance of students (Sacerdote, 2001; Hanushek and Woessmann, 2006; Feld and Zoelitz, 2017), the performance of workers and work teams (Mas and Moretti, 2009; Jackson and Bruegmann, 2009; Bandiera et al., 2010) as well as cooperation and pro-social behaviour in groups (Fershtman and Gneezy, 2001; Grimm and Mengel, 2009; Branas-Garza et al., 2010; Currarini and Mengel, 2016). One outcome that has received much less attention in this context is confidence. This is despite the fact that confidence has been shown to be an important outcome explaining gender differences in competitiveness and leadership among young adults (Alan and Ertac, 2019; Alan et al., 2020), intergenerational income mobility (Blanden et al., 2007) and academic performance of students (Golsteyn et al., 2019) among other things.[1]

In this paper we focus on how matching affects confidence. We design a lab experiment that allows us to identify the effect of being matched with others of either similar or dissimilar performance (assortative or disassortative matching) on people's confidence in their own ability. Across a variety of tasks we find that assortative matching does not have a substantial nor statistically significant effect on confidence compared to a control group with random matching. By contrast, disassortative matching has a negative effect on confidence on average that is driven by the bottom half of performers. This group becomes substantially less confident compared to random matching. However they become also more accurate, i.e. less overconfident, with disassortative compared to random matching. We also find that participants react more strongly to negative than to positive feedback on average, but there is some heterogeneity across tasks with this finding.

These results are relevant for educational tracking within schools, matching students across schools, matching workers in teams and the selection of peers more generally (Golsteyn et al., 2019). They are particularly relevant in contexts, such as education, where confidence is considered an important outcome (Ytterberg et al., 1998; Anderson et al., 2012). To the extent that our results have external validity in the

[1]Heckman et al. (2006) highlight the importance of non-cognitive skills more generally.

specific domain considered, they suggest actionable consequences for institutional design.

The paper is organized as follows. Section 2 discusses related literature. Section 3 contains details of our experimental design. Section 4 contains our main results and Section 5 concludes. A series of Online Appendices contain experimental instructions, screenshots and additional tables and figures.

2 Literature

Our paper contributes to two so far largely disjoint strands of literature. First we contribute to an active literature on how matching affects a variety of different outcomes (usually not including confidence). Second we contribute to literature aimed at understanding the various ways in which (over-) confident beliefs arise in a dynamic setting.

There is a substantial literature on how matching affects outcomes including academic performance of students (Sacerdote, 2001; Hanushek and Woessmann, 2006; Feld and Zoelitz, 2017), the performance of workers and work teams (Mas and Moretti, 2009; Jackson and Bruegmann, 2009; Bandiera et al., 2010) as well as cooperation and pro-social behaviour in groups (Fershtman and Gneezy, 2001; Grimm and Mengel, 2009; Branas-Garza et al., 2010; Currarini and Mengel, 2016). Apart from the different outcomes studied, this literature also differs by whether matching is exogenous or endogenous and it includes a large literature on peer effects. Our paper is more closely related to papers where matching, as in our case, is exogenous. Those include Feld and Zoelitz (2017) who show that university student's performance depends on the performance of other students in their class in a non-monotonic way or Mas and Moretti (2009) who show that the productivity of supermarket cashiers depends on which other cashiers work on the same shift. Hanushek and Woessmann (2006) show that early educational tracking in schools increases inequality. Their analysis also suggests that early tracking might reduce mean performance.[2] Sacerdote (2014) reviews some of the large literature on peer effects.

Our main contribution to this literature is to consider confidence as a key outcome variable. There are only very few papers on matching that include confidence as an outcome. Those can mostly be

[2]Tracking usually includes assortative matching, but also a variety of other measures (differing teaching materials etc). Hence, studying the effect of tracking does not usually identify the pure effect of matching.

found in the peer effects literature (see for example Antonio, 2004). These papers differ from ours in that they consider the influence of endogenously selected friendship groups as opposed to assortative or disassortative matching based on performance.

Our paper also contributes to literature aimed at understanding the various ways in which (over-) confident beliefs come about. In standard economic models, beliefs matter only through their instrumental value in decision making. Recent theoretical work has relaxed this assumption and assumed that people could gain direct utility by holding optimistic beliefs on qualities that are relevant for the self (Köszegi, 2006). When beliefs entail a direct utility, belief updating can depart from the Bayesian benchmark towards optimistic updating in a self-serving way.

There is a growing experimental literature on asymmetric updating, and the results are mixed. Mobius et al. (2011) and Möbius et al. (2007) find that subjects who receive positive feedback in an IQ test revise their beliefs significantly more than those who receive negative feedback. Eil and Rao (2011) also find that subjects asymmetrically update their beliefs on intelligence, and physical attractiveness. They adhere quite closely to the Bayesian benchmark in case of positive signals, but they discount or ignore the signal when it is negative. In Zimmermann (2019), subjects perform an IQ test and receive feedback. He finds little evidence for asymmetry in the short run, but subjects recall negative feedback with lower accuracy one month after receiving the feedback. Sharot et al. (2011) find that subjects updated their beliefs more in response to information that was better than expected than to information that was worse than expected. Other authors find evidence of asymmetric updating in the opposite direction - negative signals weighted more than positive ones. Ertac (2011) studies belief updating across tasks with different degrees of self-relevance. The results of her study indicate that subjects attribute more weight to negative signals than positive ones in the self-relevant context but not in the neutral one. Coutts (2019) also examines whether updating differs across ego-relevant and neutral contexts. His results show that negative signals receive more weight than positive ones but these deviations do not differ across contexts. Our main contribution to this literature is to study how matching affects belief updating and in particular confidence and accuracy of beliefs across tasks with different degrees of ego-relevance and prior strength.

3 Design

In this section we describe the experimental design and procedures. Our experiment consists of a 3×3 design where we vary the type of task and the type of matching. In all treatments participants go through the following stages: introduction, experimental task and belief updating. We will describe these stages in turn.

Introduction stage In this stage, we show participants information about the task they will be asked to complete in the next stage, and describe how their earnings from the task stage are going to be calculated. At the end of this introduction, participants are asked to answer two questions regarding their prior beliefs. In particular, they are asked to state (i) what they believe the average score is going to be in that session and (ii) where they believe they rank among the participants of that session.[3]

Task stage In the **task stage**, participants perform either one of the following tasks (i) observing pairs of contemporary paintings on the screen and guessing which painting received a higher price at an auction (ART), (ii) solving the Raven matrices task (IQ) or (iii) observing a pair of footballers on the screen and guessing which one of them scored more goals during a specified season (FOOT). We now describe details of each task.

The ART task consists in comparing pairs of paintings sold at a real auction.[4] The pictures of the paintings, the title of the work of art, and the author appear on the computer screen and participants are asked to indicate which painting was sold at the higher price. The task involves the comparison of 15 different pairs of paintings. The score is calculated as the total number of correct answers across the 15 pairs.

The IQ task consists in Raven's progressive matrices task, which is a 60-item test used in measuring abstract reasoning and regarded as a non-verbal estimate of fluid intelligence. For this task, one of the test questions is used as an example during the Introduction stage, and the task involves the 59 remaining questions. All of the questions on the Raven's test consist of visual geometric design with a missing piece.

[3]Each experimental session had 16 participants.

[4]Impressionist & Modern Art Evening Sale conducted by Christie's London on February 2, 2016.

The test taker is given six to eight choices to pick from and fill in the missing piece. Participants have 10 minutes to complete as many questions as they can. The score is calculated as the total number of correct answers given within the time limit.

The FOOT task consists in comparing the number of goals scored by two football players of the Turkish Super League. The pictures of the players, and the team they played in the season 2017-2018 appear on the computer screen and participants are asked to indicate which player scored more goals in that season. The task involves the comparison of 15 different pairs of players. The score is calculated as the total number of correct answers across the 15 pairs.

The different tasks were chosen to be able to detect if main results were driven by task-specific properties. They primarily present variation along two dimensions: (i) ego-relevance (Ertac, 2011; Coutts, 2019) and (ii) the strength of the prior about one's rank. Our assumption is that - on average - ART has both low ego-relevance and a weak prior. IQ has high ego-relevance and a strong prior and FOOT has high ego-relevance for some, and low for others and a strong prior for both groups.

Belief revision stage In the **belief revision stage**, participants are first shown their score and then asked a series of questions regarding their rank among the 16 participants in a session. Here, participants observe 8 subgroups (rank 1 or 2, rank 3 or 4,...rank 15 or 16) and for each subgroup, they specify the probability that their actual rank falls in that subgroup. In Step 2, participants receive information on the score of another participant from the session; in the assortative matching condition (*Assortative*), the score is that of a participant who is ranked similarly to them. Specifically participants with rank 1-8 are shown the score of participant ranked +1 (1 observes 2, 2 observes 3, ... , 8 observes 9) while participants with rank 9-16 are shown the score of participant ranked -1 (9 observes 8, 10 observes 9, ... , 16 observes 15). In the disassortative matching treatment (*Disassortative*), the score is that of a participant who is ranked differently to them. Specifically, participants ranked 1 to 8 are shown the score of participant ranked 16 and participants ranked 9 to 16 observe the participant ranked first. In the random condition (*Random*), they are shown the score of a randomly selected participant. Participants are then asked to state their beliefs regarding their ranking, by specifying the probabilities for 8 subgroups as in the previous step.

In subsequent steps, we disclose pieces of information while keeping track of how participants update their beliefs on their ranking. In Step 3 we reveal to each participant in which half of the distribution they performed and we ask them to guess in which one of the four remaining subgroups they think they performed. In Step 4 we reveal in which quarter of the score distribution they rank and we ask to guess in which of the two subgroups of that specific quarter they think they performed. Finally (Step 5), we tell participants in which subgroup they performed, and we ask them to guess their exact ranking. After this last step the true rank of the participant is revealed.

		Task		
		ART	IQ	FOOT
Matching	Random (C)	64	64	64
	Assortative (AT)	64	64	64
	Disassortative (DAT)	63	64	64

Table 1: Number of observations. One participant had to be dropped in ART-DAT as he had already participated in a prior session.

Payments The experiment is incentivized based on the performance in the **Task stage** and the accuracy of one of the guesses in the **Belief revision stage**. For the **Task stage**, participants are paid 1 TL for each correct answer in the ART task, 0.5 TL for each correct answer in the IQ task and 2 TL for each correct answer in the FOOT task.[5] For the **Belief revision stage**, one of the 5 steps where participants stated their beliefs was randomly selected and participants were paid based on their accuracy for this step. In particular, we used a logarithmic rule and the earnings are calculated according to the formula $32 + 32\log(p)$. Here, p is the number we find by dividing the probability that the participant assigns to the subgroup that contains her actual ranking. Since $p \in [0, 1]$, $\log(p)$ is a zero or negative number, the earnings here could be at most 32 TL from this part.[6] While novel, the scoring rule we used here is a proper scoring rule, that is, it incentivizes the participants to state their true beliefs. To see this, suppose there are two categories, *high & low*, and the participant's belief that her rank is *high* is b. When her reported belief is p, her expected payoff becomes $32(1 + blog(p) + (1 - b)log(1 - p))$. Given b, this is

[5]We chose to pay a lower amount for each correct answer in the IQ task due to higher number of questions in that task. On the other hand, the FOOT task pays a relatively higher amount because these sessions were conducted later and a there was a substantial drop in the value of Turkish Lira during the six month period between the first and last session.

[6]If the participant stated a belief of $p \leq 0.1$ for the subgroup where her rank is, then her earnings were rounded to 0. This was mainly to prevent negative earnings from this stage and was specified in the instructions, as well.

a concave function of p and the first order conditions for maximization implies $(1 - p)b + p(1 - b) = 0$, hence $p = b$. The rule and payoff consequences were explained in detail in the Experimental Instructions (see Appendix B).

Questionnaire At the end of the experiment, participants are asked to complete a post-experimental questionnaire eliciting a number of demographics as well as measures of risk attitude and trust. The full list of questions can be found in Appendix C.

Other Details The experiment was conducted at the experimental lab at Bogazici University between November 2018 (ART task) and May 2019 (FOOT task). 575 students participated in our experiment (64 per treatment).[7] Ethical approval was obtained in March 2018 by the Faculty Ethics Committee of the University of Essex (under Annex B). The experiment was programmed using the software z-tree (Fischbacher, 2007) and we used ORSEE to recruit subjects (Greiner, 2004).

4 Main Results

In this section we present our main results. We ask how matching affects confidence (Section 4.1), the accuracy of guesses (Section 4.2) and when it leads to affects over- or under-confidence (Section 4.3). Section 4.4 is dedicated to studying the dynamics of confidence over time and in Section 4.5 we discuss potential mechanisms.

4.1 Confidence

The left panel of Table 2 shows participants' average guessed rank after treatment, i.e. after they observed the score of their match. The figure shows that in all treatments participants are on average somewhat overconfident. The average guessed rank is below the actual mean of 8.5. People seem more confident in their ability in terms of the IQ and FOOT tasks compared to the ART task. It should also be noted,

[7]One participant had to be dropped ex post as it was found out that he had participated twice in the experiment.

though, that there is a substantial amount of heterogeneity in these guesses as illustrated by Appendix Figure E2.

Guessed Rank.

	Assortative	*Random*	*Disassortative*
ART	8.12	8.26	8.90
IQ	6.38	6.74	7.00
FOOT	7.01	6.62	7.64

Absolute Error.

	Assortative	*Random*	*Disassortative*
ART	1.68	1.54	1.23
IQ	1.76	1.86	1.59
FOOT	1.56	1.60	1.54

Table 2: Left: Average guessed rank after treatment (i.e., step 2 of the belief revision stage). Right: Absolute Errors (absolute difference between guessed and actual rank after treatment).

	Guessed Rank					
	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample		Rank< 9		Rank≥ 8	
Assortative	-0.183	-0.198	-0.121	-0.122	-0.117	-0.105
	(0.383)	(0.377)	(0.433)	(0.413)	(0.417)	(0.430)
Disassortative	0.560	0.560	-0.049	-0.057	1.277**	1.289**
	(0.424)	(0.414)	(0.423)	(0.428)	(0.477)	(0.476)
Constant	10.26**	10.67***	8.632***	9.251***	8.663***	8.667***
	(1.194)	(1.241)	(1.587)	(1.809)	(1.569)	(1.558)
Controls	YES	YES^{+}	YES	YES^{+}	YES	YES^{+}
Observations	575	575	287	287	288	288
R-squared	0.086	0.099	0.130	0.152	0.079	0.080

Table 3: OLS regression of Guessed Rank at step 2 of the belief revision stage on treatment dummies. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. The large set additionally controls for measures of risk aversion and trust, whether they are economics students and linearly for how many friends they have. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

To understand how confidence is affected by matching we compare our different treatments. Table 3 shows regression results showing how confidence is affected by the treatment at step 2 of the belief revision stage. For the entire sample (columns (1) and (2)) we do not see a statistically significant average treatment effect. While on average confidence tends to increase in the *Assortative* treatments (i.e. the guessed rank decreases) and decrease in the *Disassortative* treatments (i.e. the guessed rank increases), both of these effects are very imprecisely estimated. More interesting is how matching affects confidence differentially for those in the upper and lower half of the distribution. For those with a good (i.e. low) rank (columns (3) and (4)) the effect of matching on their guessed rank is close to zero and statistically insignificant. By contrast those with a high rank (columns (5) and (6)) suffer a substantial decrease in confidence (of more than one position) in *Disassortative* compared to the control condition. Disassortative matching hence seems to have a strong effect on those in bottom half of distribution. Their confidence is, as expected, lowered. The effect persists after receiving one more piece of information about one's rank but disappears after further pieces of information are revealed (Appendix Table D1).

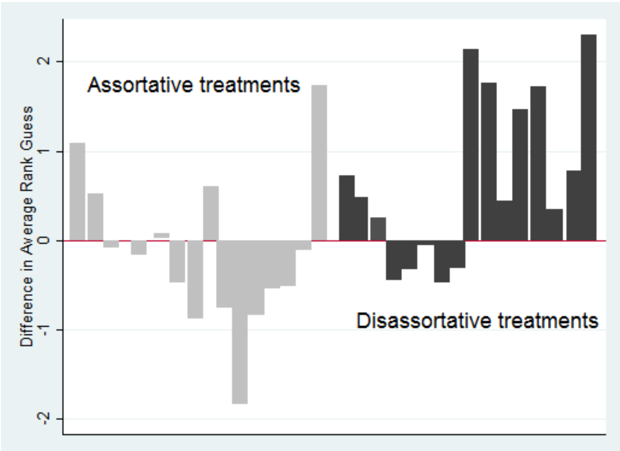


Figure 1: Difference between average rank guess in *Assortative* treatments (light gray bars) and control as well as between *Disassortative* treatments (dark gray bars) and control depending on participant's rank (1-16) at step 2 of the belief revision stage. Positive numbers indicate that participants are *less* confident than in the control (i.e. indicate a bigger rank) and negative numbers that they are *more* confident (i.e. indicate a smaller rank).

Heterogeneity Figure 1 shows this effect across the distribution of ranks. The figure shows that *Disassortative* matching almost never leads to substantial increases in confidence across this distribution, while in *Assortative* confidence tends to increase for those in the bottom half of the distribution. These treatment effects do not differ substantially neither by task nor gender (Appendix Tables D2 and D3).

4.2 Accuracy

We have seen that worse-performing participants become less confident with disassortative matching, but do they become more accurate? The right panel of Figure 1 shows averages of the variable "absolute error" defined as the absolute difference between the guessed and the actual rank of a participant. The panel shows that participants guesses are fairly accurate on average. In the control group participants guesses differ on average between 1.5 to 1.8 ranks (depending on the task) from their true rank. They seem to make somewhat smaller errors under *Disassortative* matching while there doesn't seem to be a consistent difference between the control group and the *Assortative* treatment.

	Absolute Error					
	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample		Rank< 9		Rank> 8	
Assortative	-0.020	-0.024	0.015	0.013	-0.039	-0.038
	(0.091)	(0.093)	(0.105)	(0.095)	(0.201)	(0.208)
Disassortative	-0.241**	-0.238**	0.144	0.143	-0.602***	-0.595***
	(0.098)	(0.105)	(0.097)	(0.085)	(0.197)	(0.208)
Constant	2.308***	2.367***	1.345**	1.158*	2.157**	2.199**
	(0.733)	(0.728)	(0.609)	(0.604)	(0.904)	(0.945)
Controls	YES	YES+	YES	YES+	YES	YES+
Observations	575	575	287	287	288	288
R-squared	0.014	0.029	0.033	0.063	0.061	0.068

Table 4: OLS regression of absolute error (absolute difference between guessed and actual rank) at step 2 of the belief revision task on treatment dummies. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. The large set additionally controls for measures of risk aversion and trust, whether they are economics students and linearly for how many friends they have. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4 shows regression results using absolute error as endogenous variable. The table shows that there is virtually no difference between the *Assortative* treatment and the control group neither for the high, nor for the low performers. In the *Disassortative* treatment by contrast participants make substantially smaller errors (by around 10%), i.e. the difference between guessed and actual ranks is smaller in this treatment. The table shows that this effect is driven by those with lower performance (columns (5)-(6)) for whom we observe an about 27% decrease compared to the control group.

Heterogeneity Splitting the sample by task reveals that the effect is driven by participants in the ART and IQ tasks (Appendix Table D6) with the effect size being substantially smaller in the FOOT task (F-test, p = 0.067). Worse performing women become more accurate in the *Disassortative* treatment (compared to men), while well performing men, but not women, become less accurate in this treatment (Appendix Table D5). Neither of these differences is statistically significant, though (F-test, $p > 0.197$).

We have seen that worse performers do not only become less confident, but also more accurate in the *Disassortative* treatments. This suggests that there was over-confidence in this group to start with. In the next subsection we will study effects on over- and under- confidence explicitly.

4.3 Over- or Under- confidence?

To study overconfidence we focus on two outcome variables: (i) the average amount by which participants underestimate their rank and (ii) the share of participants who believe they are ranked better than they actually are.

Table 5 shows results for the first measure. The table shows again that it is mostly the *Disassortative* treatment which has a significant treatment effect. Again the effect seems to operate mostly on the worse performers who, as anticipated, become less overconfident.

	Overconfidence					
	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample		Rank< 9		Rank> 8	
Assortative	-0.008	-0.013	0.017	0.008	0.036	0.039
	(0.182)	(0.190)	(0.212)	(0.212)	(0.215)	(0.223)
Disassortative	-0.345*	-0.346	0.008	0.001	-0.676***	-0.672**
	(0.202)	(0.208)	(0.207)	(0.209)	(0.244)	(0.255)
Constant	1.980**	2.057*	-0.802	-0.916	2.442***	2.418**
	(0.973)	(1.038)	(0.916)	(0.956)	(0.879)	(0.937)
Controls	YES	YES+	YES	YES+	YES	YES+
Observations	575	575	287	287	288	288
R-squared	0.015	0.025	0.052	0.064	0.068	0.074

Table 5: OLS regression of overconfidence (difference between actual and guessed rank) at step 2 of the belief revision stage on treatment dummies. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. The large set additionally controls for measures of risk aversion and trust, whether they are economics students and linearly for how many friends they have. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

We then turn our attention to the share of participants who believe they are ranked better than they actually are, i.e. the share of overconfident participants. Table 6 shows regressions where the endogenous variable is a dummy indicating whether a participant is overconfident. In line with the previous results we find that the share of overconfident participants decreases in the *Disassortative* treatment in the group of the worse performers. With this measure we also see for the first time an effect of the *Assortative* treatment. Among the worse performers the share of overconfident participants increases in this case. The effect size, however, is smaller (about half of the effect size of the *Disassortative* treatment).

	Overconfidence Dummy					
	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample		Rank< 9		Rank> 8	
Assortative	-0.010	-0.013	-0.059	-0.064	0.060**	0.060**
	(0.036)	(0.037)	(0.059)	(0.060)	(0.026)	(0.027)
Disassortative	-0.048	-0.049	0.024	0.020	-0.109**	-0.112**
	(0.051)	(0.051)	(0.065)	(0.065)	(0.053)	(0.055)
Constant	0.748**	0.724**	0.173	0.141	0.820***	0.746***
	(0.308)	(0.318)	(0.408)	(0.416)	(0.200)	(0.221)
Controls	YES	YES+	YES	YES+	YES	YES+
Observations	575	575	287	287	288	288
R-squared	0.010	0.024	0.020	0.036	0.069	0.080

Table 6: OLS regression of dummy indicating whether a participant is overconfident on treatment dummies at step 2 of the belief revision stage. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. The large set additionally controls for measures of risk aversion and trust, whether they are economics students and linearly for how many friends they have. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

To summarize, matching seems to predominantly affect worse performers. Disassortative matching reduces their confidence and the share of overconfident participants in this group and, as a consequence, makes their perceptions of their own rank more accurate. Assortative matching slightly increases the share of overconfident participants in this group. Whenever we see a treatment effect the direction of effect is in line with what we would expect from rational (Bayesian) learning. What is somewhat puzzling, though, is why it is predominantly the *Disassortative* treatment that affects learning and why assortative matching does not seem to differ much from random matching in terms of its effects on beliefs. Section 4.5 will focus on the mechanisms behind our results to shed some light on these questions. Before we do so, we study in some more detail the dynamics of belief revision across the different steps of our experiment.

4.4 Dynamics

The feedback provided in our experiment during the different steps of the belief revision task is highly diversified across participants, due to differences in their scores, their own ranking in the main task and the scores of other participants in their session. In this section, we introduce a normalized measure of confidence for the different steps of the belief revision task. This will allow us to compare the different steps of the revision process and study its dynamics. We start by noting that during all instances where beliefs are elicited, participants specified beliefs for an even number categories. Based on the construction used in Zimmermann (2019), we define normalized confidence as the difference between the probabilities assigned to the upper and lower half of the ranks that these categories represent. For example, a participant whose rank is 5, will learn at step 3 of the belief revision task, that her rank is somewhere between 1 to 8. Here, she is also asked to specify her beliefs for being in one of the 4 subgroups (being ranked 1st or 2nd, 3rd or 4th, 5th or 6th and 7th or 8th). In this instance, normalized confidence is measured as probabilities assigned to first two of these categories (being ranked 1st or 2nd or being ranked 3rd or 4th) minus the probabilities assigned to last two categories (being ranked 5th or 6th or being ranked 7th or 8th). By construction normalized confidence takes a value between -100 and 100 when beliefs are represented as percentages. This measure allows us to compare confidence across the different steps of belief revision in our experiment and to study its dynamics.

Next, using a regression analysis we set out to analyze the determinants of this confidence at different stages of the belief elicitation process. Our results are presented in Table 7 where confidence is regressed on a series of dummy variables. The analysis provides us with the following insights. In Step 1 of the belief revision task, the participant observes only her own score, and not surprisingly this score has a substantial and highly significant effect on confidence, as those scoring lower than average (Score < mean = 1) have lower normalized confidence levels. While not incentivized, participant's prior on the average score from the task and their prior about their own rank in the task also have significant and persistent effects on confidence. In particular, we observe higher normalized confidence for those who have lower than average expectations for the average score in the associated task (Prior for average score < mean = 1), and lower confidence for those who have higher than average, i.e. worse, expectations for their ranking

(Prior for personal rank > mean = 1). These effects are also not surprising, since an expectation of a higher average score among participants or a worse personal rank would both mitigate personal confidence.

	Normalized Confidence				
	(1)	(2)	(3)	(4)	(5)
	Step 1	Step 2	Step 3	Step 4	Step 5
Score < mean	-46.67***	-36.36***	-34.08***	-2.184	3.838
	(5.941)	(7.387)	(6.213)	(4.312)	(5.973)
Prior for average score < mean	15.18***	17.89***	16.62***	13.92***	9.141**
	(5.244)	(4.547)	(3.485)	(3.980)	(3.581)
Prior for personal rank > mean	-38.61***	-25.55***	-28.77***	-20.03***	-17.85***
	(4.364)	(4.116)	(4.690)	(4.207)	(4.251)
IQ	41.66***	38.41***	30.72***	15.53***	9.386***
	(6.073)	(5.011)	(4.717)	(4.580)	(3.324)
FOOT	13.97**	23.69***	25.61***	12.14***	15.91***
	(6.338)	(4.725)	(4.071)	(4.334)	(4.567)
Assortative		-7.187	1.843	-18.61**	-5.007
		(6.578)	(7.537)	(6.893)	(5.943)
Disassortative		-2.913	7.593	-10.19	-14.43***
		(7.191)	(7.538)	(6.922)	(5.120)
Rank > 8		-28.87***	68.01***	14.82	4.009
		(9.468)	(7.078)	(9.621)	(9.983)
Assortative × Rank > 8		13.72	-3.507	16.00	11.06
		(10.72)	(11.76)	(13.63)	(10.22)
Disassortative × Rank > 8		-24.70**	-30.52***	-8.032	10.50
		(11.12)	(11.06)	(10.57)	(9.729)
Constant	55.18***	58.21***	0.0482	19.36***	11.72**
	(5.721)	(6.600)	(6.984)	(5.021)	(5.039)
Observations	575	575	575	575	575
R-squared	0.329	0.414	0.215	0.097	0.079

Table 7: Normalized Confidence regressed on treatment and other dummies across the five steps of belief revision. Standard errors clustered at the session level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

In Step 2 of the belief revision task, where participants learn the score of another participant, we observe a significant negative effect of *Disassortative* treatment for the participants ranked in the bottom 50% (Rank > 8) of the session, as evidenced by the coefficient for "*Disassortative*× Rank > 8". This effect is also significant and has the same sign and similar magnitude at Step 3 of the belief revision task, but not at Steps 4 and 5. On the other hand, the coefficient for "Rank > 8" is positive and significant at Step 3. This observation has an intuitive explanation. At this step, a participant learns whether she is ranked in the bottom half or the upper half of the distribution. Upon learning this news, subjects with moderate beliefs would assign higher weights to categories close to the middle of the overall distribution.

This means subjects learning that they are in the lower 50 % would assign a higher belief to being at the 3rd quartile compared to being at the 4th quartile. On the other hand, subjects learning that they are in the upper 50% would assign a higher belief to to being at the 2nd quartile compared to being at the 1st quartile. These in turn would imply that normalized confidence at Step 3 would be higher for subjects ranked at the bottom 50%. The nature of tasks also seem to affect normalized confidence. In particular, the coefficients for IQ and FOOT are both positive and significant. During the initial stages of belief elicitation, normalized confidence seems to be highest in the IQ task compared to the FOOT and ART task (0.43 vs. 0.64 and 0.63, respectively). As subjects move to the latter steps, the effect sizes become smaller and the coefficients for IQ task and FOOT become more similar. Overall, hence, these results show again that matching has the strongest effect for those in the bottom half of the distribution. They also show that the effect is strongest at Step 2, where the treatment takes place.

4.5 Mechanisms

In this section we dig deeper into our data to gain some insight into the behavioural mechanisms underlying these patterns. Figure 2 shows by how much participants' rank guess changes on average (at Step 2) depending on the difference between their score and the score of the participant they observe (their match). The top left panel shows the entire sample. The figure shows that on average participants become more pessimistic about their rank (i.e. increase their guess) if they have a lower score than their match and become more optimistic (i.e. decrease their guess) if they have a higher score than their match. If they have the same score they become more pessimistic. This figure masks a considerable amount of heterogeneity across tasks. In the ART task (top right panel) and the IQ task (bottom left panel) the pattern is as described above. It can also be seen that participants in these tasks react much more strongly to negative feedback (having a lower score than the match) than to positive feedback. If we accept that the IQ task carries more ego-relevance than the ART task, then we can conclude that ego-relevance is not a crucial mechanism behind this, as both tasks show a very similar pattern. The FOOT task (bottom right panel) shows a different pattern, though. Here participants seem to become slightly more confident after feedback was received irrespective of whether feedback was positive or negative. This could be because participants have - on average - a more pessimistic prior in this task. For the remainder of this section we will aggregate the three tasks.

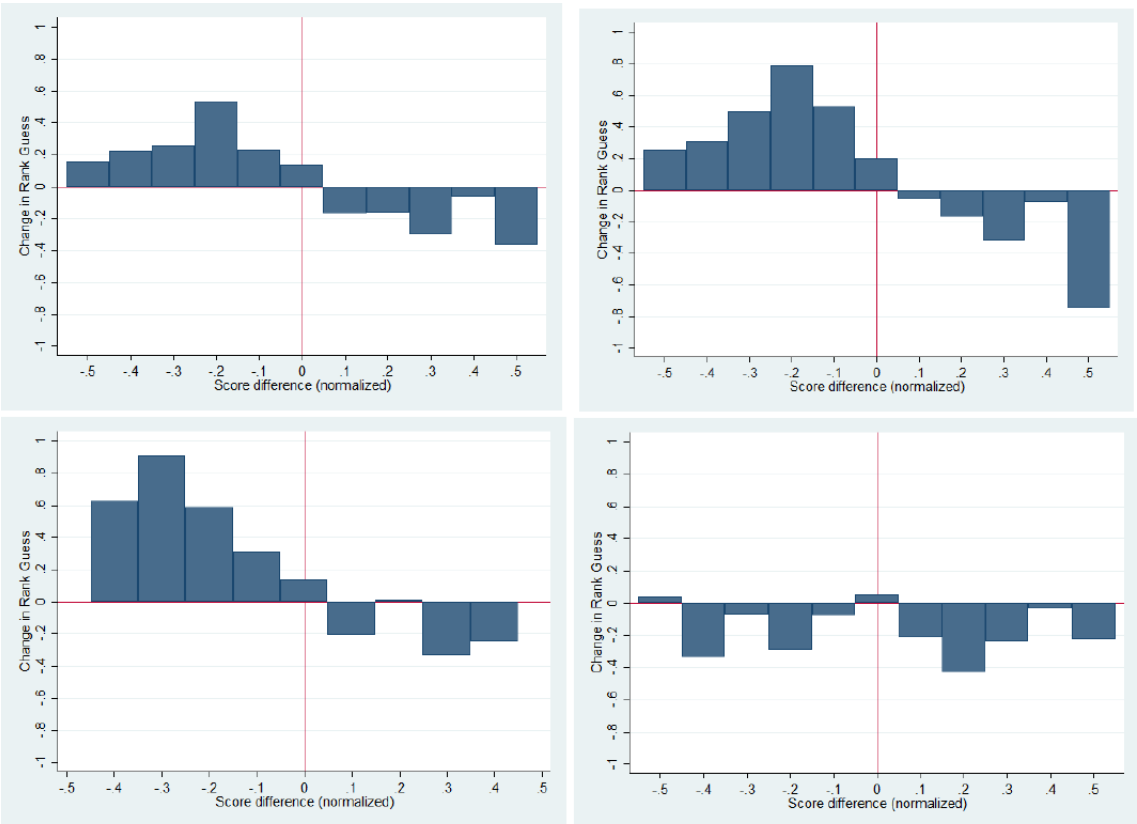


Figure 2: Change in guessed rank depending on difference between own and other participant's score. Top left: entire sample; Top right: ART task; Bottom left: IQ task; Bottom right: FOOT task.

Table 8 shows regression results where we regress the change in confidence on the observed score difference and an indicator for whether the difference is positive. The table shows that the indicator matters even when score difference is included in the regression (columns (3)-(4)). There is a discontinuous jump in participant's reaction to feedback once it changes sign. Figure 3 illustrates the regression results (Panel (b)) and also shows the raw data averages (Panel (a)). The figure illustrates that confidence changes not as much for positive feedback as it does for negative feedback.[8] It also illustrates the discontinuous change at zero.

[8]Note that there is a level shift between panels (a) and (b) which is due to the inclusion of controls in Panel (b).

	(1)	(2)	(3)	(4)
		Change in Guessed Rank		
Δ_{Score}	-1.907***		-0.919	-0.990
	(0.400)		(0.568)	(0.735)
pos		-0.866***	-0.572**	-0.587***
		(0.144)	(0.210)	(0.214)
$\Delta_{Score} \times pos$				0.191
				(1.047)
Constant	1.007	1.104	1.020	1.007
	(0.944)	(0.902)	(0.926)	(0.945)
Observations	576	576	576	576
R-squared	0.074	0.080	0.085	0.085

Table 8: Change in Guessed Rank at step 2 of the belief revision stage regressed on depending on difference between own and other participant's score (Δ_{Score}), a dummy indicating whether this difference is positive (pos) and the interaction between the two. Linear controls included for age and number of siblings as well as gender dummy, indicator for housing situation and task fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

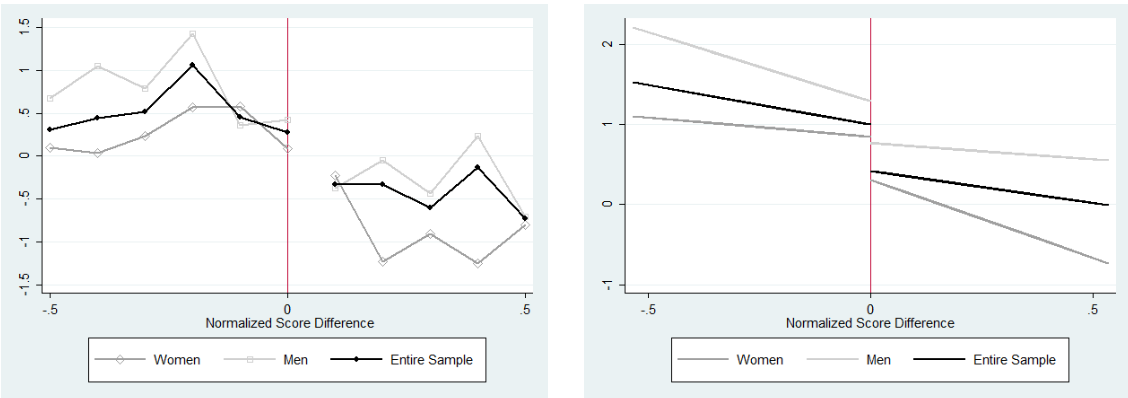


Figure 3: Change in Guessed Rank at step 2 of the belief revision stage depending on the difference between own and other participant's score and on whether that difference is negative or positive. Gray line with diamond markers show women, square markers men and the black line the entire sample. The left panel shows raw data averages and the right panel predicted values from the regression in Table 8 including controls for age, gender, task, siblings and housing situation.

As there is a large psychological literature on gender differences in reacting to feedback and attribution, which shows that (i) women tend to react more strongly to feedback (for a review see Roberts, 1991)[9] and (ii) that women are more likely to blame their ability for their failures (see for instance Dweck et al.,

[9]Interestingly, there are also experimental papers which find that women tend to update their beliefs less strongly than men in response to feedback (Mobius et al., 2011; Coutts, 2019; Buser et al., 2018; Albrecht et al., 2013). Some authors point to gender differences which depend on the valence - good or bad - of the feedback received. Ertac (2011) finds that women that completed a verbal task interpret positive feedback more conservatively than men while no gender difference is found for negative feedback. Berlin and Dargnies (2016) show that women update more pessimistically than their male counterparts after receiving negative feedback but not after positive feedback.

1978; Dweck and Reppucci, 1973; Nicholls, 1975) we also split the figures by gender. The gray lines in Figure 3 show the change in confidence after feedback separately for women and men. In our experiment, women react more strongly to feedback, but only if it is positive feedback. The effect is just outside of statistical significance, though ($p = 0.1016$). There is no (statistically significant) gender difference in the reaction to negative feedback.

Overall these results show that the direction of feedback (positive vs negative) seems more important than the extent, i.e. how different the score of the other person is from one's own score. They also show that people react more when feedback is negative. The latter fact can also explain why disassortative matching has a stronger treatment effect. While both assortative and disassortative matching have an equal share of participants exposed to positive and negative feedback, they differ in how strongly negative or positive the feedback is.[10] The fact that participants - on average - react more strongly to negative feedback means that treatment effects will be stronger under disassortative matching.

5 Conclusion

We conducted a lab experiment to study how matching affects confidence. Across a variety of tasks we find that assortative matching does not have a substantial nor statistically significant effect on confidence compared to a control group with random matching. By contrast, disassortative matching has a negative effect on confidence on average that is driven by the bottom half of performers. This group becomes substantially less confident compared to random matching. However they become also more accurate, i.e. less overconfident, with disassortative compared to random matching.

These are important findings that should be taken into account when designing policies like tracking in schools or matching peers at work. There are, however, also several caveats and open questions for future research. One important question regards the trade-off between overconfidence and accuracy. It seems obvious that there are advantages to holding accurate beliefs, however some also argue that there are benefits from being overconfident (Kahneman and Lovallo, 1993; Anderson et al., 2012; Murphy

[10]Under assortative matching many participants also see the same score as their own, i.e. receive "neutral" or "weakly positive" feedback.

et al., 2015; Radzevick and Moore, 2011). In the school setting, inaccurate self-evaluations have been found to undermine students' learning and retention (Dunlosky and Rawson, 2012) and overconfidence is believed to induce a student to allocate less time to study, resulting in poor exams grades (Bandura, 1993). However, other authors (Magnus and Peresetsky, 2018; Johnson and Fowler, 2011) find that for most students overconfidence is advantageous, possibly because it increases ambition, morale, and persistence. The positive effects of overconfidence are also stressed in innovation research. Different authors associate overconfidence with greater exploration and risk-taking (Galasso and Simcoe, 2011; Bernardo and Welch, 2001) leading to positive outcome in terms of technological development. On the other hand, the performance in simple productivity tasks has been found to benefit from a heterogeneous group composition (Mas and Moretti, 2009). Hence, which level of (over-) confidence to target is not immediately obvious and might depend on the setting. In the classroom context, improving accuracy in self-evaluation through disassortative matching might be combined with interventions that support motivation and ambition for the weakest members. In the corporate setting, group composition might need to be adjusted according to the task to be performed. Furthermore we should remember that confidence is only one outcome affected by matching and obviously the effect on other outcomes has to be taken into account.

References

Alan, S. and S. Ertac (2019). Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment. *Journal of the European Economic Association 17*(4), 1147–1185.

Alan, S., S. Ertac, E. Kubilay, and G. Loranth (2020). Understanding gender differences in leadership. *Economic Journal 130*(626), 263–289.

Albrecht, K., E. Von Essen, J. Parys, and N. Szech (2013). Updating, Self-Confidence and Discrimination. *European Economic Review 60*, 144–169.

Anderson, C., S. Brion, D. A. Moore, and J. A. Kennedy (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology 103*(4), 718—-735.

Antonio, A. (2004). The influence of friendship groups on intellectual self-confidence and educational aspirations in college. *The Journal of Higher Education 75*(4), 446–471.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *Review of Economic Studies 77*, 417–458.

Bandura, A. (1977). Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review 84*(2), 191–215.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational psychologist 28*(2), 117–148.

Berlin, N. and M. P. Dargnies (2016). Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior and Organization 130*, 320–336.

Bernardo, A. E. and I. Welch (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy 10*(3), 301–330.

Blanden, J., P. Gregg, and L. Macmillan (2007). Accounting for intergenerational income persistence: noncognitive skills, ability and education. *The Economic Journal 117*(519), C43–C60.

Buser, T., L. Gerhards, and J. van der Weele (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty 56*(2), 165–192.

Coutts, A. (2019). Good news and bad news are still news: experimental evidence on belief updating. *Experimental Economics 22*(2), 369–395.

Currarini, S. and F. Mengel (2016). Identity, homophily and in-group bias. *European Economic Review 90*, 40–55.

Dunlosky, J. and K. A. Rawson (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction 22*(4), 271–280.

Dweck, C. S., W. Davidson, S. Nelson, and B. Enna (1978). Sex Differences in Learned Helplessness: II. The Contingencies of Evaluative Feedback in the Classroom and III. An Experimental Analysis. *Developmental Psychology 14*(3), 268–76.

Dweck, C. S. and N. D. Reppucci (1973). Learned Helplessness and Reinforcement Responsibility in Children. *Journal of Personality and Social Psychology 25*, 109–116.

Eil, D. and J. M. Rao (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics 3*(2), 114–138.

Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior and Organization 80*(3), 532–545.

Feld, J. and U. Zoelitz (2017). Understanding peer effects - on the nature, estimation and channels of peer effects. *Journal of Labor Economics 35*(2), 387–428.

Fershtman, C. and U. Gneezy (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics 116*(1), 351–377.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics 10*(2), 171–178.

Galasso, A. and T. Simcoe (2011). CEO overconfidence and innovation. *Management Science 57*(8), 1469–1484.

Golsteyn, B., A. Non, and U. Zoelitz (2019). The impact of peer personality on academic achievement. *mimeo*.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen*. GDWG Bericht 63.

Grimm, V. and F. Mengel (2009). Cooperation in viscous populations-Experimental evidence. *Games and Economic Behavior 66*(1), 202–220.

Hanushek, E. A. and L. Woessmann (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal 116*(510), C63–C76.

Heckman, J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Jackson, C. and E. Bruegmann (2009). Teaching students and teaching each other: the importance of peer learning for teachers. *American Economic Journal: Applied Economics 1*(4), 85–108.

Johnson, D. D. P. and J. H. Fowler (2011). The Evolution of Overconfidence. *Nature 477*(7364), 317–320.

Kahneman, D. and D. Lovallo (1993). Timid choices and bold forecasts: A cognitve perspectve on risk and risk taking. *Management Science 39*, 17—31.

Köszegi, B. (2006). Ego Utility, Overconfidence, and Task Choice. *Journal of the European Economic Association 4*(4), 673–707.

Magnus, J. R. and A. A. Peresetsky (2018). Grade expectations: Rationality and overconfidence. *Frontiers in Psychology 8*(JAN), 1–10.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–145.

Matthews, Gerald, M. Zeidner, and R. D. Roberts (2004). *Emotional Intelligence: Science and Myth*. MIT press.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2007). Gender Differences in Incorporating Performance Feedback. *Mimeo, Harvard University 1*, 1–27.

Mobius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2011). Managing self-confidence: Theory and experimental evidence.

Murphy, S. C., W. von Hippel, S. L. Dubbs, M. J. Angilleta Jr, R. S. Wilson, R. Trivers, and F. K. Barlow (2015). The Role of Overconfidence in Romantic Desirability and Competition. *Personality and Social Psychology Bulletin 41*(8), 1036–1052.

Nicholls, J. G. (1975). Causal attributions and other achievement-related cognitions: Effects of task outcome, attainment value and sex. *Journal of Personality and Social Psychology 31*, 379–389.

Radzevick, J. R. and D. Moore (2011). Competng to be certain (but wrong): Social pressure and overprecision in judgment. *Management Science 57*(1), 93—-106.

Roberts, T. A. (1991). The influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin 109*(2), 297–308.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics 116*(2), 681–704.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics 6*(1), 253–272.

Sharot, T., C. W. Korn, and R. J. Dolan (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience 14*(11), 1475–1479.

Ytterberg, S., I. Harris, S. Allen, D. Anderson, P. Kofron, J. Kvasnicka, J. McCord, and J. Moller (1998). Clinical confidence and skills of medical students: use of an OSCE to enhance confidence in clinical skills. *Academic Medicine 73*(10), S103–105.

Zimmermann, F. (2019). The dynamics of motivated beliefs. *American Economic Review 110*(2), 337–361.

Online Appendix

A Appendix: Sample

We perform a balancing check for our treatments. We regress the observables collected in the post-experimental questionnaire on treatment dummies. The systematic presence of significant coefficients in the regression would reveal differences in the pool of participants of *Assortative* and *Disassortative* treatments compared to the random treatment.

We consider the following dependent variables: in column (1) age, in (2) gender (male = 1), in (3) living arrangement of subject (0= dorm; 1= with family; 2= with friends; 3= alone), in (4) number of siblings, (5) risk attitude measured by the question "How willing are you to take risks in general?" (10=high; 1=low), (6) trust measured by the question "Would you say that most people can be trusted?" (1=yes), (7) number of economics classes taken, (8) number of friends among participants in the session.

Table A1 presents the results of the balancing check. Only two coefficients are significant - *Assortative* when regressed on age and *Disassortative* when regressed on siblings - indicating that participants do not systematically differ across treatments. The data support the hypothesis of random assignment to treatments.

Dep variable:	(1) Age	(2) Male	(3) Living	(4) Siblings	(5) Risk	(6) Trust	(7) Econ.	(8) Friends
Assortative	-0.474***	-0.062	0.052	-0.042	-0.125	-0.005	-0.130	0.083
	(0.159)	(0.050)	(0.010)	(0.098)	(0.195)	(0.037)	(0.149)	(0.090)
Disassortative	-0.252	-0.033	-0.021	0.180*	-0.034	0.006	-0.046	0.108
	(0.160)	(0.050)	(0.010)	(0.098)	(0.195)	(0.037)	(0.150)	(0.090)
Observations	575	575	575	575	575	575	575	575
R-squared	0.015	0.003	0.001	0.010	0.001	0.000	0.001	0.003

Table A1: Balancing check. Regression of observables on treatment dummies. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

B Appendix: Instructions

All participants within a treatment received the same instructions. We handed out written instructions at the beginning of each session. In B.1 and B.2, we report written instructions and screenshots for the ART task. For the FOOT and IQ task, the instructions were modified according to the specificity of each task. We provide further details on the FOOT task in B.4. The original instructions were in Turkish. Here, we provide the English translation.

B.1 Written instructions - ART task

General Information

Welcome! You are about to participate in an experiment on decision-making. If you follow the instructions carefully, you can earn a significant amount of money based on your choices.
This instruction set is for your private use only. You cannot communicate with anyone during the experiment. If you have any questions, please raise your hand. Then we'll come and answer your questions. Violation of this rule requires that we immediately exclude you from the experiment.
With the decisions you will make in the experiment, you will earn a profit. Below, we will explain the details of this. All your decisions will be handled confidentially, both during and after the experiment. This means that none of the other participants will know the decisions you make.

Part 1:

In this first part of the experiment we ask you to answer the questions in a test. To perform the test, we will show paintings sold in an auction in February 2016. In addition to the paintings, we present the name of the painting and the painter. You will see something similar to the following images:

Pablo Picasso, Mandoline

Egon Schiele, Oesterreichisches Mäderl

We then ask you to specify which one of these two paintings was sold at a higher price in this auction. To complete the task, you are asked to compare 15 different pairs of paintings. Your test score is calculated as the total number of correct answers in 15 pairs.

The maximum score you can get is 15 (if all your answers are correct). The minimum score is 0 (if all your answers are incorrect). You will earn 1 TL for each correct answer in this section.

Part 2:

We will rank the participants in this room based on your test scores in the previous section. If two participants get the same score, their rankings will be random. In this part of the experiment you are asked to answer some questions about the performance of the other participants in the experiment and the comparison of your performance vs. others performances.

In this section, we'll ask you questions about your beliefs regarding your ranking in the test. We'll ask you to do this a total of 5 times, and give you new information every time. As you know, there are 16 people in the experiment, and your rankings range from 1st to 16th.

Each time, you will be asked to guess how well you think you did the test compared to the rest of the students in the lab. You will do this by specifying your probability estimates for the ranking groups you see on the screen. We ask you to enter numbers from 0 to 100 for the probabilities for each ranking group.

Please note that the odds allocated to the groups you see on the screen should always add up to 100%. For example:

Your rank	1. or 2.	3. or 4.	5. or 6.	7. or 8.	9. or 10.	11. or 12.	13. or 14.	15. or 16.	SUM
Probability	5%	10%	40%	10%	5%	0%	0%	0%	100%

If the total of the odds is not 100%, the computer will inform you and you will need to correct your answer before proceeding to the next stage.

Payment

As we explained above, we will ask you to specify the possibilities for your ranking 5 times in total. One of the 5 predictions you make will be chosen randomly and we will pay you based on the accuracy of that prediction. Payment is made according to the formula $32 + 32\log(p)$. Here, p is the number we find by dividing the probability that you assign to the category that contains your actual ranking. Note that $\log(p)$ is a zero or negative number, and your payment will be between 0 and 32, depending on how accurate the prediction is.

Example: For example, if you scored the best score in the test, that is, if your ranking is 1 and f you correctly guess the probability of being 1st or 2nd as 100%, you earn the highest possible amount. Your payment is 32 TL (calculation: $32 + 32 \times \log(1) = 32$).

Your rank	1. or 2.	3. or 4.	5. or 6.	7. or 8.	9. or 10.	11. or 12.	13. or 14.	15. or 16.	SUM
Probability	100%	0%	0%	0%	0%	0%	0%	0%	100%

However, if your ranking is 1, and you assign a probability lower than 100% for this category, your earnings will be reduced. This reduction is proportional to the inaccuracy of the probability assigned to the correct group.

For example, if your ranking is 1 and you set the probability of being '1st or 2nd' to 90%, your payment will be reduced by 5% and will be $32 + 32 \times \log(1.01) = 32 - 32 \times 0.05 = 30.54$ TL.

Your rank	1. or 2.	3. or 4.	5. or 6.	7. or 8.	9. or 10.	11. or 12.	13. or 14.	15. or 16.	SUM
Probability	90%	10%	0%	0%	0%	0%	0%	0%	100%

If your ranking is 1 and you set the probability of being '1st or 2nd' to 80%, your payment will be reduced by 9.7% and will be $32 + 32 \times \log(0.8) = 32 - 32 \times 0.097 = 28.89$ TL.

If your ranking is 1 and you set the probability of being '1st or 2nd' to 70%, your payment will be reduced by 15.5% and will be $32 + 32 \times \log(0.7) = 32 - 32 \times 0.155 = 27.04$ TL.

If your ranking is 1 and you set the probability of being '1st or 2nd' to 60%, your payment will be reduced by 22.2% and will be $32 + 32 \times \log(0.6) = 32 - 32 \times 0.222 = 24.9$ TL.

If your ranking is 1 and you set the probability of being '1st or 2nd' to 50%, your payment will be reduced by 30.1% and will be $32 + 32 \times \log(0.5) = 32 - 32 \times 0, 301 = 22.37$ TL.

If your ranking is 1 and you set the probability of being '1st or 2nd' to 40%, your payment will be reduced by 39.8% and will be $32 + 32 \times \log(0.4) = 32 - 32 \times 0.398 = 19.27$ TL.

If your ranking is 1 and you set the probability of being '1st or 2nd' to 30%, your payment will be reduced by 52.3% and will be $32 + 32 \times \log(0.3) = 32 - 32 \times 0.523 = 15.27$ TL.

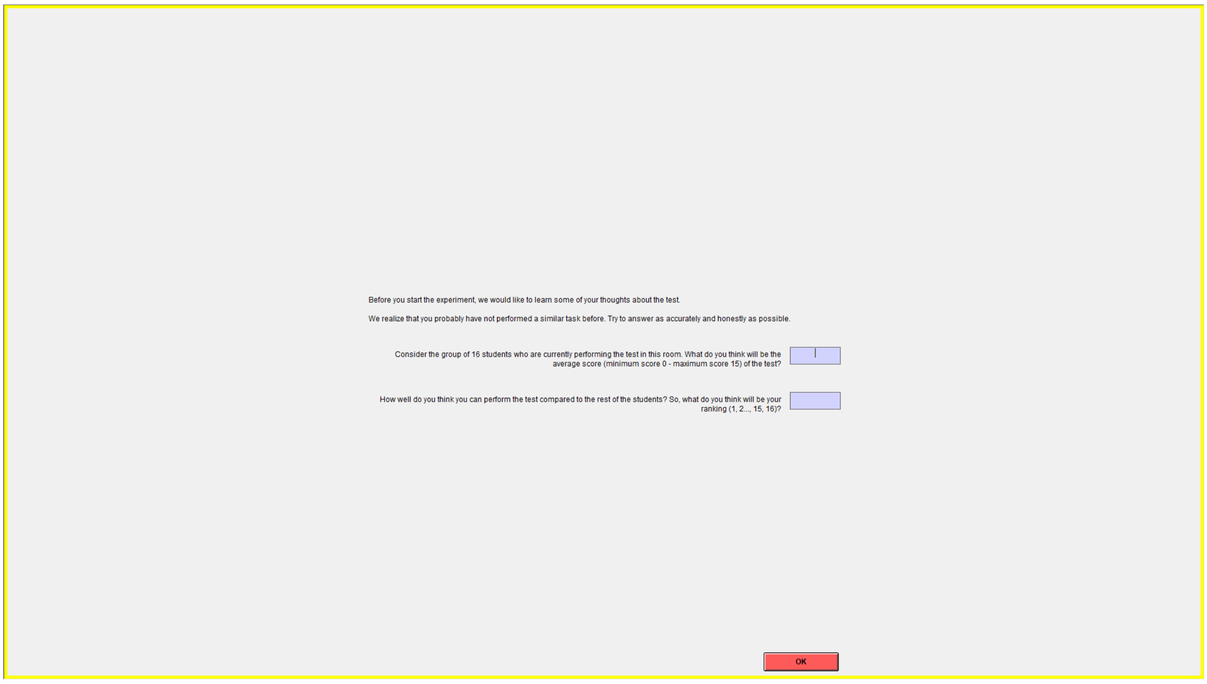
If your ranking is 1 and you set the probability of being '1st or 2nd' to 20%, your payment will be reduced by 69.9% and will be $32 + 32 \times \log(0.2) = 32 - 32 \times 0.699 = 15.27$ TL.

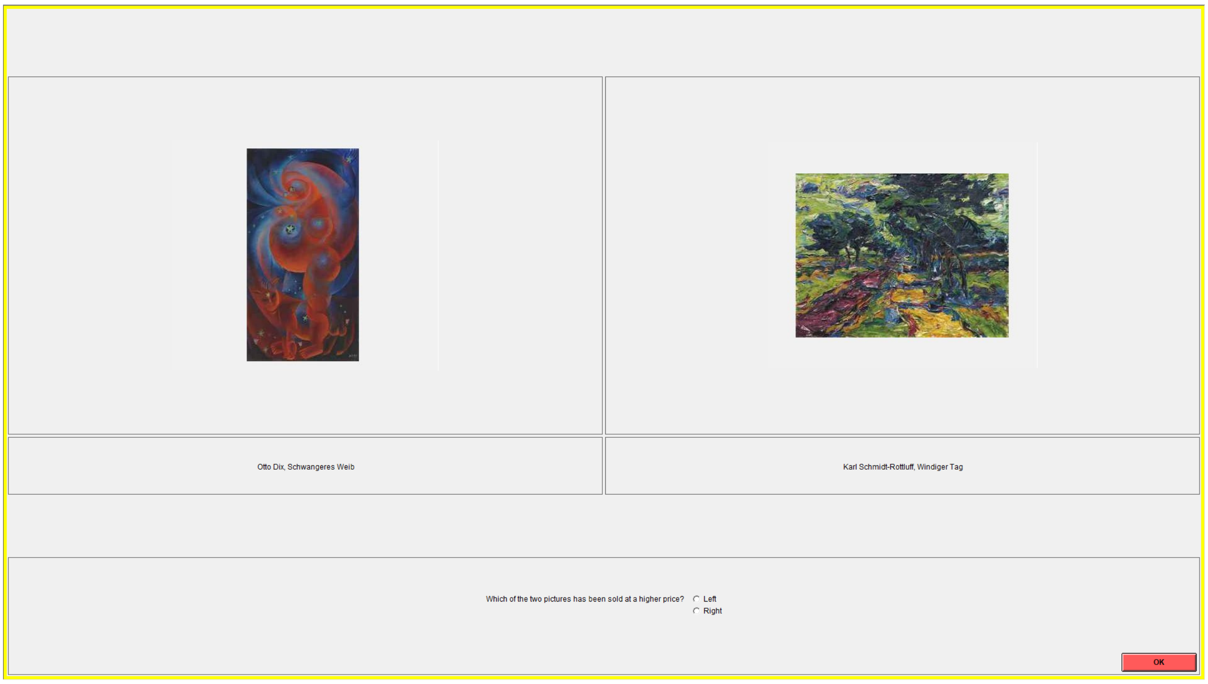
Finally, if your ranking is 1, if you set the probability of being 1 or 2 as 10% or less, your payment will be reduced to 0.

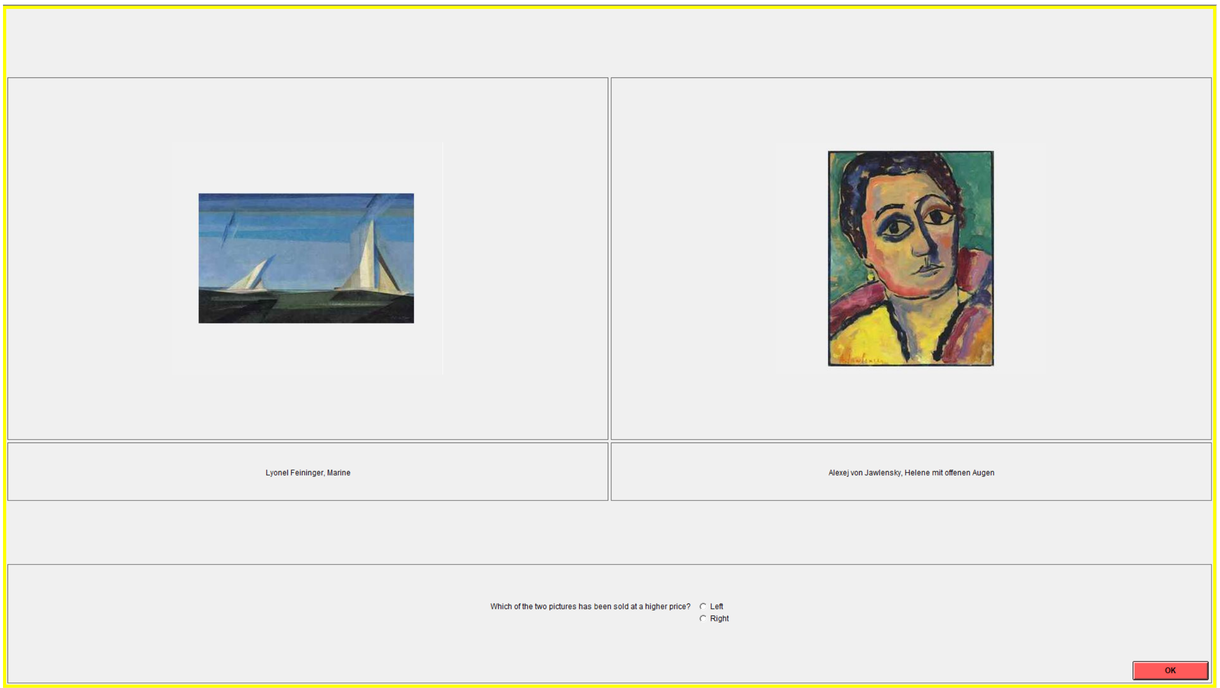
You should try to be as accurate as possible in your predictions. A good estimate will give you the best benefit.

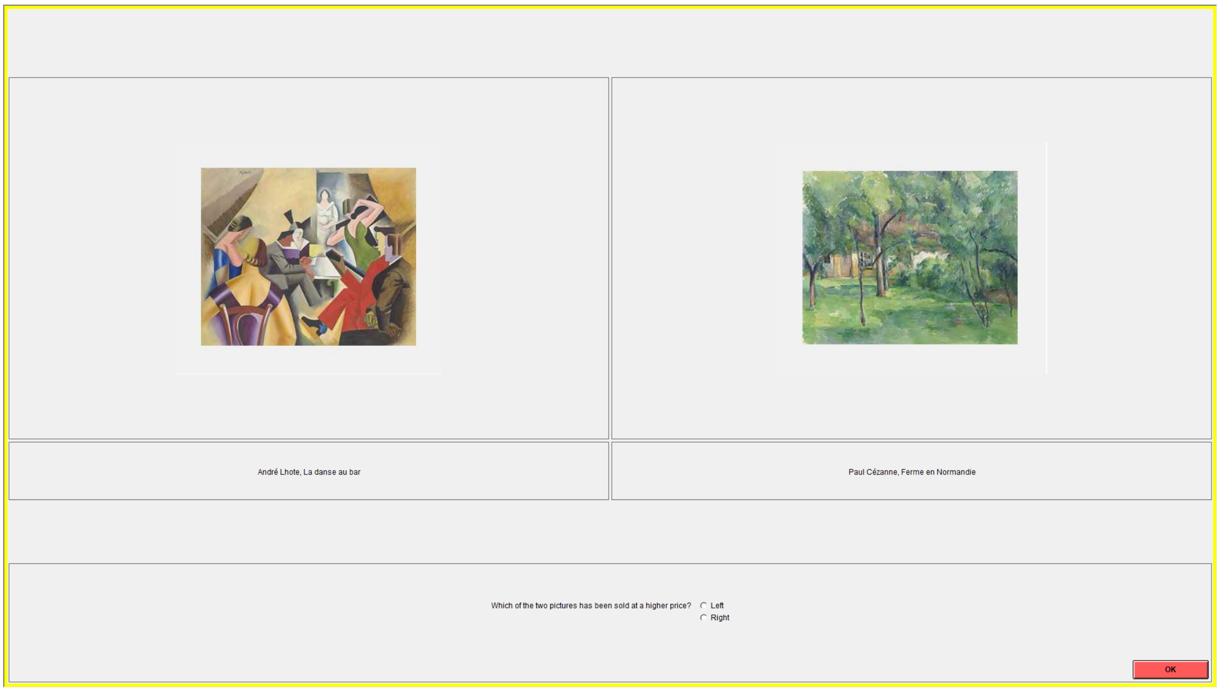
At the end of the experiment, all your winnings will be paid in cash.

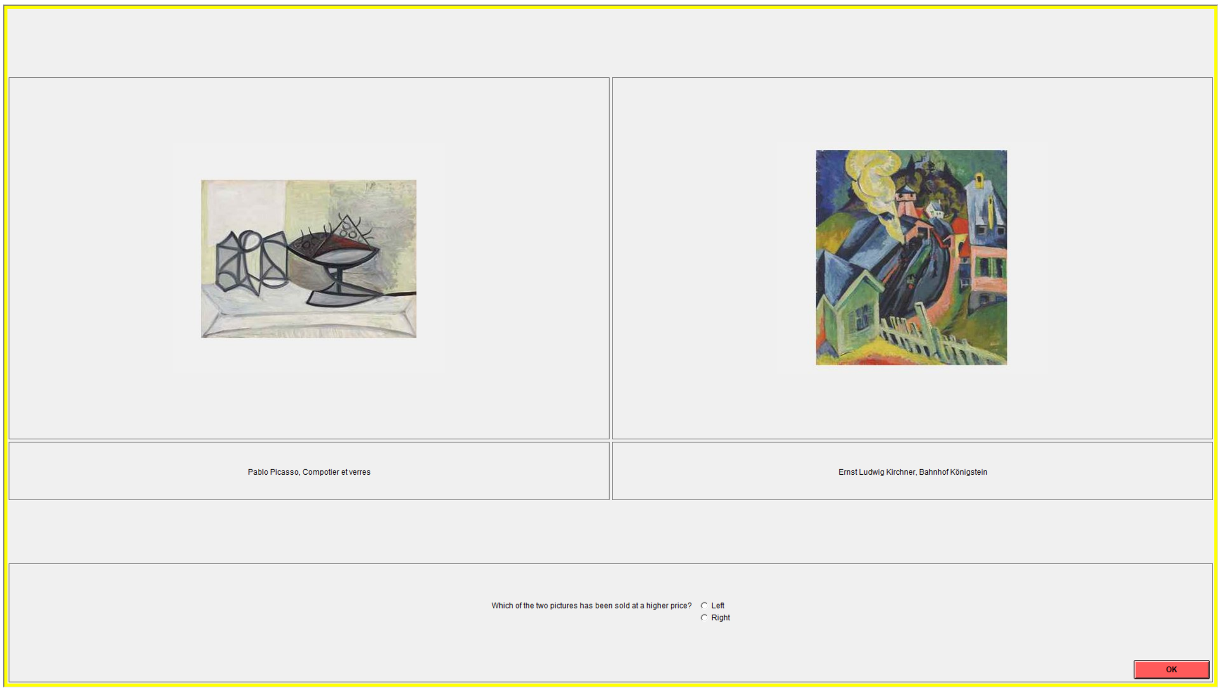
B.2 Screenshots - ART task

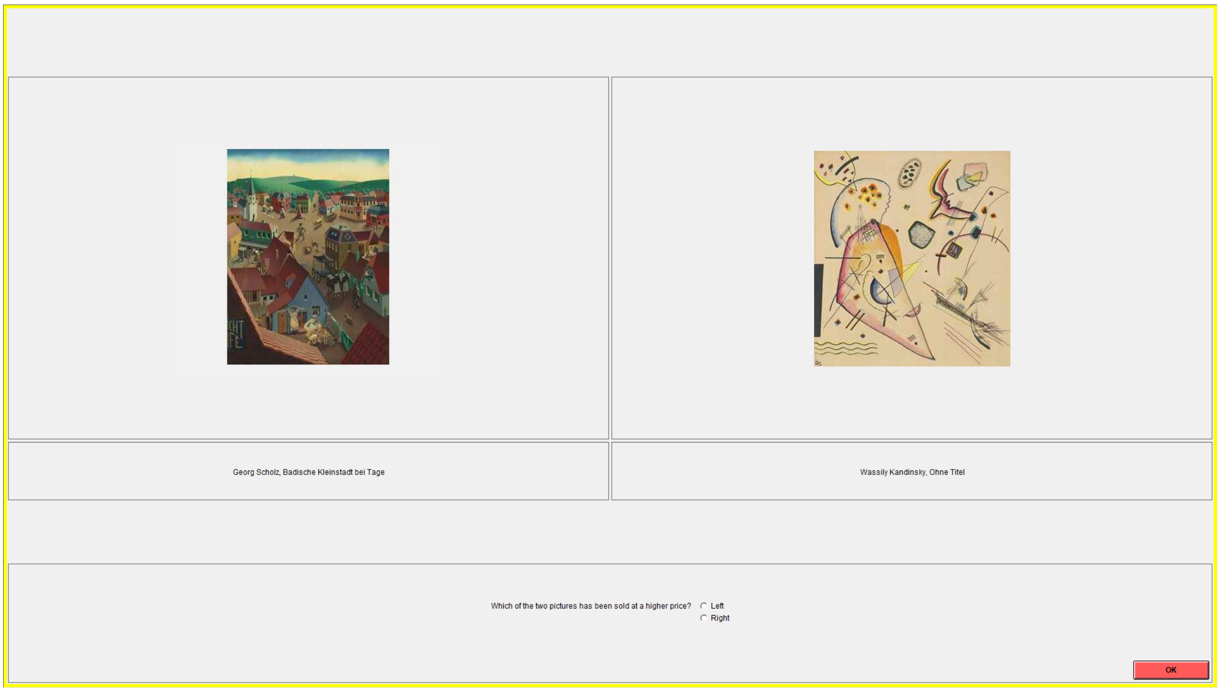


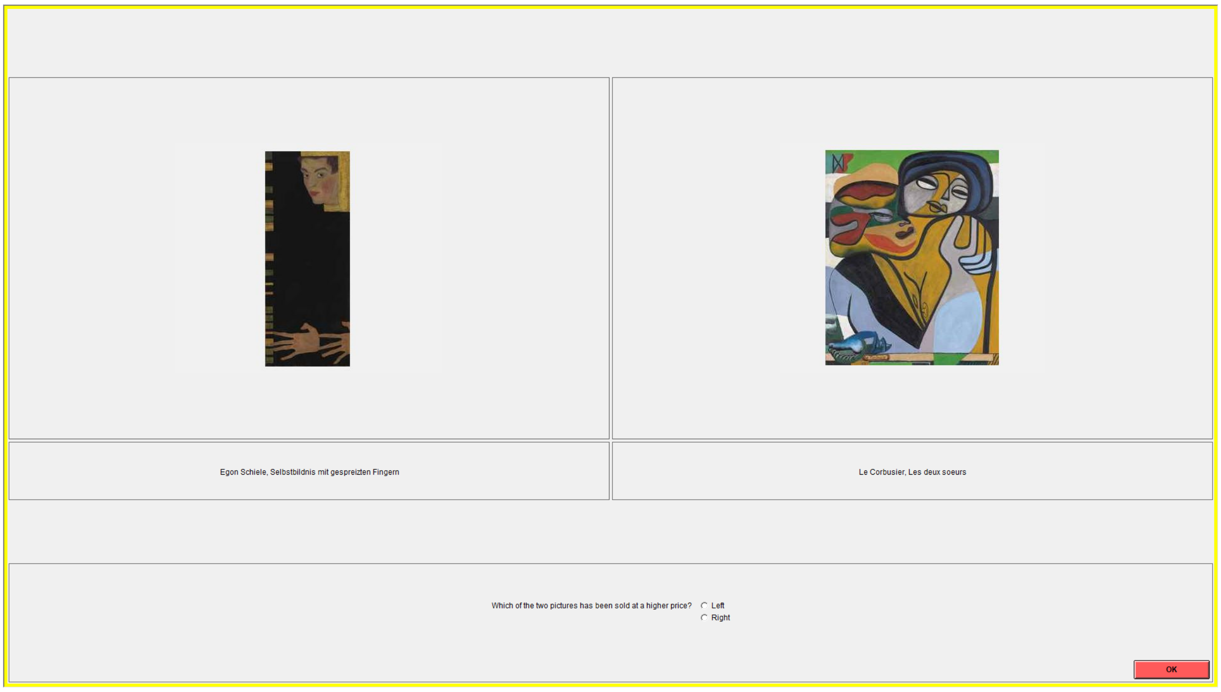


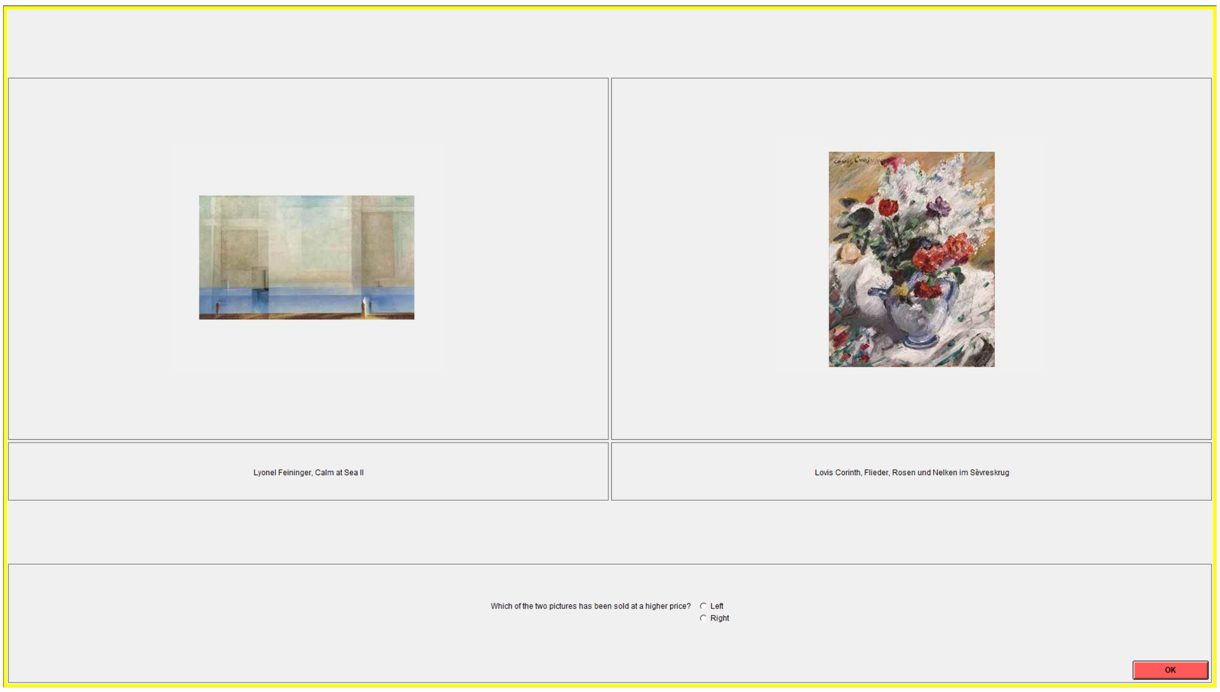

Lyonel Feininger, Marine
Alexej von Jawlensky, Helene mit offenen Augen
Which of the two pictures has been sold at a higher price?
Left
Right
OK

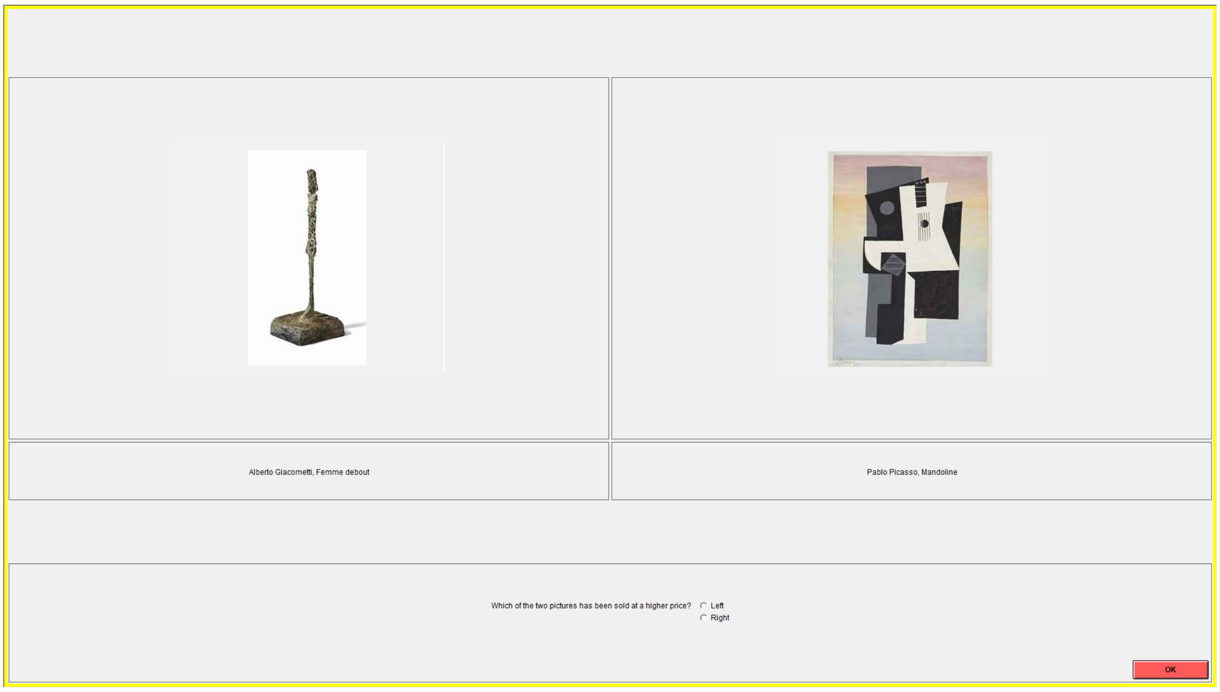

André Lhote, La danse au bar
Paul Cézanne, Ferme en Normandie
Which of the two pictures has been sold at a higher price?
Left
Right
OK

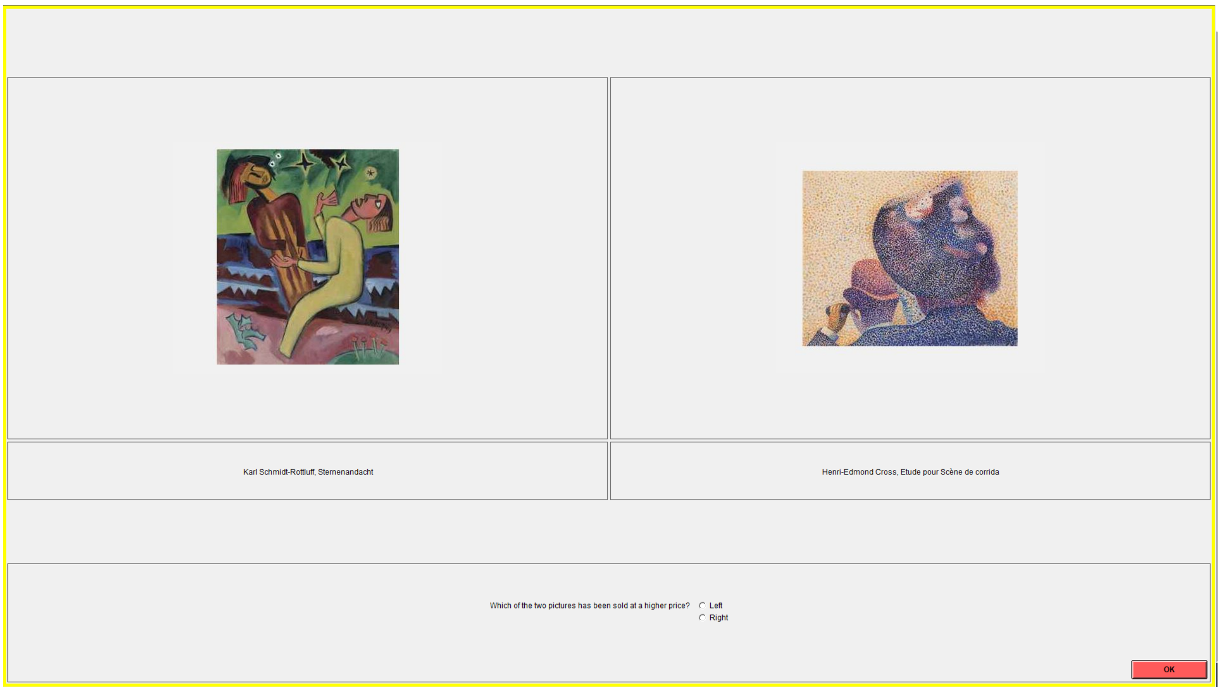

Pablo Picasso, Compotier et verres
Ernst Ludwig Kirchner, Bahnhof Königstein
Which of the two pictures has been sold at a higher price?
Left
Right
OK

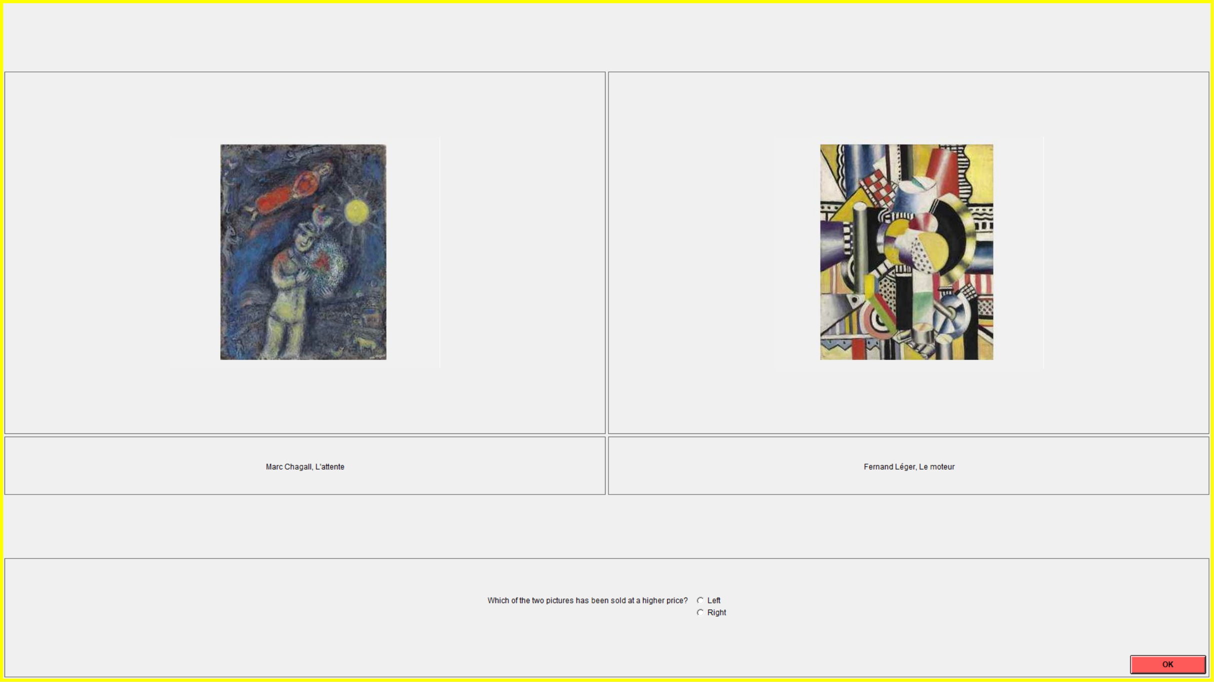

Georg Scholz, Badische Kleinstadt bei Tage
Wassily Kandinsky, Ohne Titel
Which of the two pictures has been sold at a higher price?
Left
Right
OK

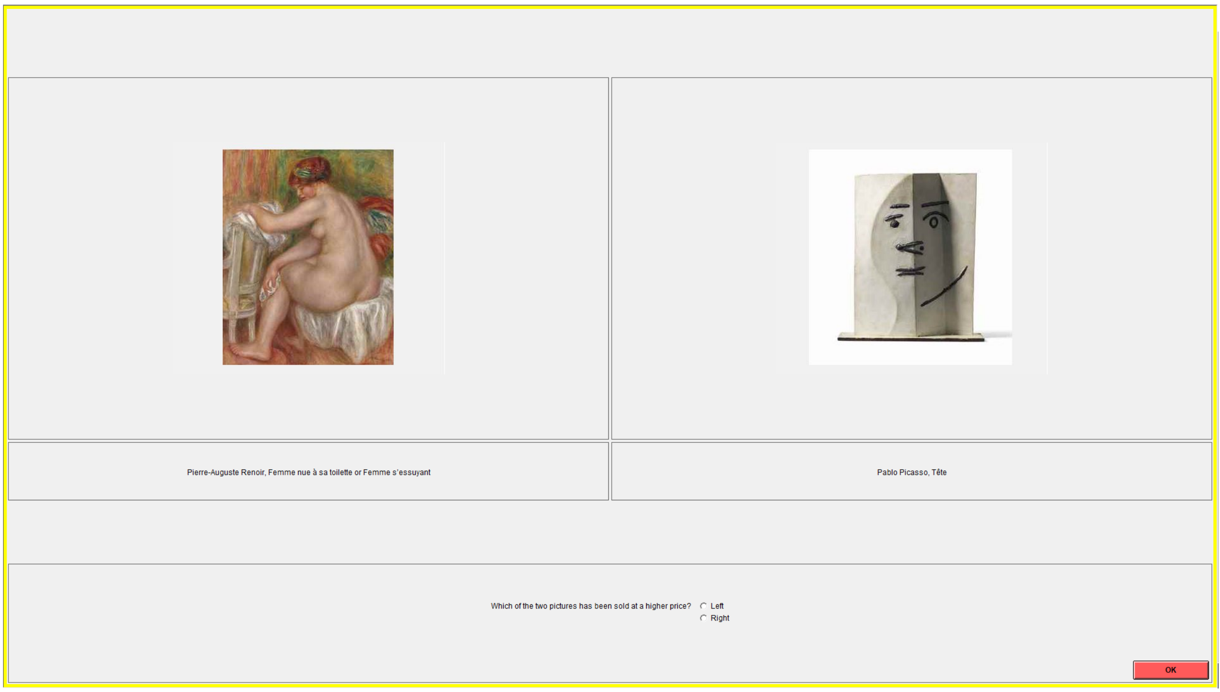

Egon Schiele, Selbstbildnis mit gespreizten Fingern
Le Corbusier, Les deux soeurs
Which of the two pictures has been sold at a higher price?
Left
Right
OK

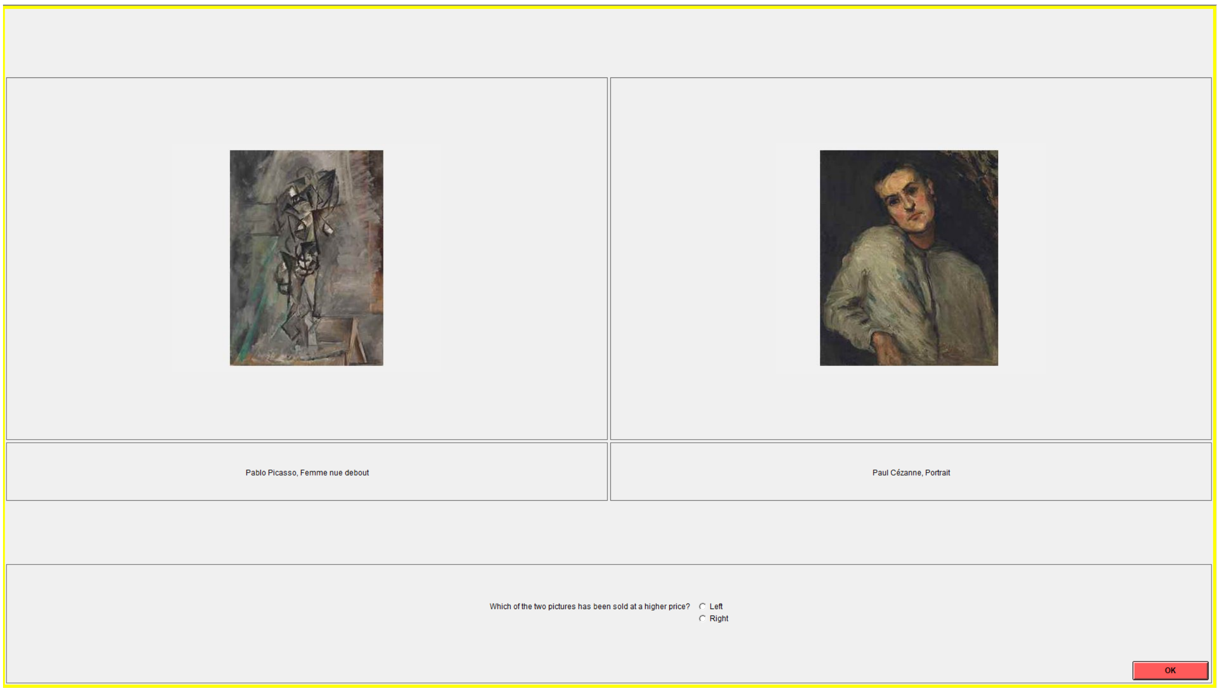

Lyonel Feininger, Calm at Sea II
Lovis Corinth, Flieder, Rosen und Nelken im Sèvreskrug
Which of the two pictures has been sold at a higher price?
Left
Right
OK

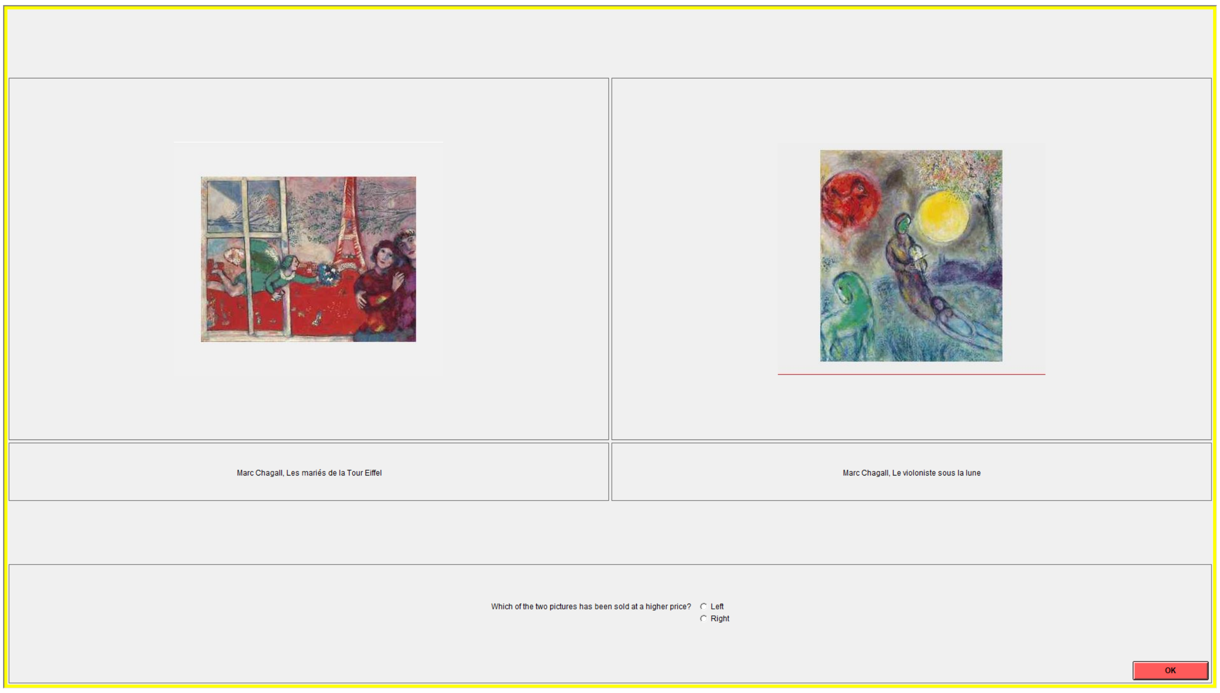

Alberto Giacometti, Femme debout
Pablo Picasso, Mandoline
Which of the two pictures has been sold at a higher price?
Left
Right
OK


Karl Schmidt-Rottluff, Sternenandacht
Henri-Edmond Cross, Etude pour Scène de corrida
Which of the two pictures has been sold at a higher price?
Left
Right
OK

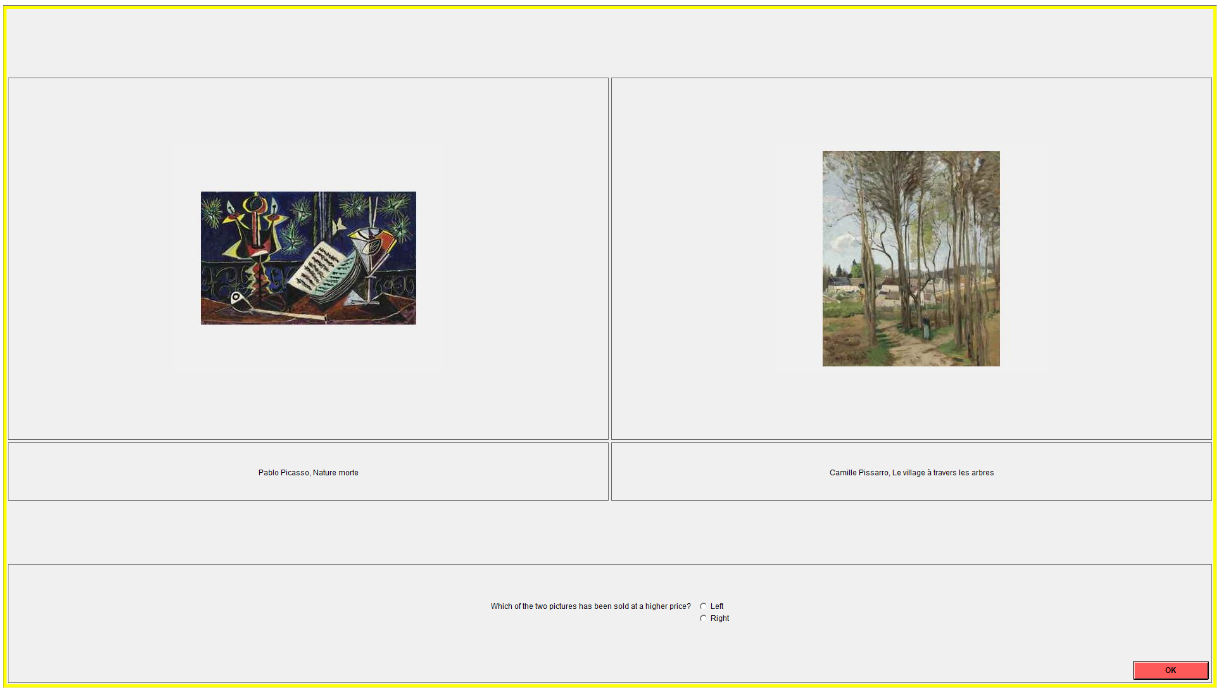

Marc Chagall, L'attente
Fernand Léger, Le moteur
Which of the two pictures has been sold at a higher price?
Left
Right
OK

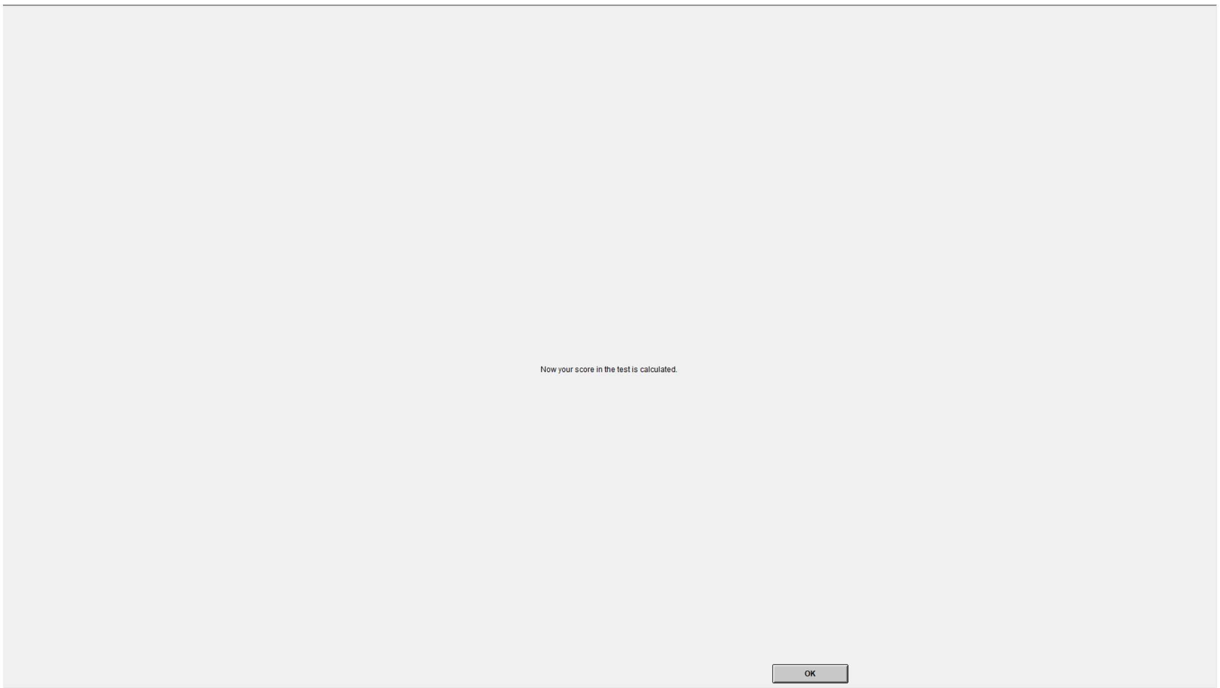

Pierre-Auguste Renoir, Femme nue à sa toilette or Femme s'essuyant
Pablo Picasso, Tête
Which of the two pictures has been sold at a higher price?
Left
Right
OK

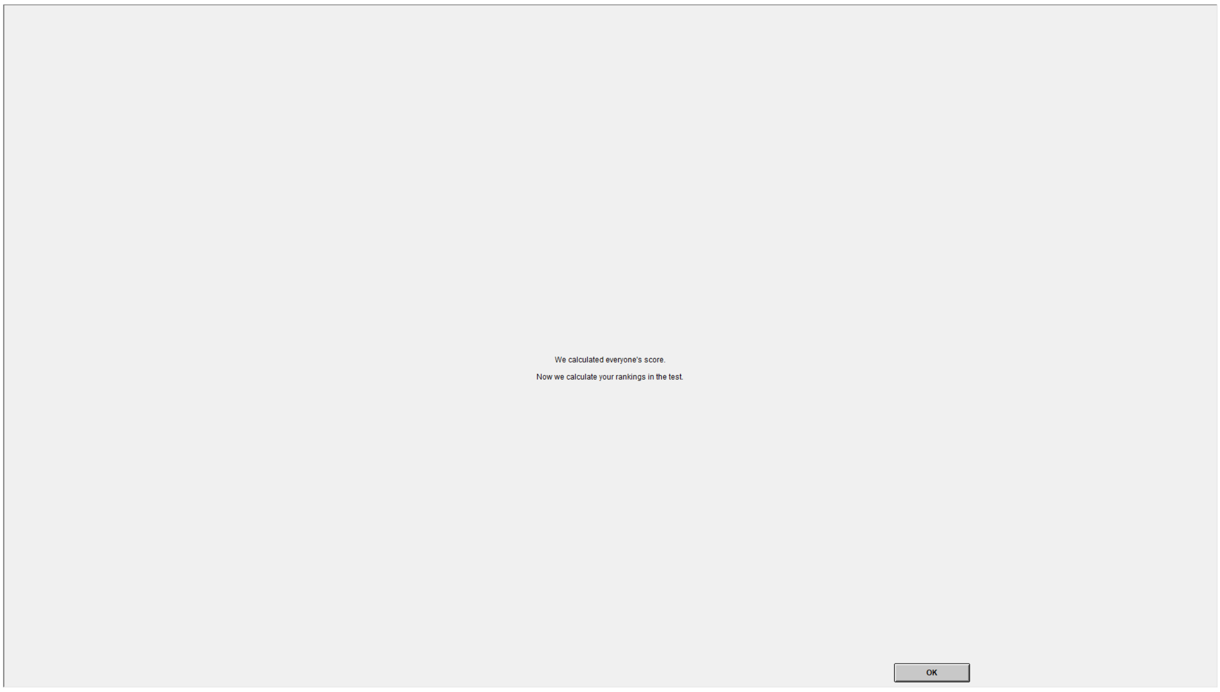

Pablo Picasso, Femme nue debout
Paul Cézanne, Portrait
Which of the two pictures has been sold at a higher price?
Left
Right
OK


Marc Chagall, Les mariés de la Tour Eiffel
Marc Chagall, Le violoniste sous la lune
Which of the two pictures has been sold at a higher price?
Left
Right
OK


Egon Schiele, Oesterreichisches Mäderl
Pablo Picasso, Picador et personnage
Which of the two pictures has been sold at a higher price?
Left
Right
OK

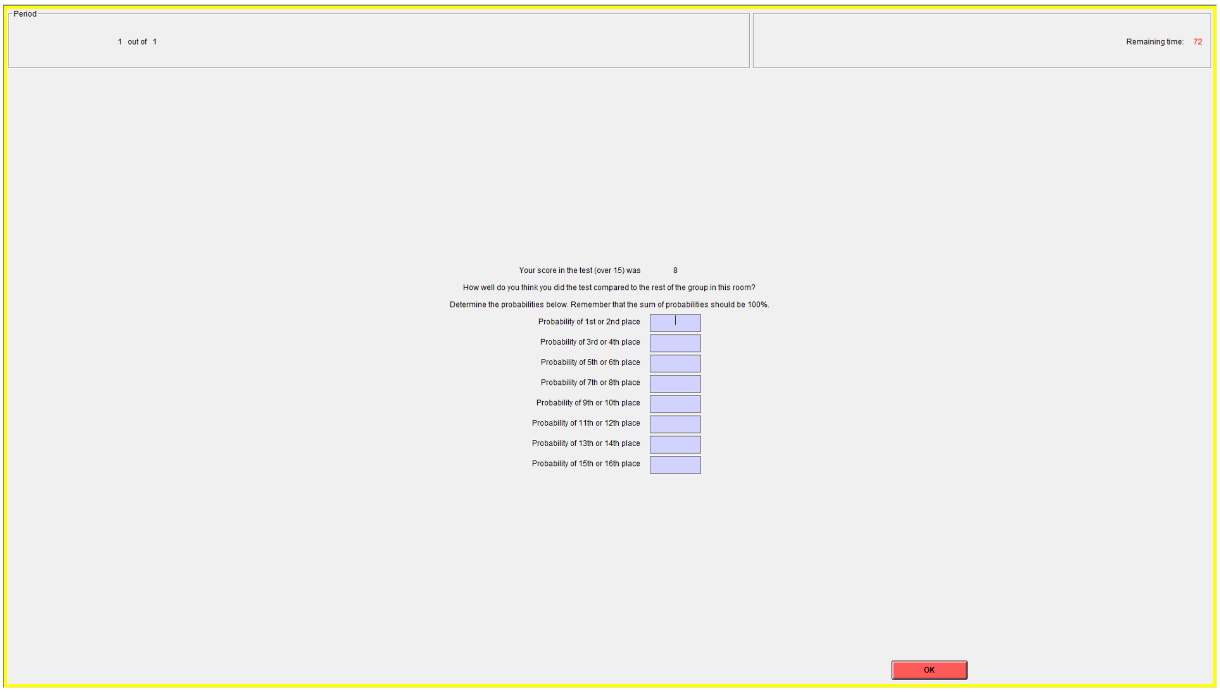

Pablo Picasso, Nature morte
Camille Pissarro, Le village à travers les arbres
Which of the two pictures has been sold at a higher price?
Left
Right
OK

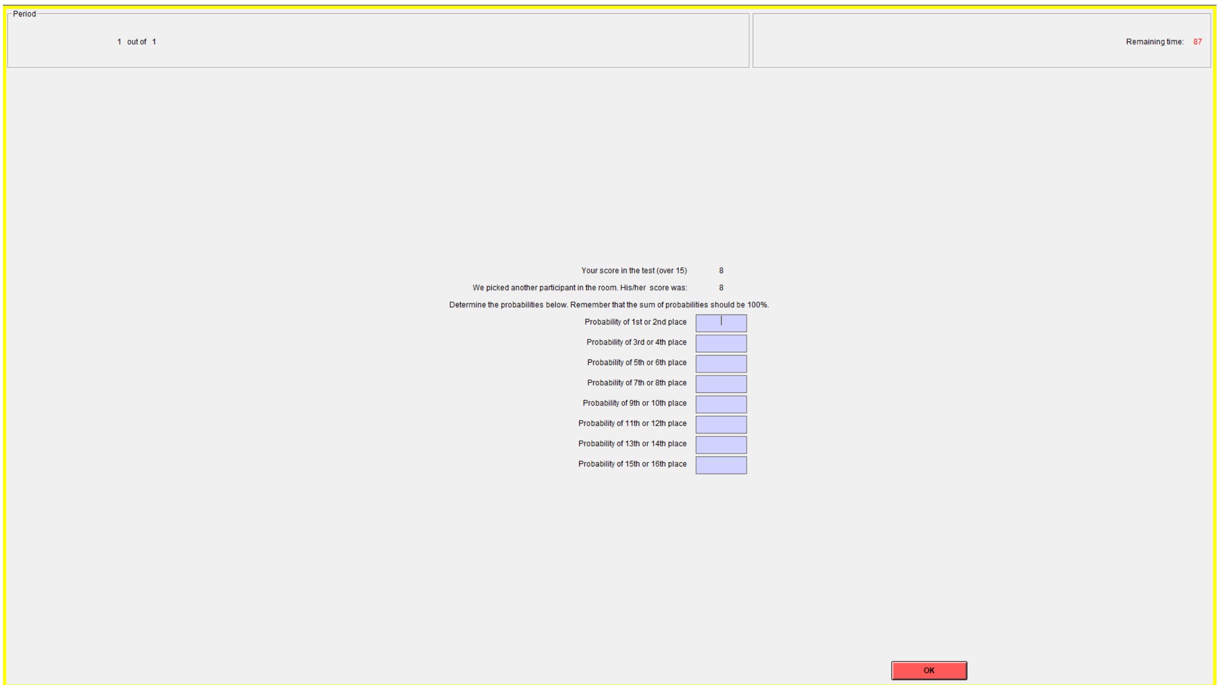

Now your score in the test is calculated.
OK

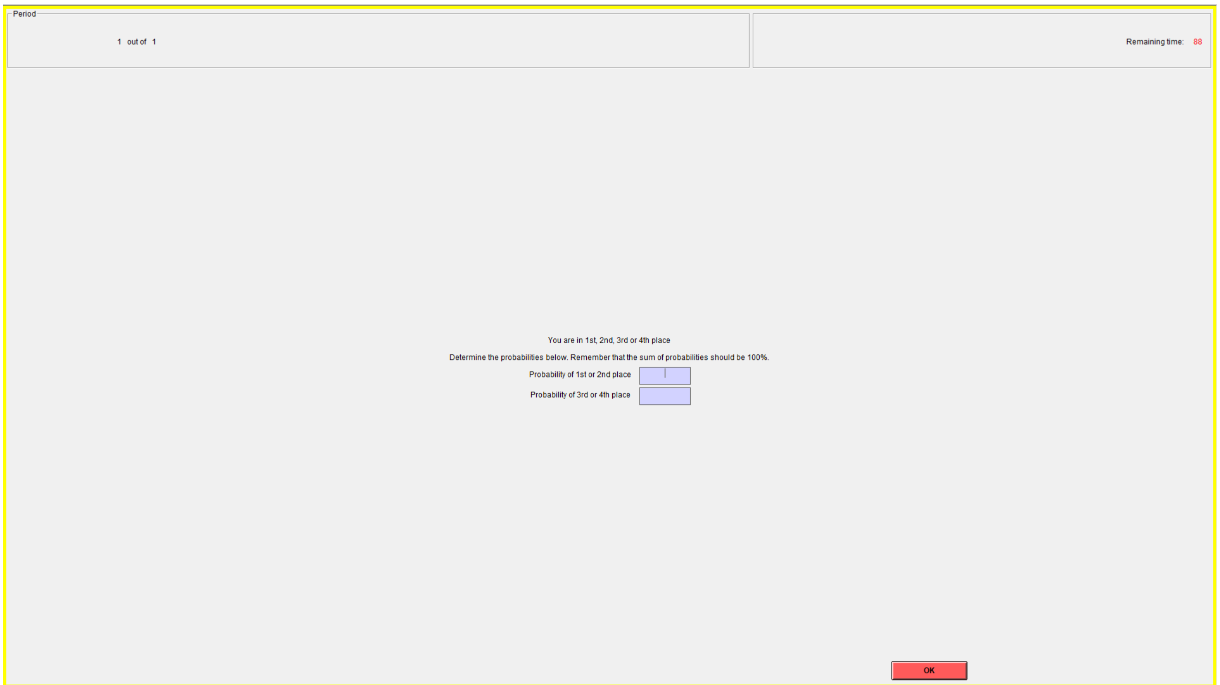

We calculated everyone's score.
Now we calculate your rankings in the test.
OK

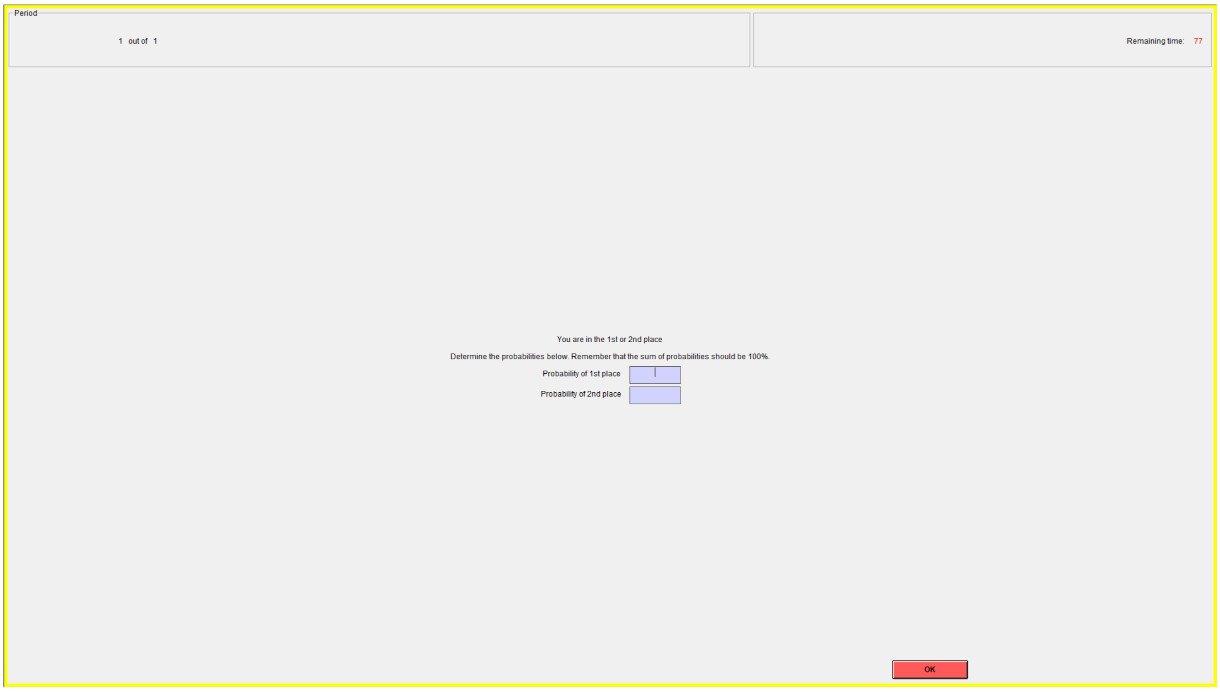
B.3 Screenshots - Probability updating

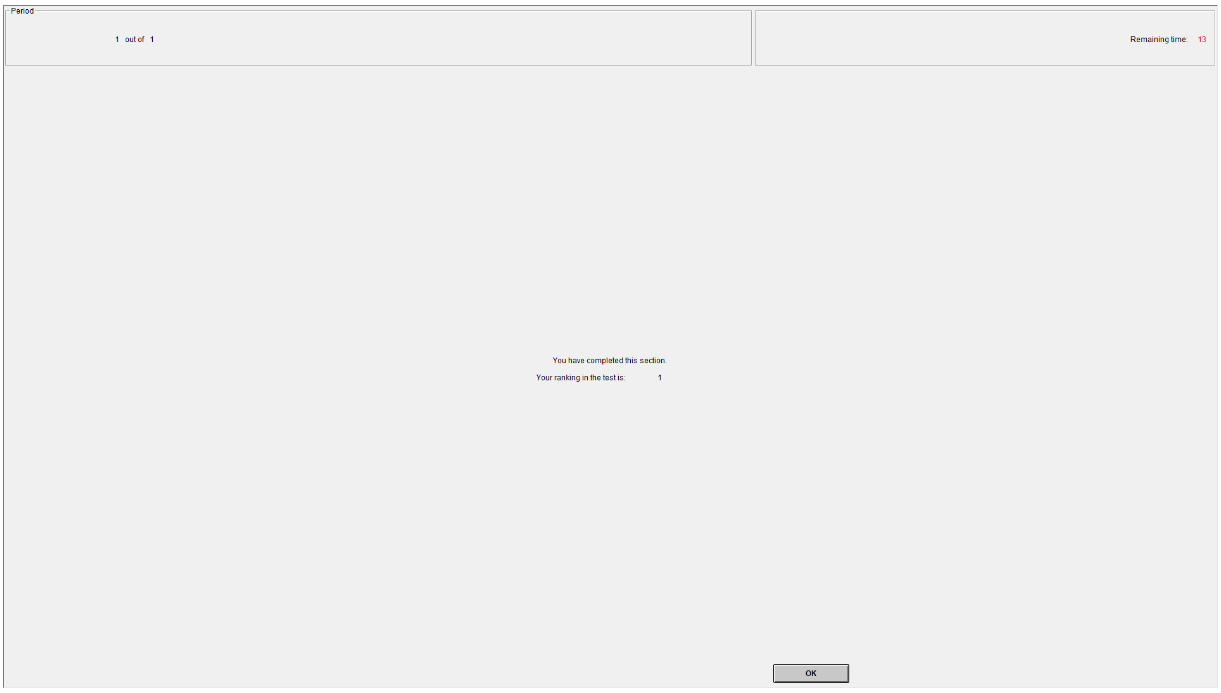
We report screenshots of the probability updating task for a participant that did the best score in the session. For participants with a different position in the ranking the instructions are modified accordingly.




Period
1 out of 1
Remaining time: 87
Your score in the test (over 15) 8
We picked another participant in the room. His/her score was: 8
Determine the probabilities below. Remember that the sum of probabilities should be 100%.
Probability of 1st or 2nd place
Probability of 3rd or 4th place
Probability of 5th or 6th place
Probability of 7th or 8th place
Probability of 9th or 10th place
Probability of 11th or 12th place
Probability of 13th or 14th place
Probability of 15th or 16th place
OK


Period
1 out of 1
Remaining time: 88
You are in 1st, 2nd, 3rd or 4th place
Determine the probabilities below. Remember that the sum of probabilities should be 100%.
Probability of 1st or 2nd place
Probability of 3rd or 4th place
OK


Period
1 out of 1
Remaining time: 77
You are in the 1st or 2nd place
Determine the probabilities below. Remember that the sum of probabilities should be 100%.
Probability of 1st place
Probability of 2nd place
OK


Period
1 out of 1
Remaining time: 13
You have completed this section.
Your ranking in the test is: 1
OK

B.4 FOOT task

In the FOOT task, we ask participants to compare 15 pairs of players for whom we provide names, team, and a close-up picture. Table B1 reports the pairs of players that participants face during the experiment. The team refers to the Turkish League 2017/2018.

Taliska (Besiktas JK)	Serdar Aziz (Galatasaray SK)
Garry Rodrigues (Galatasaray SK)	Adriano Correia Claro (Besiktas JK)
Mathieu Valbuena (Fenerbahce SK)	Aziz Behich (Burnaspor)
Pepe (Besiktas JK)	Sofiane Feghouli (Galatasaray SK)
Mustafa Yumlu (Akhisarspor)	Maicon Pereira Roque (Galatasaray SK)
Hasan Ali Kaldirim (Fenerbahce SK)	Bafétimbi Gomis (Galatasaray SK)
Bogdan Stancu (Burnaspor)	Juraj Kucka (Trabzonspor)
Pablo Batalla (Burnaspor)	Deniz Kadah (Antalyanspor)
Dusko Tosic (Besiktas JK)	Ricardo Quaresma (Besiktas JK)
André Castro (Göztepe)	Serginho (Akhisarspor)
Younès Belhanda (Galatasaray SK)	Emre Akbaba (Alanyaspor)
Titi (Burnaspor)	Burak Yilmaz (Trabzonspor)
Martin Skrtel (Fenerbahce SK)	Mehmet Topal (Fenerbahce SK)
Ryan Babel (Besiktas JK)	Josef de Souza Dias (Fenerbahce SK)
Roman Neustädter (Fenerbahce SK)	Emmanuel Adebayor (Basaksehir)

Table B1: Pairs of players in FOOT test

C Appendix: Questionnaire

We list here the different variables elicited in our post-experimental questionnaire.

- Age of the subject (in years).
- Sex of the subject (1=male, 0=female).
- Living: living arrangement for the subject (0=student housing, 1=with family, 2= with friends, 3=alone).
- Siblings: number of siblings of subject.
- Older siblings: number of siblings who are older than the subject.
- Trust: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" (Be careful 0 ... 10 can be trusted)
- Risk: "How willing are you to take risks in general?" (0 lowest – 10 highest)
- Major: subject's major (2=economics, 1=business, political science or international trade, 0=other).
- Econ: number of economics classes (censored at 4).
- Friends: number of people known in the session
- Rely: "How much can we trust the data coming from you in this experiment?" (0 lowest – 10 highest)

D Appendix: Additional tables

	Persistence of Guessed Rank					
	(1)	(2)	(3)	(4)	(5)	(6)
	Step 3		Step 4		Step 5	
Assortative	-0.091	-0.058	0.094	-0.000	-0.123	-0.157
	(0.232)	(0.164)	(0.121)	(0.0931)	(0.171)	(0.132)
Disassortative	-0.089	0.579***	0.062	0.149	0.080	-0.125
	(0.227)	(0.193)	(0.102)	(0.0937)	(0.118)	(0.142)
Constant	4.708***	12.25***	6.240***	13.31***	12.60***	26.10***
	(0.930)	(0.852)	(1.594)	(1.213)	(3.559)	(2.637)
Controls	YES	YES+	YES	YES+	YES	YES+
Observations	287	288	287	288	287	288
R-squared	0.068	0.136	0.054	0.006	0.028	0.008

Table D1: OLS regression of confidence (guessed rank) on treatment dummies across the various steps of the belief revision process. Bottom half of performers only. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. The large set additionally controls for measures of risk aversion and trust, whether they are economics students and linearly for how many friends they have. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	Guessed Rank - split by gender					
	(1)	(2)	(3)	(4)	(5)	(6)
	All Ranks		Rank< 9		Rank> 8	
	F	M	F	M	F	M
Assortative	0.006	-0.290	0.177	-0.277	0.007	-0.265
	(0.536)	(0.582)	(0.627)	(0.576)	(0.578)	(0.636)
Disassortative	0.808	0.444	-0.180	0.054	1.378***	1.208
	(0.493)	(0.632)	(0.703)	(0.534)	(0.493)	(0.723)
Constant	10.34***	8.678***	11.07***	5.769**	9.995***	7.320***
	(1.993)	(1.584)	(3.166)	(2.150)	(3.056)	(1.991)
Controls	YES	YES	YES	YES	YES	YES
Observations	231	344	97	190	134	154
R-squared	0.029	0.012	0.028	0.006	0.076	0.057

Table D2: OLS regression of confidence (guessed rank) at step 2 of the belief revision stage on treatment dummies split by gender. The small set of controls includes age, whether the participant lives with their parents and whether they have siblings. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	Guessed Rank - split by task			
	(1)	(2)	(3)	(4)
	all	ART	IQ	FOOT
Assortative	-0.183	-0.121	0.155	0.155
	(0.383)	(0.361)	(0.578)	(0.578)
Disassortative	0.560	0.550	0.405	0.405
	(0.424)	(0.481)	(0.763)	(0.763)
Constant	10.26***	9.902***	8.976***	8.976***
	(1.194)	(1.620)	(1.836)	(1.836)
Controls	YES	YES	YES	YES
Observations	575	191	192	192
R-squared	0.086	0.037	0.330	0.330

Table D3: OLS regression of confidence (guessed rank) at step 2 of the belief revision stage on treatment dummies split by task. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	Change in Confidence					
	(1)	(2)	(3)	(4)	(5)	(6)
	Entire Sample		Rank< 9		Rank> 8	
Assortative	0.265	0.261	0.428**	0.421**	0.126	0.124
	(0.255)	(0.257)	(0.178)	(0.174)	(0.392)	(0.389)
Disassortative	0.0720	0.0743	-0.173	-0.169	0.329	0.365
	(0.267)	(0.268)	(0.210)	(0.210)	(0.416)	(0.423)
Constant	1.463	1.530*	1.324	1.061	0.802	1.320
	(0.885)	(0.849)	(0.842)	(0.929)	(1.630)	(1.497)
Controls	YES	YES^{+}	YES	YES^{+}	YES	YES^{+}
Observations	575	575	287	287	288	288
R-squared	0.014	0.017	0.063	0.080	0.014	0.023

Table D4: OLS regression of the Change in Confidence between step 1 and 2 of the belief revision stage on treatment dummies. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. The large set additionally controls for measures of risk aversion and trust, whether they are economics students and linearly for how many friends they have. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	Absolute Error - split by gender					
	(1)	(2)	(3)	(4)	(5)	(6)
	All Ranks		Rank< 9		Rank> 8	
	F	M	F	M	F	M
Assortative	-0.052	-0.001	0.158	-0.051	-0.141	0.061
	(0.194)	(0.133)	(0.226)	(0.113)	(0.307)	(0.249)
Disassortative	-0.430*	-0.134	0.053	0.205*	-0.756***	-0.479*
	(0.212)	(0.141)	(0.202)	(0.114)	(0.271)	(0.279)
Constant	1.142	2.630***	0.514	1.247*	1.805	2.556**
	(1.080)	(0.891)	(1.400)	(0.710)	(1.482)	(1.046)
Controls	YES	YES	YES	YES	YES	YES
Observations	231	345	97	191	134	154
R-squared	0.039	0.007	0.025	0.028	0.077	0.035

Table D5: OLS regression of Absolute Error at step 2 of the belief revision stage on treatment dummies split by gender. The small set of controls includes age, whether the participant lives with their parents and whether they have siblings. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	Absolute Error - split by task			
	(1)	(2)	(3)	(4)
	all	ART	IQ	FOOT
Assortative	-0.014	0.125	-0.152	-0.079
	(0.092)	(0.110)	(0.149)	(0.174)
Disassortative	-0.245**	-0.359***	-0.325	-0.080
	(0.101)	(0.037)	(0.185)	(0.138)
Constant	2.299***	2.184**	2.923*	2.427*
	(0.732)	(0.743)	(1.475)	(1.329)
Controls	YES	YES	YES	YES
Observations	576	192	192	192
R-squared	0.013	0.036	0.024	0.005

Table D6: OLS regression of Absolute Error at step 2 of the belief revision stage on treatment dummies split by task. The small set of controls includes age, gender, whether the participant lives with their parents and whether they have siblings. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

E Appendix: Additional figures

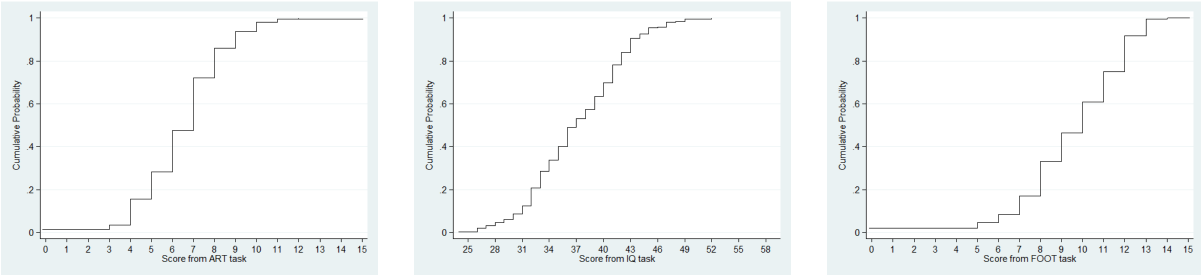


Figure E1: Score distributions for the three tasks.

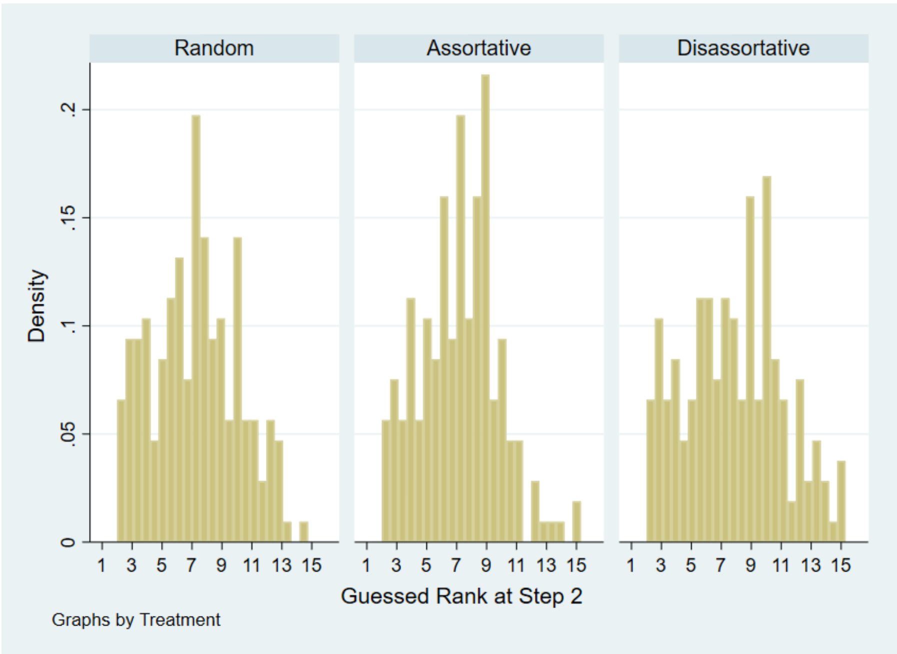


Figure E2: Histogram of Average Guessed Rank at step 2 of the belief revision stage.

Incidental Affects, Belief Management, and Decision-Making

Elisabetta Leni*
University of Essex

Abstract

This paper uses a lab experiment to study how incidental affects influence decisions and beliefs in two classical choice situations - a dictator game and an effort task. We employ music to provoke an exogenous shift in participants' affective state after they enter the lab. The experimental design is between-subject. We vary the valence - positive or negative - of the affect induced and the moment, in the timeline of the experiment, when participants listen to the music. Decisions and beliefs under the effect of the exogenous affective activation are compared to a baseline where participants' affective state remains untouched. The analysis reveals that incidental affects reduce altruism. Moreover, our treatments show that participants with high trait emotional intelligence (Petrides et al., 2007) strategically manage their beliefs when under the influence of the exogenous affective activation. Both in the dictator game and effort task, they appear to overestimate the likelihood of the negative outcome. We conjecture that they adopt a strategic pessimism approach as a way of protecting themselves from the emotional reaction should the negative outcome materialize. We also find that high trait emotional intelligence is predictive of the outcome in the effort task. Here, the decrease in confidence induced by incidental affects in participants with high trait emotional intelligence generates a decrease of 9% in effort provision. This result confirms the importance of confidence for performance.

*Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK; *E-mail address*: eleni@essex.ac.uk

1 Introduction

Affects are general feelings of pleasure or displeasure accompanied by arousing or quieting bodily activation. Several studies in cognitive psychology suggest that affects can influence what we think and remember (Bower, 1981; Forgas and Bower, 1987; Kavanagh and Bower, 1985; Bower et al., 1981), and how we reason (Schwarz and Clore, 1983; Isen and Reeve, 2005). Neuroscientific data show that affective states and higher cognition are integrated in the brain (Barrett, 2017a; Schaefer et al., 2002) indicating that rational decision-making and the processing of affects are intertwined. Those studies suggest that affective states are relevant for beliefs formation and choices even when unrelated to the judgement at hand (i.e., they are incidental).

Despite the relevance the topic, few researchers in economics have tried to model the influence of incidental affects on decision-making (Loewenstein and Lerner, 2003; Loewenstein, 2000) and the experimental evidence is still very limited (Charness and Grosskopf, 2001; Kirchsteiger et al., 2006; Capra, 2004; Oswald et al., 2015; Drouvelis and Grosskopf, 2016; Proto et al., 2019; Oswald et al., 2015). A fundamental question not yet addressed in the literature is whether incidental affects influence the way people manage their beliefs. This question is particularly relevant in light of the theories on belief-dependent preferences (Battigalli and Dufwenberg, 2009; Köszegi and Rabin, 2006) that suggest that expectations could act as a reference point affecting people's decisions both in interactive games and individual tasks. In this framework, identifying what is preference and what is belief-driven becomes crucial. The present paper aims at filling this gap and investigates the influence of incidental affects on decisions and beliefs in two classical choice situations - a dictator game and an effort task - using a lab experiment.

We design an Affect Induction Procedure (AIP) to provoke an exogenous shift in participants' affective state after they enter the lab. Through a series of pilot sessions, we identify two music tracks that induce high arousal and feelings with positive or negative valence. During the experiment, we ask participants to listen to one of the music tracks to produce the incidental affects with the desired valence. Besides the AIP, the experiment has two other main stages: a dictator game and an effort task with belief elicitation. Then, we collect extensive data on participants ability to recognize others' emotions and self-evaluation on their capacity to manage and be aware of their own and others' emotional states. Finally, we collect measures of cognitive ability, risk attitudes, demographics and music tastes or experience.

The experimental design is between-subject. We vary the valence - positive or negative - of the affect induced and the moment, in the timeline of the experiment, when participants experience the AIP. The variation in timing aims at inducing incidental affects at different stages during the experiment to identify the influence of our treatments on each choice, performance and elicited belief in the two main tasks. Decisions and beliefs under the effect of the AIP are compared to a neutral baseline where participants' affective state remains untouched. The design allows us to separately study the effect of a change in valence combined with an increase in arousal on belief management and decision-making.

Our study finds its main theoretical underpinnings in the psychological literature, with few contributions from the field of economics. Different authors theorize that both positive and negative affects should promote helpfulness and altruism (Carlson et al., 1988; Isen and Simmonds, 1978). Loewenstein and O'Donoghue (2007) attempt to explain helping behavior by modelling the interaction between an affective and a deliberative system. Based on a review of empirical findings, Lyubomirsky et al. (2005) suggest a conceptual model that explains the association between positive affects and confidence, self-efficacy, and success. Oswald et al. (2015) propose a microeconomic model of emotional internal resource allocation that links (un)happiness and (decreased) productivity. Several theories link incidental affects to beliefs management strategies (Isen and Simmonds, 1978; Cialdini et al., 1973; Weber, 1994) and the recalling of events from memory (Schwarz and Clore, 1983; Bandura, 1977). Despite a large number of experimental studies (the vast majority focusing on the affect-helpfulness link), the evidence is still controversial and the problems associated with the various AIPs adopted are often neglected. We believe our paper adds to the existing literature by providing new evidence based on a sound experimental methodology. Moreover, we explore the link between incidental affects and belief management which has never been explored in an incentivized setting.

We believe that people might react differently to incidental affects and heterogeneity might be one of the causes of the inconclusive evidence encountered in the literature. Some scholars (Yip and Côté, 2013) argue that the ability to understand their own and others' emotions might help to pinpoint the correct source of incidental feelings and become aware of their irrelevance for the decision at hand. We explore this heterogeneity by collecting extensive data on participants' 'emotional intelligence'. We are well aware of the debate concerning the nature of the construct and we choose to adopt measures pertaining to the two main competing approaches on the topic - the 'ability' approach (Salovey and Mayer, 1990) and trait emotional intelligence (Petrides et al., 2007). We conjecture that participant with high emotional intelligence should be the ones better equipped to recognize the incidental nature of affects attenuating their influence on decisions and beliefs.

To a large extent, the results of our experiment do not confirm the theoretical predictions. Positive affects significantly *decrease* altruism in the dictator game. Also, dictators believe more recipients expect them to be selfish after listening to positive music compared to the baseline. We find no significant effect of the AIP on recipients' beliefs, participants' confidence or performance in the effort task. When we study the interaction between treatments and emotional intelligence, we find no effect for the 'ability measure but a significant heterogeneity related to trait emotional intelligence. Participants with high trait emotional intelligence are the ones that seem to be affected the most by the treatments. Our results suggest that they might strategically overestimate the likelihood of a negative outcome as a way of protecting themselves from the emotional reaction should the negative outcome materialize (Shepperd et al., 2000; Risen and Gilovich, 2007).

We test the hypothesis that beliefs act as a reference point affecting decisions in the dictator game and performance in the effort task by conducting a mediation analysis (Mackinnon et al., 2007; Imai et al., 2010). We assume that dictators might be guilt averse (Battigalli and Dufwenberg, 2007) and bear a psychological cost when they let their partner down. It follows that the observed effect of the positive music on dictators' choices might be indirectly and partly driven by the shift in beliefs. In fact, the above-described increase of second order beliefs in the dictator game should reduce the psychological cost of guilt associated with the selfish choice thus making that option more attractive. However, we do not find a significant mediation effect of beliefs on altruism. We perform a similar analysis for the effort task assuming that confidence might act as a motivational boost (Bénabou and Tirole, 2002) and sustain effort. The analysis confirms that confidence is a mediator of performance for participants with high trait emotional intelligence. For them, the decrease in confidence generates a decrease of 9% in effort provision.

This study is the first that tries to separately identify the effect of incidental affects on decision-making and beliefs and explore the relation between belief management and decision-making. Our results are relevant because they point to the capacity, for a relevant proportion of the population, of strategically managing their beliefs in an attempt to cope with the consequences of the choices that they and others make. This study supports the idea that beliefs not always reflect the individuals' knowledge concerning the state of the world but the mechanism through which we form beliefs might be more complex (Schwarz and Clore, 1983) and deserve further investigation. Moreover, our experiment lends partial support to belief-dependent theories by means of a novel experimental design. The mediation analysis indicates that performance can be significantly affected by beliefs. Understanding how incidental affects influence preferences and beliefs adds crucially on the debate concerning stability of preference (Stigler and Becker, 1977) and belief-dependent motivation (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009; Köszegi, 2006).

In section 2 we describe the AIP and test its efficacy. Section 3 gives details on the experimental design and procedure, describes the experimental treatments, and introduces our conjectures. Section 4 is devoted to the statistical analysis and section 5 speculates on the possible mechanism driving the results paving the ground for future research. Section 6 concludes. The appendices report experimental instructions, additional analysis and robustness checks.

2 Affect Induction Procedure (AIP)

In this section, we discuss the AIP and present an assessment of its validity. First, we motivate the choice of the AIP adopted in the study. Second, we present the treatment manipulation check and the experimental procedures adopted in the pilot sessions. Third, we present the results of the pilot sessions. Finally, we comment on our methodological contribution.

2.1 Why affects?

The decision to focus on affects, rather than emotions or moods, stems from three considerations. The first has its roots in the scientific discussion about emotions that calls into questions the existence of emotions intended as discrete categories. Our research refers to the theory of constructed emotions (Barrett, 2017b) that questions the assumption that emotions are genetically endowed and produced by dedicated circuits in the brain. According to this theory, what *exists* in the brain and body is affect while emotions are complex and culturally driven concepts that can be controlled through perspective taking and re-categorization. As such emotions are not ‘reactions’ to external events nor they are universal. We thus expect a standardised procedure to reliably induce affects but not emotions.

The other two considerations are methodological. We argue that variations in arousal and valence can be accurately measured using self-reported questionnaires. Several studies show that people are usually able to give an explicit account of pleasant and unpleasant feelings using a variety of self-rating scales (Barrett and Russell, 1998; Bradley and Lang, 1994; Carroll et al., 1999; Lang et al., 1993). Moreover, facial expressions highly correlate with affective valence judgements and skin conductance magnitude with arousal ratings (Greenwald et al., 1998; Lang et al., 1993). Thus, surveys provide a cheap and easy way to test the efficacy of the procedure. The measurement of variations in specific emotions or moods requires the use of biometrics or more sophisticated forms of self-assessment which we cannot expect from all participants. Finally, we are confident that a laboratory procedure could effectively induce short term variations in affects with a standardized procedure. The same might not be true for emotions or moods. Some authors (Martin, 1990) argue that inducing emotions might encounter the issue of specificity since emotional states can overlap (e.g., sadness and anxiety) (Goldberg and Huxley, 1980) or depend on idiosyncratic cognitive assessment (Barrett, 2017a). Other authors (Marston et al., 1984) note that only a low level of intensity of mood can be induced. Another peculiarity of moods - their duration - might be hard to create in the lab as the effect of the procedure tends to fade out quickly.

2.2 Why music as induction procedure?

Psychologists use many different techniques to induce affective states in the lab. Most studies rely on recollection of autobiographical events, short films, or imaging of emotional situations - like the so-called Velten procedure (Velten, 1968).[1] The idea of using music to induce affective states in the lab was first introduced by Clark (1983). Studies reviewing the effectiveness of several AIPs have concluded that music is among the most effective (Westermann et al., 1996; Gerrards-Hesse et al., 1994; Martin, 1990). The original procedure contains instructions about how participants should go about obtaining the desired affective states, but later studies minimize such instructions (Pignatiello et al., 1986). In the experiment, we also use minimal instructions.[2] An important advantage of our procedure is that it minimizes demand effects and cognitive priming which significantly interfere with other techniques (Westermann et al., 1996). Also, it appears to be particularly suitable to induce affects, rather than specific emotions.

2.2.1 Selection of music tracks

The starting point of our research is the review of Västfjäll (2001) which provides a list of music tracks used in previous experimental studies. We select 2 (2) tracks from the list to induce high arousal and positive (negative) valence. The tracks for the positive valence are Delibes, L. (1870) *Coppelia (Mazurka, Act I, No. 3)* and Bach, J. S. (1977) *Brandenburg Concerto No. 2, BWV 1047 (First Movement; Allegro)*. The tracks for the negative valence are Moussorgsky, M. (1988) *Night on Bald Mountain* and Holst, G. (1986) *The Planets: Mars, the bringer of war*. Given the poor results concerning the efficacy of the music to induce negative valence (see section 2.5.2), we test an additional track. This is *Farsa Del Buen Vivir* by Merzbow, a track of noise music[3] that has never been tested or used in lab experiments before. All the music tracks have been cut to a duration of just over 4 minutes.

[1] For a comprehensive list of studies using affect induction procedures see the reviews by Westermann et al. (1996); Gerrards-Hesse et al. (1994); Martin (1990).

[2] We ask subjects to ‘listen to the music and tune with it’.

[3] Noise music is a category of music that is characterized by the expressive use of noise within a musical context.

2.3 Treatment manipulation check

We test the efficacy of the music tracks in a series of pilot sessions using several self-reported questionnaires. First, we use the Affect grid (Russell et al., 1989), a scale designed as a quick means of assessing affects along the dimensions of pleasure-displeasure and arousal-sleepiness. The Affect grid is a kind of map for feelings (see Figure 1) where the centre of the square represents a neutral, average, everyday feeling. The (left) right half of the grid represents (un)pleasant feelings. The farther to the (left) right the more (un)pleasant. The vertical dimension of the map represents the degree of arousal. The top half is for feelings that are above average in arousal. The lower half for feelings below average. Participants should select the square that best describes how they felt after listening to music.

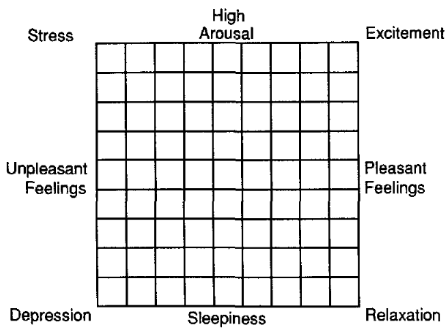


Figure 1: Affect grid Russel et al. (1989)

The Affect grid is easy to administer. It is a 9-point scale offering substantial variations in the two dimensions under analysis. However, to the best of my knowledge, there is a lack of studies confronting the evaluation given with the Affect grid with biometric measures. We thus decided to administer a second questionnaire - the Self-Assessment Manikin (SAM) (Bradley and Lang, 1994) - that was proven to have a strong correlation with skin conductance and facial expressions (Greenwald, Cook, and Lang, 1988; Lang et al., 1993). The SAM provides a non-verbal pictorial representation of the concepts of arousal and valence and a variation on a 5-points scale across the two dimensions. Participants should select the images (see Figure 2) that best describe their reaction to the music.[4] The comparison of the

[4]The test also includes an assessment of Dominance. The data concerning the Dominance scale are not reported as they are beyond the scope of our research.

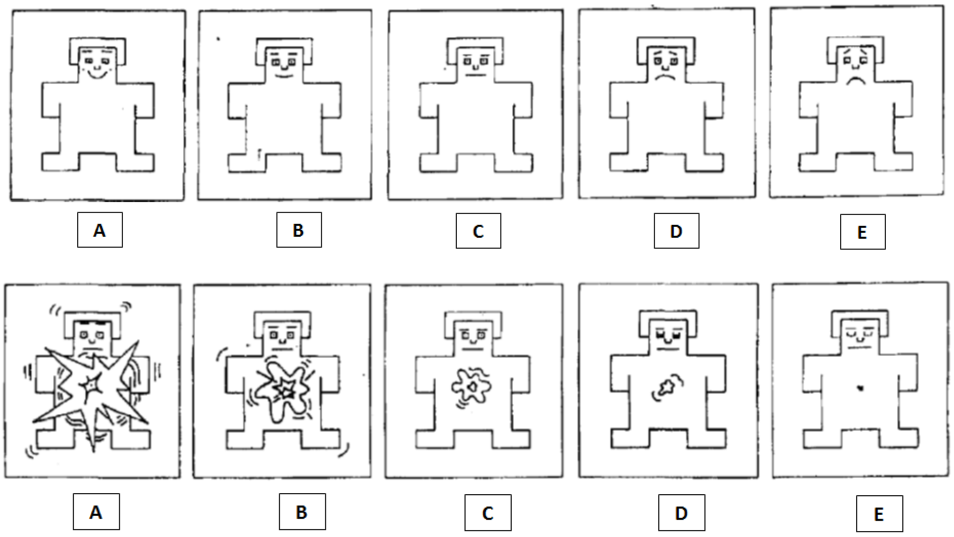


Figure 2: Self-assessment Manikin (Bradley and Lang, 1994) - Valence scale on top, arousal below

data extracted from the two questionnaires could provide a test of the internal consistency of the data and a sensible treatment manipulation check.

To further understand how participants perceive the music, they complete the Positive Affect Negative Affect Schedule (PANAS)(Watson et al., 1988), a self-report questionnaire that consists of two 10-item scales to measure both positive and negative affects. Participants are asked to indicate to what extent they felt in the way described by the items after listening to the music. The words related to negative affects are: distressed, upset, hostile, irritable, scared, afraid, ashamed, guilty, nervous, and jittery. The words related to positive affects are attentive, interested, alert, excited, enthusiastic, inspired, proud, determined, strong, and active. Participants choose one option out of a 5-Likert scale, from ‘very slightly or not at all’ to ‘extremely’.

Finally, we use a post-experimental questionnaire to collect demographics and ask some control questions regarding the music tracks and their appreciation (see A for details).

2.4 Pilot sessions - procedure

The efficacy of the music tracks was tested in a series of pilot sessions. The sessions took place in June 2018 at the EssexLab of the University of Essex. The participants were recruited through the ESSEXLab

Online Recruitment System and invited to take part in a computerized session at the lab on campus. The research received Ethical Approval from the Social Sciences Faculty Ethics Sub-Committee of the University of Essex. Conforming to departmental ethics guidelines and current practice in experimental research with human subjects, participants read and signed informed consent before taking part in the sessions. The experiment was programmed and conducted with the software z-tree (Fischbacher, 2007).

The pilot sessions involved 66 participants; more than 90% were students. Experimental sessions consisted of 8 to 20 participants. They were randomly assigned to cubicles and told that all the instructions would appear on the computer screen. For the screenshots of the experiment, see Appendix A.

Participants in all AIP conditions were asked to wear headphones and listen to the music before completing the questionnaires. Participants in the baseline condition were asked to complete the questionnaires right after the assignment to cubicles without listening to any audio recording. The sessions lasted on average 30 minutes, and participants earned a flat fee of 7 Pounds.

2.5 Pilot sessions - results

2.5.1 Positive valence

Figure 3 shows the mean values for the Affect grid and SAM obtained in the pilot sessions testing Bach and Delibes's tracks. Delibes's *Coppelia* consistently generates higher arousal and positive valence with respect to the baseline condition. The data of both the SAM and Affect grid indicate a significant increase in arousal and the difference in valence is highly significant when considering the SAM. The track by Bach seems to be increasing positive valence and arousal but the effect is sensibly weaker.

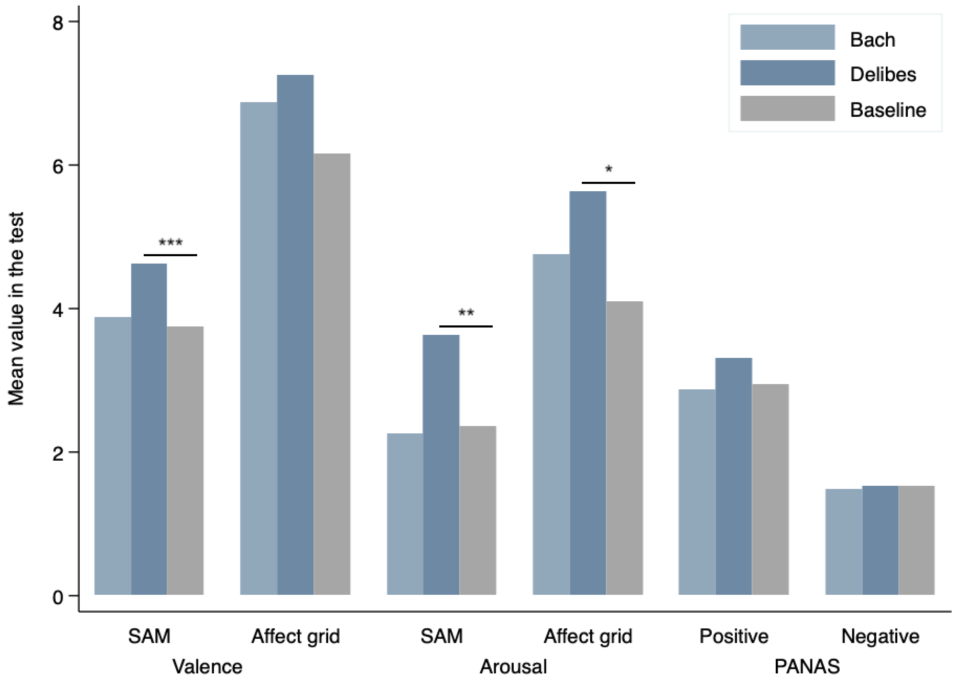


Figure 3: Mean scores in the tests for Bach, Delibes, and Baseline. Stars refer to one-sided t-test with null hypothesis Baseline < Delibes; $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Sample size is 8 for Bach, 8 for Delibes, and 20 for baseline.

Delibes's track also provokes an increase in the mean of all the evaluations concerning positive words in the PANAS, but the overall variation is not significant. Enthusiastic, proud, and inspired are the items that increase the most, displaying a marginally significant increase. On average, there is no variation in the overall evaluation concerning the negative words of the PANAS and none of the items is affected significantly. The pilot sessions give support to the efficacy of Delibes's track in inducing high arousal and positive affect. We thus use Delibes's *Coppelia* in the main experiment.

2.5.2 Negative valence

Figure 4 shows the mean values for the Affect grid and SAM in the pilot sessions testing Holst, Moussorgosky and Merzbow's tracks. The two pieces of classical music seem ineffective in generating negative valence feelings. In fact, participants reported more positive affects compared to subjects in the baseline after listening to Holst or Moussorgosky. This unexpected result appears to be driven by the interaction between music taste and valence: participants who appreciate the music also report positive feelings (see Appendix B for analysis).

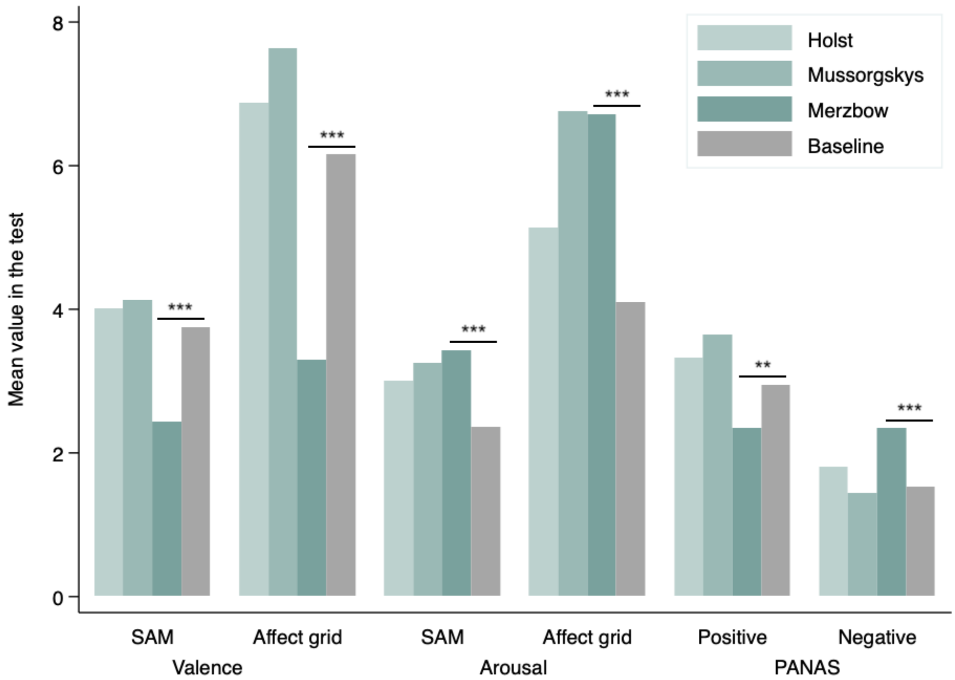


Figure 4: Mean scores in the tests for Holst, Moussorgosky, Merzbow, and Baseline. Stars refer to one-sided t-test with null hypothesis Baseline > Merzbow for valence, Baseline < Merzbow for arousal; $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Sample size is 8 for Holst, 8 for Moussorgosky, 14 for Merzbow, and 20 for baseline.

The music track by Merzbow seems to be highly effective as participants report significantly higher arousal and lower valence than subjects in the baseline. Merzbow also provokes a significant decrease in the positive items of the PANAS and increase in the negative ones. Interested, excited, and enthusiastic are the positive items that decrease the most; distressed, irritable and jittery are the negative items that increase the most. The variation in all the above-mentioned items of the PANAS is significant at 1% level. We thus use Merzobow's *Farsa Del Buen Vivir* in the main experiment.

2.6 Methodological contribution

A review of the experimental literature on incidental affects in incentivized settings reveal several areas of weakness. First, many studies (Bosman and Riedl, 2003; Kirchsteiger et al., 2006; Capra, 2004) compare positive and negative mood/emotion conditions, without using a neutral baseline. However, the independence of positive and negative affect is well-established: affects with opposite valence do not necessarily have opposite effects on behavior (Diener and Emmons, 1984; Watson et al., 1988; Lyubomirsky et al., 2005; Isen, 1990). There is substantial evidence suggesting that emotions with opposite valence can exert similar influences (Lerner and Keltner, 2000, 2001; Keltner et al., 1993; DeSteno et al., 2000). Second, experimental studies adopt different AIPs[5] without adequately discussing their downsides. Westermann et al. (1996) warn against the risk of demand effects in procedures where participants are explicitly instructed to try to enter a specified mood state - like in the Velten procedure, or memory elicitation - or when the affective states is measured right after the procedure (as in Drouvelis and Grosskopf, 2016; Drichoutis and Nayga, 2013; Capra, 2004; Capra et al., 2010; Kirchsteiger et al., 2006).

Our study addresses several drawbacks observed in the experimental literature. First, we use a neutral baseline to separately study the effect of positive and negative affective states. Then, we use a pilot session to test the efficacy of the procedure and avoid measuring the affective state right during the experimental sessions. Finally, we use a procedure - music - with minimal demand effect or possibility for priming.

3 Experimental design, procedure, and conjectures

3.1 Design

Besides the AIP, the experiment involves four incentivized stages. See Appendix C for the screenshots of the experiment and the full set of instructions.

STAGE 1: Dictator Game - Participants are matched to play a one-shot dictator game. Dictators are endowed with 10 Pounds and have to decide how to split the amount choosing one of the following allocations: (7, 3), (5, 5) or (3, 7). After the decision stage, players enter the belief elicitation stage. Recipients are asked to guess the percentage of dictators in the lab making each of the three choices.

[5]The economic literature relied on video clips (Kirchsteiger et al., 2006; Drouvelis and Grosskopf, 2016; Ifcher and Zarghamee, 2011), Velten procedure Oswald et al. (2015), experience of success/failure (Capra, 2004; Capra et al., 2010; Drichoutis and Nayga, 2013), memory elicitation (Capra, 2004; Capra et al., 2010).

Dictators are asked to guess on the guesses of the recipients. Each guess within 5% from the realization is rewarded with 2 Pounds.[6] The payoff of STAGE 1 is the sum of the outcome of the game and the rewards.

STAGE 2: Effort Task - Participants are asked to count the number of zeros in matrices containing only 0 and 1 as in Abeler et al. (2016). Each matrix contains 50 numbers in total. Participants have to enter the number of zeros they think is right and press the 'ok' button. Then, a new table appears on the screen, regardless of whether the entry was correct or wrong. Participants have to count as many tables as possible in 4 minutes. Each correct table is worth 50 cents and nothing is subtracted from participants' earnings if the entry is wrong. Before the effort task participants are asked to reveal their beliefs regarding how well they will do in the task. They are asked to state the probability of belonging to each of the quartiles of the score distribution. Correct guesses are incentivized using a logarithmic scoring rule. The payoff of STAGE 2 is calculated either according to the score in the task or to the guess with equal probability.

STAGE 3: Cognitive skills and risk attitudes - We use two IQ-type questions as in Bigoni et al. (2015) to measure cognitive skills. Each correct answer is worth 2 Pounds. For the risk task, we use the Eckel and Grossman task (Eckel and Grossman, 2002). The payoff of STAGE 3 is calculated by summing the outcome of the lottery with that of the IQ-type task.

STAGE 4: Emotional intelligence - The measurement of emotional intelligence is a cumbersome task as there is no consensus in the literature regarding the nature of the construct. There are two main competing approaches. Some authors believe that there is a set of abilities to validly reason with emotions and to use emotions to enhance thought (Salovey et al., 2004). According to this view, individuals can be more or less capable of solving certain categories of problems about emotions and thus they are more or less 'emotionally intelligent' (Salovey and Mayer, 1990). However, the concept of emotional 'ability' has been heavily criticized. Many scholars argue that it is poorly defined and measured (see Matthews, Gerald et al., 2004 for a comprehensive critical review on the topic). We refer to this set of ideas as the 'ability' approach.

A competing approach (Petrides et al., 2007; Tett et al., 2005) construes emotional intelligence as

[6]We choose this incentive scheme instead of other proper scoring rules because it is simple and easy to describe in instructions.

a set of personality traits. Scholars of *trait* emotional intelligence propose to use a categorization of personality traits that is domain-specific rather than based on five broad and orthogonal dimensions. Thus trait emotional intelligence refers to personality traits in the emotional domain.

It is not the purpose of this research to contribute to the discussion on emotional intelligence. We thus choose to rely on measures pertaining to both approaches. The 'ability' approach has the advantage of relying on measures of maximal performance. For this reason, we could construct an incentive compatible test based on a selection of items from 'Reading the Mind in the Eyes,' (Baron-Cohen et al., 2001) 'Reading the Mind in Films,' (Golan et al., 2006) and 'Reading the Mind in Voices' (Golan et al., 2007) tests. The tests are intended to measure the ability to recognize emotional states and consist of a series of eyes pictures, audio recordings and video where participants have to attribute the correct emotion to the eyes, speaker and actor.[7] Beside the emotion recognition task, we use a small selection of questions (4 items in total) on the verbal knowledge of emotions. The payoff of STAGE 4 is calculated by randomly selecting two subtasks and summing the correct items in those subtasks. Each correct item is worth 1 Pound.

The measure of trait emotional intelligence consists in a self-reported questionnaire - the Trait Emotional Intelligence Short Form questionnaire or TEIQue-SF (Petrides, 2009) - which measures participants' self-perception of their own emotional self-efficacy, i.e., how good they believe they are in terms of understanding, regulating, and expressing emotions in order to adapt to their environment and maintain well-being. The TEIQue-SF is a 30-item questionnaire where responses are expressed in a 7-Likert scale. Appendix C.4 reports the full questionnaire.

In the remaining of the experiment, participants complete a short survey to assess their proneness to guilt - the Guilt and Shame Proneness Scale (GASP) (Cohen et al., 2011). In the post-experimental questionnaire, we collect demographics and some control questions that are intended to measure participants understanding of the goal of the experiment and their appreciation of the music tracks.[8]

Participants are paid two randomly chosen payoffs among the incentivized stages 1 to 4 described above. A show-up fee of 4 Pounds is added to the final payoff of the experiment.

[7]We refer to the above-mentioned papers for details on the items and a definition of what 'correct emotion' means.

[8]This last item could provide a useful check on the efficacy of the AIP as argued in Appendix B.

3.2 Experimental treatments

The experimental design is between-subject with randomization at session-level. There are 8 treatment conditions that vary along two dimensions: the timing and the valence of the affect induced with the AIP. Table 1 reports an overview of our treatments where DG stands for dictator game, ET for effort task. Each column describes the timeline of one treatment in which the coloured cells represent the exact time when the participants listen to the music track (blue cells correspond to Delibes's *Coppelia*, green cells to Merzbow *Farsa del Buen Vivir*). Besides the 8 treatments, we collect data in a neutral condition where participants complete the stages described in 3.1 without listening to any music.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Timing 1	Timing 1	DG-Choice	DG-Choice	DG-Choice	DG-Choice	DG-Choice	DG-Choice
DG-Choice	DG-Choice	Timing 2	Timing 2	DG-Beliefs	DG-Beliefs	DG-Beliefs	DG-Beliefs
DG-Beliefs	DG-Beliefs	DG-Beliefs	DG-Beliefs	Timing 3	Timing 3	DG-Beliefs	DG-Beliefs
ET-Beliefs	ET-Beliefs	ET-Beliefs	ET-Beliefs	ET-Beliefs	ET-Beliefs	Timing 4	Timing 4
ET-Task	ET-Task	ET-Task	ET-Task	ET-Task	ET-Task	ET-Task	ET-Task

Table 1: Experimental treatments

The variation in timing aims at provoking the exogenous affective activation at different stages during the experiment in order to identify its effect on elicited beliefs, choices and effort without concerns about the duration of the effect of the AIP. Moreover, it allows us to analyse the persistence of the effects generated by the music in the effort task. A significant effect of Timing 1 and/or 2 treatment on the beliefs and/or performance in the effort task, would suggest that the variation in affect is persistent. Such analysis also provides tentative information on the duration of the effect of the AIP.

3.3 Procedures

The experiment took place in 2018 and 2019 at the EssexLab of the University of Essex. The recruitment procedure followed the same practice and guidelines described for the pilot sessions. The experiment was programmed and conducted with z-tree (Fischbacher, 2007).

The research involved 364 participants, 91% were students, 55% females. We ran 14 experimental sessions consisting of 20 to 32 participants. Participants could read the instructions on the computer screen and were asked to wear headphones to listen to the music and complete the emotional ability tests.

The sessions lasted on average 86 minutes, including random assignment to cubicles, answering the post-experimental questionnaire and receiving payments. The participants earned an average of 16 pounds (including the show-up fee). See Appendix D for the schedule of the experimental sessions, Appendix E for the random assignment check and Appendix F for a discussion on the independence between the emotional intelligence measures and the treatments.

3.4 Conjectures

The theory of constructed emotions (Barrett, 2017b) posits that the primary purpose of an organism's brain is to regulate the physiological resources required to meet the organism's short and long-term needs. The energy regulation process is optimized when the brain anticipates bodily needs (Sterling, 2015).

Our AIP targets affects inducing a surge in arousal and increasing the pleasantness/unpleasantness of the bodily activation. The heightened bodily arousal should shape predictions concerning the presence or likelihood of threats that would necessitate the current bodily arousal, directing perception and action accordingly (Fridman, Barrett, Wormwood, and Quigley, Fridman et al.). How arousal and valence relate to one another is still an open question (see Kuppens, Tuerlinckx, Russell, and Barrett, Kuppens et al. for a discussion). Several studies suggest the prevalence of a V-shaped relation of arousal as a function of valence but warn about the presence of a variety of relations depending on person or circumstances (Kuppens, Tuerlinckx, Russell, and Barrett, Kuppens et al.; Barrett, 1998; Ito and Cacioppo, Ito and Cacioppo).

Our conjectures rely on existing experimental evidence on incidental affects which has traditionally focused on discrete emotions or positive/negative affects without commenting upon the role played by arousal.

3.4.1 Incidental affects and altruism

A substantial body of research in psychology shows that positive affective states promote helpfulness and altruism (for instance Isen, 1970, 1987; Isen and Levin, 1972; Weyant, 1978). The same result occurs using music to induce affect (Fried and Berkowitz, 1979). Several models have been proposed to explain the results (for a review see Carlson et al., 1988). Experimental studies that investigate the effect of incidental emotion in one-shot economic decisions confirm that individuals in a positive affective state are more altruistic and trusting (Kirchsteiger et al., 2006; Drouvelis and Grosskopf, 2016; Capra, 2004; Dunn and Schweitzer, 2005).

In contrast to what might have been expected on the intuitive ground, there is strong empirical evidence in support of the thesis that also negative affects increase altruism and helping behavior (for instance Weyant, 1978; Manucia et al., 1984; Cialdini et al., 1987, for reviews Cialdini and Fultz, 1990 and Carlson and Miller, 1987). The dominant model is the mood maintenance hypothesis (Isen et al., 1978; Batson et al., 1989) according to which people in a negative mood try to restore an equilibrium condition through positive action. Experimental economic studies mainly compare a positive affect condition with a negative one, without using a neutral baseline. People under negative affects are generally considered less altruistic and helping than subjects in a positive affective state.

However, the evidence is not uncontroversial. For instance, Proto et al. (2019) find that players with an induced positive mood tend to cooperate less than players in a neutral mood setting. Also, there are studies showing reduced helping in subjects in a negative mood (Isen, 1970; Berkowitz and Connor, 1966). Moreover, it must be noted that the extreme variety of experimental methodologies adopted leave room for different interpretations of the results. Nonetheless, our first conjecture follows the majority of the results encountered in the literature.

Conjecture 1 - Both positive and negative affects induce altruistic behavior.

3.4.2 Incidental affects and performance

People experiencing positive affects have been associated with confidence, optimism, self-efficacy, and energy (Lyubomirsky et al., 2005). Those characteristics are typically associated with success. However, the experimental evidence on incidental affects and task performance is rather mixed and context-specific (see Lyubomirsky et al., 2005 for a review). Some studies have shown that those put in a pleasant mood outperform others. For example, subjects in a positive mood solve more anagrams correctly (Erez and Isen, 2002) and perform better in a clerical coding task (Baron, 1990). Subjects in induced happy moods also appear to persist longer at tasks in which perseverance is required (Erez and Isen, 2002; Kavanagh, 1987) and show greater creativity that those in a neutral state (Isen et al., 1987). Other authors argue that positive affective states can be detrimental for performance as they tend to favour the use of heuristics, whereas negative affective states promote analytical thinking (the sadder-but-wiser hypothesis by Forgas, 1995; George and Zhou, 2002). There is indeed evidence that participants placed in a negative affective state outperformed participants in a neutral/positive state (Melton, 1995; Bless et al., 1990; Elsbach and Barr, 1999). The above-mentioned research has concentrated on unpaid experimental settings.

Economists and management scientists still know relatively little about the causal linkages between incidental affects and effort in incentivized settings. There is a general expectation that happy workers are more productive and those in a negative affective state are less productive but the evidence is scarce. In a series of experiments, Oswald et al. (2015) consistently find that happiness raises productivity by comparing subject in an induced happy state to a neutral control group. Dalton et al. (2017) find a positive correlation between negative affective states and a decrease in productivity. We believe that our experiment is likely to reproduce those results.

Conjecture 2 - Positive (negative) affects increase (decrease) performance.

3.4.3 Incidental affects and beliefs

In the psychological literature, we find many theories according to which incidental affects should influence beliefs directly or indirectly through the recalling of events from memory. According to the affect-as-information model (Schwarz and Clore, 1983), affective states have a direct effect on beliefs. In fact, people use their affective states as a sample of experience of the object of judgement. Affects are used to produce a quick judgement of a situation without trying to integrate the external characteristics with their own memory and internal associations. For instance, if people feel good about something they are more likely to believe in a good outcome happening. Bandura (1977) stresses the importance of mood congruence in memory activation. The emotional state activates memories consistent with the affect: positive emotions facilitate the recall of successful experiences, negative emotion of failure.

Some authors link affective states to belief management strategies. Isen and Simmonds (1978) introduce the mood maintenance hypothesis - people in a positive affective state try to maintain their good-mood states by recalling positive material in memory or making actions that make them feel good. Cialdini et al. (1973) propose the negative state relief hypothesis suggesting that people in negative affective states might try to change their emotional state by adopting behaviors similar to those of people in a positive affective state. Weber (1994) argue that people might strategically overestimate the likelihood of a negative outcome as a way of protecting themselves from the emotional reaction should the negative outcome materialize (Shepperd et al., 2000; Risen and Gilovich, 2007). In fact, unexpected outcomes have a greater emotional impact than expected ones (Mellers et al., 1997).

There is a lack of studies on the influence of incidental affects on beliefs in incentivized settings. To the best of my knowledge, the only one is Proto et al. (2019). The authors elicit participants beliefs in a prisoner's dilemma and find that participants in the neutral mood treatment have more accurate beliefs about their partners' choices compared to participants that were induced a positive mood. Given the scarcity of experimental evidence and the conflicting predictions of the theories, we do not make conjecture concerning how incidental affect might influence beliefs. However, we do claim that beliefs might mediate choices in the dictator game - as suggested by Psychological Game theory (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2007) - and performance in the effort task - as suggested by motivational theory (Bandura and Cervone, 1986; Cervone, 1993; Locke and Latham, 1990) and reference- dependent models (Köszegi, 2006).

Conjecture 3 - If incidental affects significantly influence beliefs, the variation should partly explain the influence of incidental affects on altruism and performance.

3.4.4 Incidental affects and heterogeneity

People might react differently to exogenous affective activations and heterogeneity might be one of the causes of the mixed evidence encountered in the literature. Yip and Côté (2013) argue that the ability to read emotional states might help in understanding the biasing effects that incidental affects exert on cognition. Following this line of reasoning, our conjecture is that people with low emotional intelligence could be more influenced by incidental affects and the mechanism underlying this effect involves pinpointing the correct source of the feeling and, in turn, becoming aware that incidental affects are irrelevant to decisions at hand. On the other hand, higher levels of emotional intelligence should attenuate the impact of incidental affect on decisions and beliefs.

Conjecture 4 - The influence of incidental affects on choices, performance and beliefs is heterogeneous across participants with different level of emotional intelligence. Participants with high emotional intelligence should not be biased.

4 Data Analysis

4.1 Descriptives

Table 2 reports summary statistics. In the dictator game, nearly 45% of the dictators choose the selfish option (7,3), all the others (except 1) the fair split (5, 5). Almost half of the recipients expect the dictators to make the selfish choice; almost half of the dictators believe the recipients expect them to be selfish.

Variable name	Range	Mean	SD	Min	Max
Dictator game					
Dictator's choice (discrete)	1-3	1.55	0.50	1	3
$FOB_{(7,3)}$: recipients' beliefs on dictators choosing (7,3)	0-1	0.49	0.22	0.05	1
$SOB_{(7,3)}$: dictators' beliefs on $FOB_{(7,3)}$	0-1	0.49	0.19	0	1
Effort task					
Beliefs on performing in the top quartile	0-1	0.45	0.28	0	1
Beliefs on performing in the second best quartile	0-1	0.33	0.19	0	1
Number of correct tables completed	0-26	13.66	4.09	1	26
Number of attempted tables	1-40	16.97	3.96	3	29
Emotional 'ability'	0-16	11.21	2.34	5	16
Trait emotional intelligence	0-7	4.96	0.75	2.7	6.53

Table 2: Summary statistics. Dictator's choice coding is as follows: 1 corresponds to (7,3), 2 to (5,5), and 3 to (3,7).

In the effort task, participants are on average overconfident. Nearly 45% of the participants believe they would perform in the top 25%, and more than 75% in the top half. On average, participants completed a little less than 14 tables correctly out of almost 17 attempted. Only 3 participants completed less than 3 tables in 4 minutes.

On average, participants correctly answered 11 questions of the emotional 'ability' task, with a standard deviation of 2.34. The trait emotional intelligence average score lies very near to the values obtained in other studies that used the TEIQue-SF with representative samples (Cooper and Petrides, 2010). High scores in the test indicate that the person views themselves as flexible in their approach to life and willing and able to adapt to new environments and conditions. To the contrary, low scorers perceive themselves as change-resistant and have fixed ideas and views. Additional descriptive statistics are available in Appendix G. In Appendix H, we explore the correlation between emotional intelligence and demographics and find no systematic correlation.

4.2 Average treatment effect

We estimate the following equation:

$$Y_i = \alpha + \beta \text{Treatment}_i + \gamma_1 X_{i,1} + \ldots + \gamma_M X_{i,M} + \epsilon_i \tag{4}$$

where Y_i is one of the variables of interest, Treatment_i is a discrete variable indicating the treatment where the excluded category is the baseline, and $X_{i,1} \ldots X_{i,M}$ are control variables.

We use an Ordinary Least Squares specification with robust standard errors and control for age, gender, number of siblings, education, cognitive skills, and risk attitudes. Table 3 reports the raw coefficients of the regression analysis. The dependent variables are as follows: in column (1) a dummy where 1 corresponds to the selfish choice[9]; in column (2) dictator's beliefs on the percentage of recipients expecting them to be selfish - $\text{SOB}_{(7,3)}$; in column (3) recipients beliefs on the percentage of dictators being selfish - $\text{FOB}_{(7,3)}$; in column (4) participants' beliefs on performing in the top quartile of the ranking in the effort task - Beliefs Top25; in column (5) the number of correct tables solved.

The analysis indicates that altruism is significantly affected by positive AIP. Dictators in treatment Positive - Timing 1 are 25% more likely to make the selfish choice compared to those in the baseline. Dictators' $\text{SOB}_{(7,3)}$ are also marginally affected by that treatment - dictators that listened to the positive music before making their choice believe that a higher proportion of recipients expect them to be selfish

[9]The results of the regression in (1) do not change using a Probit or Logit specification.

compared to the baseline. All the other coefficients are not significantly different from zero. It is worth noting that, in general, positive and negative AIPs do not produce opposite effects on the variables under analysis.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Selfish choice	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25	Correct tables
Timing of treatment:	Timing 1	Timing 2	Timing 2	Timing 3	Timing 4
Negative	0.0239	0.0185	-0.0416	0.0345	-1.3427
	(0.1059)	(0.0499)	(0.0530)	(0.0660)	(1.1563)
Positive	0.2484***	0.0823*	-0.0592	-0.0436	-0.2836
	(0.1040)	(0.0432)	(0.0451)	(0.0543)	(1.0244)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	181	181	183	364	364
R-squared	0.0578	0.0418	0.1008	0.0686	0.0814

Table 3: OLS Regressions - Average Treatment Effect. Robust standard errors in parentheses. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes. Note: the baseline for the treatment in column (1) encompasses subjects in the neutral condition, treatment at Timing 2, 3 and 4; in column (2) and (3) neutral condition, treatment at Timing 3 and 4; in column (4) neutral condition and treatment at Timing 4; in column (5) neutral condition only.

Result 1 - On average, positive affects cause more selfish behavior. Negative affects have no impact on altruistic behavior.

Result 2 - On average, incidental affects have no impact on effort.

Result 3 - On average, incidental affects have no impact on beliefs, nor in the dictator game nor in the effort task.

4.3 Incidental affects and heterogeneity

To test our conjecture that emotional intelligence can be relevant for the way people manage their beliefs and make choices or perform under affective activation, we conduct a heterogeneity analysis using the measure of emotional intelligence introduced in section 3. We estimate:

$$Y_i = \alpha + \beta_1 \text{Treatment}_i + \beta_2 \text{Treatment}_i \times \text{TEIQue-SF}_i + \beta_3 \text{TEIQue-SF}_i + \gamma_1 X_{i,1} + \ldots + \gamma_M X_{i,M} + \epsilon_i \quad (5)$$

where TEIQue-SF is the centred measure of trait emotional intelligence (Petrides, 2009). The rest of the equation is as described in section 4.2. The analysis using the emotional ‘ability’ approach does not yield significant results and it is presented in Appendix I.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Selfish choice	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25	Correct tables
Timing of treatment:	Timing 1	Timing 2	Timing 2	Timing 3	Timing 4
Negative × TEI	-0.0311	-0.0580	0.1084*	-0.2398**	1.0033
	(0.1343)	(0.0522)	(0.0603)	(0.1036)	(1.5207)
Positive × TEI	0.2000*	-0.1047**	0.0201	-0.1465**	-1.0168
	(0.1165)	(0.0512)	(0.0678)	(0.0714)	(1.7374)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	181	181	183	364	364
R-squared	0.0859	0.0846	0.1467	0.0943	0.1036

Table 4: OLS Regressions - Heterogeneity. Robust standard errors in parentheses. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes. Note: the baseline for the treatment in column (1) encompasses subjects in the neutral condition, treatment at Timing 2, 3 and 4; in column (2) and (3) neutral condition, treatment at Timing 3 and 4; in column (4) neutral condition and treatment at Timing 4; in column (5) neutral condition only.

The results in table 4 suggest that participants with different levels of trait emotional intelligence might manage their beliefs differently in the presence of an exogenous affective activation. Interestingly, both AIPs move beliefs in the same direction suggesting that the shift in arousal (and not valence) might be driving the results. We thus merge treatments with opposite valence and same timing to gain statistical power. Moreover, we split the sample into three percentile according to the distribution of the TEIQue-SF scores and run the regressions for beliefs on those tertiles. Table 5 reports the results for the top and bottom tertiles.

	Bottom TEIQue-SF tertile			Top TEIQue-SF tertile		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25
Timing 1	-0.0235	-0.0431	0.0796	0.0263	0.0535	-0.1704**
	(0.0656)	(0.0685)	(0.0722)	(0.0625)	(0.0544)	(0.0708)
Timing 2	0.0790	-0.1136*	0.0482	-0.1229*	0.1953***	-0.1444**
	(0.0624)	(0.0662)	(0.0673)	(0.0645)	(0.0656)	(0.0730)
Timing 3			0.1543*			-0.2982***
			(0.0848)			(0.0722)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56	69	125	57	59	116
R-squared	0.1134	0.2162	0.1445	0.1762	0.3204	0.1755

Table 5: OLS Regressions - Sub-samples. Robust standard errors in parentheses. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes. Note: the baseline for the treatment in column (1), (2), (4), and (5) encompasses subjects in the neutral condition, treatment at Timing 3 and 4; in column (3) and (6) neutral condition and treatment at Timing 4.

The analysis suggests that the participants with high scores in TEIQue-SF (columns (3) to (6)) are significantly affected by the AIP both in the dictator game and effort task. The recipients believe that an additional 20% of the dictators made the selfish option compared to the baseline; the dictators believe that -12% of the recipients expect them to be selfish; and the participants in the effort task believe they are almost 30% less likely to perform in the top quartile. The results for the recipients and beliefs in the effort task are highly significant, at least when the AIP is delivered right before the belief elicitation (Timing 2 for $FOB_{(7,3)}$ and Timing 3 for beliefs in the effort task). For participants with low scores in TEIQue-SF, the AIP generates at most weak effects.

The effect of the AIP seems to have little persistence: treatment at timing 1 does not have any significant effect on beliefs in the dictator game, and treatment at timing 1 and 2 generate a smaller effect compared to treatment at timing 3 on beliefs in the effort task. The analysis has been repeated considering TEIQue-SF score quartiles, and considering different sub-samples of participants and the results are substantially confirmed (see Appendix J.1 and Appendix J.2).

In figure 5, we plot average beliefs by score in the TEIQue-SF. We observe a clear pattern for participants with high trait emotional intelligence for which the AIP induces a shift with respect to the baseline, confirming the results of the regression. In the following section, we explore three possible mechanisms to explain the observed pattern.

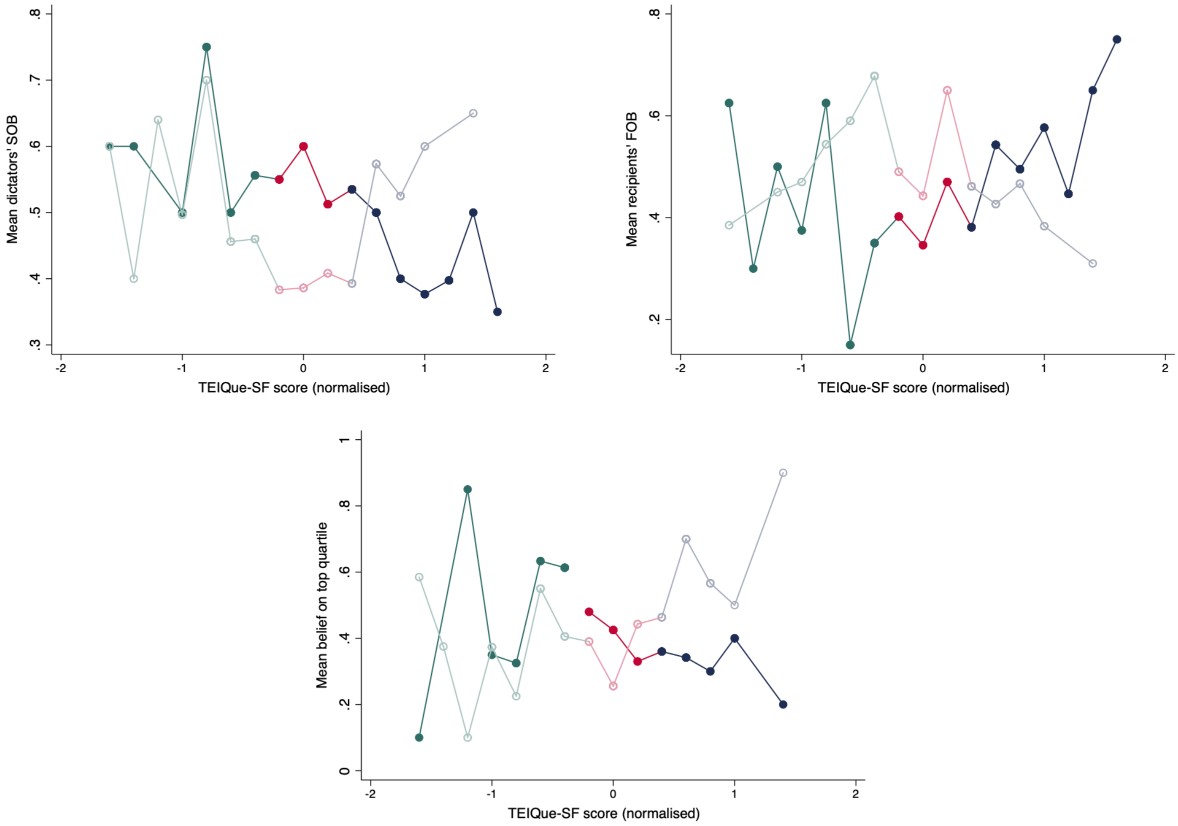


Figure 5: Beliefs in the dictator game and effort task depending on the score in the TEIQue-SF. Green lines/indicators refer to the bottom TEIQue-SF percentile, red to the middle, blue to the top tertile. Light coloured lines with circle hollow markers show baseline, bright lines and circle markers treated. The top left panel shows mean dictators' SOB (treatment=Timing 2), the top right panel shows mean recipients' FOB (treatment=Timing 2), and the bottom panel shows beliefs of performing in the top quartile in the effort task (treatment=Timing 3).

Result 4 - Beliefs of participants with high trait emotional intelligence are significantly affected by incidental affects.

5 Possible mechanisms

5.1 Beliefs

In the previous section, we show that the AIP induces significant variation in beliefs for participants scoring high in TEIQue-SF. We explore three possible mechanisms.

5.1.1 Delaying

The music introduces a 4-minute delay between the belief elicitation instructions and the response. It might be this delay and not the affective activation, that drives our results. This hypothesis would be consistent with the fact that positive and negative AIPs generate effects in the same direction.

Dictator game - Table 6 reports average beliefs split by response time and TEIQue-SF tertiles. Fast response indicates participants whose response time is in the top tertile of the response time distribution, slow response in the bottom. On average, slower response times are associated with beliefs nearer to the equilibrium of the game - recipients expect dictators to be more selfish and dictators believe that more recipients expect them to be selfish. However, we also notice that the difference is significant using the whole sample and mainly consistent across TEIQue-SF tertiles. It is thus unlikely for delay to be the main drive behind the results presented in tables 4 and 5.

	Whole sample		Bottom TEI tertile		Top TEI tertile	
	$SOB_{(7,3)}$	$FOB_{(7,3)}$	$SOB_{(7,3)}$	$FOB_{(7,3)}$	$SOB_{(7,3)}$	$FOB_{(7,3)}$
Fast response	.459	.435	.416	.464	.506	.411
Slow response	.592	.568	.575	.636	.484	.545
Difference	.070**	.133***	.159***	.172**	-.021	.134**
Observations	125	121	42	44	37	40

Table 6: Dictator game - Average beliefs by response time. Stars refer to two-sided t-test of difference in means. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Effort task - We conduct a similar analysis for beliefs about own performance in the effort task. Slower response times are associated with confidence and the effect is consistent across TEIQue-SF tertiles, although stronger for participants with low trait emotional intelligence. In our main analysis, we find that the AIP generates *less* confidence in high TEIQue-SF scorers (see table 5) suggesting that the AIP should not be considered a simple delay in response.

	Whole sample		Bottom TEI tertile		Top TEI tertile	
Belief of scoring:	Top 25%	Top 50%	Top 25%	Top 50%	Top 25%	Top 50%
Fast response	.397	.748	.325	.706	.477	.772
Slow response	.543	.834	.581	.845	.550	.855
Difference	.146***	.086***	.256***	.138***	.073	.083*
Observations	243	243	79	79	82	82

Table 7: Effort task - Average beliefs by response time. Stars refer to two-sided t-test of difference in means. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

5.1.2 Rational anticipation

Are participants with high trait emotional intelligence best responding to how they expect the exogenous affective activation has influenced the decision/performance of others?

Dictator game - If we consider the experimental timeline outlined in table 1, we notice that Timing 2 treatments are happening *after* dictators make their choices and this is common knowledge for participants. As a consequence, the change that we observe in recipients' beliefs cannot be driven by them anticipating that dictators become more selfish because of the AIP. And dictators, knowing that recipients have no reason to change their first order beliefs, should keep their second order beliefs unchanged. Moreover, if participants were trying to anticipate the effect of the AIP on the counterpart, then Timing 1 (and not Timing 2) treatments should be a significant predictor of dictator's SOB. However, the regressions in table 5 shows that coefficients for Timing 1 are not significantly different from zero suggesting that the shift in beliefs is not due to a rational anticipation mechanism.

Effort task - The results reported in table 5 suggest that participants with high trait emotional intelligence are less confident after the AIP compared to the baseline. They might be anticipating that they will do worse in the task compared to other participants. If rational anticipation is at play, beliefs regarding whether other participants listened to the same music should be relevant as the ranking depends on others' performance. In the post-experimental questionnaire, we ask participants if they believe that everybody in the lab listened to the same music during the experiment. Table 8 reports average confidence in the effort task by belief regarding other participants' AIP. The variable seems to play no role for confidence in the effort task. We thus think that rational anticipation is unlikely to be the mechanism driving our results.

Belief of scoring:	Whole sample Top 25%	Top 50%	Bottom TEI tertile Top 25%	Top 50%	Top TEI tertile Top 25%	Top 50%
Others listened to same music	.471	.796	.564	.836	.394	.750
Others did not listen to same music	.393	.742	.494	.800	.275	.700
Difference	.078	.054	.070	.036	.119	.050
Observations	51	51	20	20	16	16

Table 8: Average beliefs by beliefs on others' AIP. Stars refer to two-sided t-test of difference in means. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

5.1.3 Strategic pessimism

Some authors in the psychological literature argue that people might strategically overestimate the likelihood of a negative outcome (Weber, 1994) as a way of protecting themselves from the emotional reaction should the negative outcome materialize (Shepperd et al., 2000; Risen and Gilovich, 2007). In fact, unexpected outcomes have a greater emotional impact than expected ones (Mellers et al., 1997). Importantly, strategic pessimism would depend on how aroused a decision-maker is. Because arousal intensifies emotional reactions, aroused decision-makers should overestimate the likelihood of a negative outcome to a greater extent than non-aroused decision-makers (Vosgerau, 2010). In the context of our experiment, strategic pessimism would predict that high TEIQue-SF scorers would overestimate the negative outcome in the game since they are the ones that react more to the exogenous affective activation (Furnham and Petrides, 2003).

Dictator game - For the recipients, the negative outcome is the selfish choice of the dictators as this generates the lowest payoff for them. Hence, strategic pessimism would predict an increase in $\text{FOB}_{7,3}$. The sign of the coefficient in column (5) of table 5 is consistent with what the theory would predict.

For the dictators, we assume that they can be averse to guilt (Battigalli and Dufwenberg, 2007). As such, they bear a psychological cost if they disappoint their partner (i.e., if they keep more money than their partner expects). For guilt averse dictators, the payoffs associated with the available options are:

$$\pi_{(7,3)} = 7 - \theta_i[\text{SOB}_{(5,5)}(5-3) + \text{SOB}_{(3,7)}(7-3)]$$
$$\pi_{(5,5)} = 5 - \theta_i[\text{SOB}_{(3,7)}(7-5)]$$
$$\pi_{(3,7)} = 3$$

where θ_i is guilt sensitivity - an individual time-invariant characteristic - and the terms in square brackets correspond to the expected disappointment of the counterpart following each of the choice. If a dictator makes the selfish choice, it has to be that $\pi_{(7,3)} \geq \pi_{(5,5)}$ and $\pi_{(7,3)} \geq \pi_{(3,7)}$. Thus, a 'bad news' corresponds to recipients expecting them to be more altruistic as this would increase the psychological cost of guilt (or change the ranking of the payoffs). Strategic pessimism would thus predict an increase in $\text{SOB}_{(5,5)}$ and/or $\text{SOB}_{(3,7)}$. If we consider dictators choosing the fair split, their guilt would increase with higher $\text{SOB}_{(3,7)}$. In table 9 we compare $\text{SOB}_{(5,5)}$ and $\text{SOB}_{(3,7)}$ in treatment Timing 2 to the baseline across dictators' choices and TEIQue-SF tertiles.

	Bottom TEI tertile			Top TEI tertile		
Dictator choice:	Selfish		Altruistic	Selfish		Altruistic
Average beliefs:	$\text{SOB}_{(5,5)}$	$\text{SOB}_{(3,7)}$	$\text{SOB}_{(3,7)}$	$\text{SOB}_{(5,5)}$	$\text{SOB}_{(3,7)}$	$\text{SOB}_{(3,7)}$
Baseline	.279	.117	.166	.298	.087	.142
Timing 2	.271	.064	.109	.289	.139	.177
Difference	-.008	-.053	-.057	-.009	.052*	.036
Observations	21	21	35	28	28	29

Table 9: Average dictators' SOB by response time. Stars refer to one-sided t-test of difference in means. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

The results are weakly consistent with the strategic pessimism hypothesis. The $\text{SOB}_{(5,5)}$ remain statistically unchanged, regardless of the dictators' decision and TEIQue-SF score. For dictators in the bottom TEI tertile, we see that beliefs $\text{SOB}_{(3,7)}$ tend to decrease after the treatment compared to the control condition. To the contrary, for dictators in the top TEI tertile, we find that dictators have higher $\text{SOB}_{(3,7)}$ after the treatment. The effect, however, is only marginally significant. It is worth noticing that participants with high trait emotional intelligence are also the ones more sensitive to guilt as we observe a significant positive correlation between TEIQue-SF scores and the two scales of the GASP questionnaire - negative behavior evaluation (Pearson's correlation=0.1085, p-value<0.05, sample size=364) and repair responses to private transgressions (Pearson's correlation=0.2385, p-value<0.001, sample size=364).[10] Participants in the top tertile have a significantly higher score in GASP negative behavior evaluation (difference in means=0.3991, p-value<0.05) and GASP repair (difference in means=0.5951, p-value<0.001) compared

[10] Guilt negative behavior evaluation items describe feeling bad about how one acted. Guilt–repair items describe action tendencies (i.e., behavior or behavioral intentions) focused on correcting or compensating for the transgression.

to participants in the bottom tertile.

Effort task - Strategic pessimism would predict a decrease in the probability of performing in the first and second quartile for high TEIQue-SF scorers. The results presented in table 5 support this hypothesis as we notice a significant decrease in the probability attributed to scoring in the top quartile in the task for high TEIQue-SF scorers. We repeated the regression using beliefs on scoring in the top half of the distribution and the coefficient of Timing 3 is still significant (coefficient Timing 3=-13.4959, p-value<0.05).

In conclusion, we believe that the reaction to the AIP of participants with high trait emotional intelligence might be due to strategic pessimism.

5.2 Choices

Dictator game - According to guilt aversion theories (Battigalli and Dufwenberg, 2007), the psychological cost of guilt associated with the available options (and thus the final payoffs) depends on the dictators' SOB. The design of the experiment allows us to use Timing 1 treatment to test whether dictators' SOBs are a 'mediating variable' (Mackinnon et al., 2007; Imai et al., 2010) for dictators' choices. We estimate the following system of equations:

$$\mathrm{SOB}_i = \alpha_3 + \tau \mathrm{Timing1}_i + \gamma_{3,1} X_{3,1,i} + \ldots + \gamma_{3,M} X_{3,M,i} + \epsilon_{3,i} \tag{6}$$

$$Y_i = \alpha_2 + \beta_2 \mathrm{Timing1}_i + \delta \mathrm{SOB}_i + \gamma_{2,1} X_{2,1,i} + \ldots + \gamma_{2,M} X_{2,M,i} + \epsilon_{2,i} \tag{7}$$

where Y_i is the dictator's choice (1=selfish choice, 0 otherwise), SOB the mediating variable. The estimates of this system of equations intend to isolate two effects. The first is given by $\hat{\beta}_2$, which measures the direct effect of treatment on Y_i. This corresponds to the effect on choices of being assigned to treatment Timing 1 once the relationship between variations in beliefs and choices is accounted for. The second effect is the indirect effect of treatment on choices via changes in beliefs $\hat{\mu} = \hat{\tau}\hat{\delta}$. The total effect of treatment is the sum of the direct and indirect effects. We use the Quasi-Bayesian Monte Carlo approximation of King et al. (2000) to make statistical inference about $\hat{\mu}$. The parameters reported are the average of 100 draws.

	Whole sample		Bottom TEIQue-SF percentile		Top TEIQue-SF percentile	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Selfish choice Mediator: Dictators' SOB						
$\hat{\gamma}$ (Mediation effect)	0.0576	0.0575	-0.0298	-0.2318	0.0471	0.0093
$\hat{\beta}_2$ (Direct effect)	0.1797*	0.1865*	0.2824	0.2363	0.3869**	0.4590**
Total effect	0.2385**	0.2440**	0.2527	0.2131	0.4340**	0.4683**
Controls		Yes		Yes		Yes
Observations	123	123	38	38	37	37

Table 10: Mediation analysis - Dictator game

Table 10 reports the results of the analysis using the whole sample and the top and bottom tertile of the TEIQue-SF score distribution. We observe a significant direct effect of the Timing 1 treatment on dictator's choice driven by the top percentile. The coefficient of the mediation effect is in the expected direction but does not reach statistical significance. It is worth emphasizing that the variation induced by the AIP Timing 1 on dictators' SOB is weak (see table 3 and 5) and the regression on the sub-sample relies on a small number of observations making it difficult to precisely identify a mediation effect.

Result 5 - We do not find evidence of a mediation effect of beliefs on choices in the dictator game.

Effort task - Various authors have argued that confidence might act as a motivational boost. Bénabou and Tirole (2002) stress the importance of self-confidence as it improves individuals' motivation to undertake projects and persevere in the pursuit of goals. The link between self-confidence and motivation is also pervasive in the psychology literature (see, for instance, James, 1890; Bandura, 1977; Deci, 1975). In motivation research, expectations that fall below one's goals lead to reduced effort, goal abandonment, and lower achievement (Bandura and Cervone, 1986; Cervone, 1993; Locke and Latham, 1990). Confidence has been shown to be an important outcome explaining leadership among young adults (Alan and Ertac, 2019), intergenerational income mobility (Blanden et al., 2007) and academic performance of students (Golsteyn et al., 2017).

Our experimental design allows us to run a mediation analysis and estimate the indirect impact that a reduction in confidence has on performance. We estimate equations 6 and 7 using the number of correct table as Y_i and beliefs on performing in top quartile as the mediating variable.

	Whole sample		Bottom TEIQue-SF percentile		Top TEIQue-SF percentile	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Correct tables						
Mediator: Beliefs on top quartile						
$\hat{\gamma}$ (Mediation effect)	-0.0461	-0.0450	0.1489	0.1477	-0.8972*	-1.1970**
$\hat{\beta}_2$ (Direct effect)	0.0041	-0.0031	0.2186	-0.0277	-0.3367	-0.3383
Total effect	-0.0420	-0.0481	0.3676	0.1200	-1.2340**	-1.5353**
Controls		Yes		Yes		Yes
Observations	79	79	29	29	26	26

Table 11: Mediation analysis - Effort task

Table 11 reports the result of the estimation. We observe a significant mediation effect only for participants scoring high in TEIQue-SF. The mediation effect corresponds to a decrease of a little more than 1 correct table solved in the game. Considering that the average number of tables that high TEIQue-SF scorers solve correctly is 13.08, the decrease in confidence induced by the AIP generates a reduction of 9% in performance.

Result 6 - We find evidence of a mediation effect of confidence on performance in the effort task for participants scoring high in TEIQue-SF. The mediation effect corresponds to a reduction of 9% in performance.

6 Conclusions

This research uses a lab experiment to explore how incidental affects influence beliefs management and decision-making in two economically relevant situations - a bargaining game and an effort task. The analysis reveals that incidental affects make dictators significantly more selfish. Our treatments also show that participants with high trait emotional intelligence strategically manage their beliefs when under the influence of an exogenous affective activation.

Our contribution is empirical as well as methodological. We design and test a novel affect induction procedure with minimal demand effect or priming potential, and strong efficacy. We compare decisions and beliefs under the effect of incidental affects with a baseline where participants' affective states remain untouched. Our experimental design allows us to separately study the effect of a change in valence combined with an increase in arousal on belief management and decision-making. To a large extent, the results of our experiment *disconfirm* the existing evidence on the topic and point to a note of caution in the study of emotions and decision-making as many factors can confound the results. We believe that our results are particularly robust as our design is minimizing such confounding.

Our results point to the capacity, for a substantial proportion of people, of strategically managing their beliefs in an attempt to cope with the emotional consequences of the choices that they and others make. Our study provides evidence that beliefs do not always reflect the individuals' knowledge concerning the state of the world and strategic pessimism might be one of the mechanisms through which people form beliefs. Finally, our study lends support to belief-dependent theories through a novel experimental design and by using a mediation analysis to show how altruism and performance can be significantly affected by beliefs.

References

Abeler, B. J., A. Falk, L. Goette, and D. Huffman (2016). Reference Points and Effort Provision. *The American Economic Review 101*(2), 470–492.

Alan, S. and S. Ertac (2019). Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment. *Journal of the European Economic Association 17*(4), 1147–1185.

Bandura, A. (1977). Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review 84*(2), 191–215.

Bandura, A. and D. Cervone (1986). Differential Engagement of Self-Reactive Cognitive Motivation. *Organizational Behavior & Human Decision Processes 38*, 92–113.

Baron, R. A. (1990). Environmentally Induced Positive Affect: Its Impact on Self-Efficacy, Task Performance, Negotiation, and Conflict. *Journal of Applied Social Psychology 20*(5), 368–384.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning.

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*(2), 241–251.

Barrett, L. F. (1998). Discrete emotions or dimensions? the role of valence focus and arousal focus. *Cognition and Emotion 12*(4), 579–599.

Barrett, L. F. (2017a). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization. *Social Cognitive and Affective Neuroscience 12*(1), 1–23.

Barrett, L. F. and J. A. Russell (1998). Independence and Bipolarity in the Structure of Current Affect. *Journal of Personality and Social Psychology 74*(4), 967–984.

Batson, C. D., J. G. Batson, C. A. Griffitt, S. Barrientos, J. R. Brandt, P. Sprengelmeyer, and M. J. Bayly (1989). Negative-State Relief and the Empathy-Altruism Hypothesis. *Journal of Personality and Social Psychology 56*(6), 922–933.

Battigalli, P. and M. Dufwenberg (2007). Guilt in Games. *American Economic Review 97*(2), 170–176.

Battigalli, P. and M. Dufwenberg (2009). Dynamic Psychological Games. *Journal of Economic Theory 144*(1), 1–35.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3), 871–915.

Berkowitz, L. and W. H. Connor (1966). Success, Failure, and Social Responsibility. *Journal of Personality and Social Psychology 4*(6), 664–669.

Bigoni, M., M. Casari, A. Skrzypacz, and G. Spagnolo (2015). Time Horizon and Cooperation in Continuous Time. *Econometrica 83*(2), 587–616.

Blanden, J., P. Gregg, and L. Macmillan (2007). Accounting for intergenerational income persistence: noncognitive skills, ability and education. *The Economic Journal 117*(519), C43–C60.

Bless, H., G. Bohner, N. Schwarz, and F. Strack (1990). Mood and Persuasion. *Personality and Social Psychology Bulletin 16*(2), 331–345.

Bosman, R. and A. Riedl (2003). Emotions and economic shocks in a first-price auction. Technical report, Tinbergen Institute discussion paper.

Bower, G. H. (1981). Mood and Memory. *American Psychologist 36*(2), 129–148.

Bower, G. H., S. G. Gilligan, and K. P. Monteiro (1981). Selectivity of Learning Caused by Affective States. *Journal of Experimental Psychology: General 110*(4), 451–473.

Bradley, M. M. and P. J. Lang (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Different. *Journal of Behavior Therapy and Experimental Psychiatry 25*(1), 49–59.

Capra, C. M. (2004). Mood-Driven Behavior in Strategic Interactions. *The American Economic Review 94*(2), 367–372.

Capra, C. M., K. F. Lanier, and S. Meer (2010). The Effects of Induced Mood on Bidding in Random Nth-Price Auctions. *Journal of Economic Behavior and Organization 75*(2), 223–234.

Carlson, M., V. Charlin, and N. Miller (1988). Positive Mood and Helping Behavior: A Test of Six Hypotheses. *Journal of Personality and Social Psychology 55*(2), 211–229.

Carlson, M. and N. Miller (1987). Explanation of the Relation Between Negative Mood and Helping. *Psychological Bulletin 102*(1), 91–108.

Carroll, J. M., M. S. Yik, J. A. Russell, and L. F. Barrett (1999). On the Psychometric Principles of Affect. *Review of General Psychology 3*(1), 14–22.

Cervone, D. (1993). The Role of Self-Referent Cognitions in Goal Setting, Motivation, and Performance. In *Cognitive Science Foundations of Instruction* (M. Rabinow ed.)., pp. 57–95. Hillsdale, NJ: Erlbaum.

Charness, G. and B. Grosskopf (2001). Relative Payoffs and Happiness: An Experimental Study. *Journal of Economic Behavior and Organization 45*(3), 301–328.

Cialdini, R. B., B. L. Darby, and J. E. Vincent (1973). Transgression and Altruism: A Case for Hedonism. *Journal of Experimental Social Psychology 9*(6), 502–516.

Cialdini, R. B. and J. Fultz (1990). Interpreting the Negative Mood-Helping Literature via 'Mega'-Analysis: A Contrary View. *Psychological Bulletin 107*(2), 210–214.

Cialdini, R. B., M. Schaller, D. Houlihan, K. Aips, J. Fultz, and A. L. Beaman (1987). Empathy-Based Helping: Is It Selflessly or Selfishly Motivated? *Journal of Personality and Social Psychology 52*(4), 749–758.

Clark, D. M. (1983). On the Induction of Depressed Mood in the Laboratory: Evaluation and Comparison of the Velten and Musical Procedures. *Advances in Behaviour Research and Therapy 5*(1), 27–49.

Cohen, T. R., S. T. Wolf, A. T. Panter, and C. A. Insko (2011). Introducing the GASP Scale: A New Measure of Guilt and Shame Proneness. *Journal of Personality and Social Psychology 100*(5), 947–966.

Cooper, A. and K. V. Petrides (2010). A Psychometric Analysis of the Trait Emotional Intelligence Questionnaire-Short Form (TEIQue-SF) Using Item Response Theory. *Journal of Personality Assessment 92*(5), 449–457.

Dalton, P. S., V. H. Jimenez, and C. N. Noussair (2017). Exposure to Poverty and Productivity. *PLoS ONE 12*(1), 1–19.

Deci, E. L. (1975). *Intrinsic Motivation*. New York, NY, US.: Plenum Press.

DeSteno, D., R. E. Petty, D. T. Wegener, and D. D. Rucker (2000). Beyond valence in the perception of likelihood: The role of emotion specificity. *Journal of Personality and Social Psychology 78*(3), 397–416.

Diener, E. and R. A. Emmons (1984). The Independence of Positive and Negative Affect. *Journal of Personality and Social Psychology 47*(5), 1105–1117.

Drichoutis, A. and R. Nayga (2013). Eliciting Risk and Time Preferences Under Induced Mood States. *The Journal of Socio-Economics 45*, 18–27.

Drouvelis, M. and B. Grosskopf (2016). The Effects of Induced Emotions on Pro-Social Behaviour. *Journal of Public Economics 134*, 1–8.

Dunn, J. R. and M. E. Schweitzer (2005). Feeling and Believing: The Influence of Emotion on Trust. *Journal of Personality and Social Psychology 88*(5), 736–748.

Eckel, C. C. and P. J. Grossman (2002). Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. *Evolution and Human Behavior 23*(4), 281–295.

Elsbach, K. D. and P. S. Barr (1999). The Effects of Mood on Individuals' Use of Structured Decision Protocols. *Organization Science 10*(2), 181–198.

Erez, A. and A. M. Isen (2002). The Influence of Positive Affect on the Components of Expectancy Motivation. *Journal of Applied Psychology 87*(6), 1055–1067.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics 10*(2), 171–178.

Forgas, J. P. (1995). Mood and Judgment: The Affect Infusion Model (AIM). *Psychological Bulletin 117*(1), 39–66.

Forgas, J. P. and G. H. Bower (1987). Mood Effects on Person-Perception Judgments. *Journal of Personality and Social Psychology 53*(1), 53–60.

Fridman, J., L. F. Barrett, J. B. Wormwood, and K. S. Quigley. Applying the theory of constructed emotion to police decision making. *Frontiers in Psychology 10*

Fried, R. and L. Berkowitz (1979). Music Hath Charms... And Can Influence Helpfulness. *Journal of Applied Social Psychology 9*(3), 199–208.

Furnham, A. and K. V. Petrides (2003). Trait Emotional Intelligence: Behavioural Validation in Two Studies of Emotion Recognition and Reactivity to Mood Induction. *European Journal of Personality 57*(August 2002), 39–57.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior 1*(1), 60–79.

George, J. M. and J. Zhou (2002). Understanding When Bad Moods Foster Creativity and Good Ones Don't: The Role of Context and Clarity of Feelings. *Journal of Applied Psychology 87*(4), 687.

Gerrards-Hesse, A., C. Spies, and F. W. Hesse (1994). Experimental Inductions of Emotional States and Their Effectiveness: A Review. *British Journal of Psychology 85*, 55–78.

Golan, O., S. Baron-Cohen, J. J. Hill, and Y. Golan (2006). The 'Reading the Mind in Films' Task: Complex Emotion Recognition in Adults With and Without Autism Spectrum Conditions. *Social Neuroscience 1*(2), 111–123.

Golan, O., S. Baron-Cohen, J. J. Hill, and M. D. Rutherford (2007). The 'Reading the Mind in the Voice' Test-Revised: A Study of Complex Emotion Recognition in Adults With and Without Autism Spectrum Conditions. *Journal of Autism and Developmental Disorders 37*(6), 1096–1106.

Goldberg, D. and P. Huxley (1980). *Mental Illness in the Community: The Pathway to Psychiatric Care London*. New York: Tavistock.

Golsteyn, B., A. Non, and U. Zölitz (2017). The impact of peer personality on academic achievement. *University of Zurich, Department of Economics, Working Paper* (269).

Greenwald, A. G., D. E. McGhee, and J. L. K. Schwartz (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology 74*(6), 1464–1480.

Ifcher, J. and H. Zarghamee (2011). Happiness and Time Preference: The Effect of Positive Affect in a Random-Assignment Experiment. *The American Economic Review 101*(7), 3109–3129.

Imai, K., L. Keele, and D. Tingley (2010). A General Approach to Causal Mediation Analysis. *Psychological Methods 15*(4), 309–334.

Isen, A. M. (1970). Success, Failure, Attention, and Reaction to Others: The Warm Glow of Success. *Journal of Personality and Social Psychology 15*(4), 294–301.

Isen, A. M. (1987). Positive Affect, Cognitive Processes, and Social Behavior. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, Volume 20 of *Advances in Experimental Social Psychology*, pp. 203–253. Academic Press.

Isen, A. M. (1990). The Influence of Positive and Negative Affect on Cognitive Organization: Some Implications for Development. In *Psychological and BIological Approaches to Emotion.*, pp. 75–94. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Isen, A. M., K. A. Daubman, and G. P. Nowicki (1987). Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology 52*(6), 1122–1131.

Isen, A. M. and P. F. Levin (1972). Effect of Feeling Good on Helping: Cookies and Kindness. *Journal of Personality and Social Psychology 21*(3), 384–388.

Isen, A. M. and J. Reeve (2005). The Influence of Positive Affect on Intrinsic and Extrinsic Motivation: Facilitating Enjoyment of Play, Responsible Work Behavior, and Self-Control. *Motivation and Emotion 29*(4), 295–323.

Isen, A. M., T. E. Shalker, M. Clark, and L. Karp (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology 36*(1), 1–12.

Isen, A. M. and S. F. Simmonds (1978). The Effect of Feeling Good on a Helping Task that is Incompatible with Good Mood. *Social Psychology 41*(4), 346–349.

Ito, T. A. and J. T. Cacioppo. Variations on a human universal: Individual differences in positivity offset and negativity bias. *Cognition and Emotion* (1), 1–26.

James, W. (1890). *The Principles of Psychology*. Cleveland, OH: World Publishing.

Kavanagh, D. J. (1987). Mood, Persistence, and Success. *Australian Journal of Psychology 39*(3), 307–318.

Kavanagh, D. J. and G. H. Bower (1985). Mood and Self-Efficacy: Impact of Joy and Sadness on Perceived Capabilities. *Cognitive Therapy and Research 9*(5), 507–525.

King, G., M. Tomz, and J. Wittenberg (2000). Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science 44*(2), 347.

Kirchsteiger, G., L. Rigotti, and A. Rustichini (2006). Your Morals Might be your Moods. *Journal of Economic Behavior and Organization 59*(2), 155–172.

Köszegi, B. (2006). Ego Utility, Overconfidence, and Task Choice. *Journal of the European Economic Association 4*(4), 673–707.

Köszegi, B. and M. Rabin (2006). A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics 121*(4), 1133–1165.

Kuppens, P., F. Tuerlinckx, J. A. Russell, and L. F. Barrett. The relation between valence and arousal in subjective experience. *Psychological Bulletin* (4), 917–940.

Lang, P. J., M. K. Greenwald, M. M. Bradley, and A. O. Hamm (1993). Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions. *Psychophysiology 30*(3), 261–273.

Lerner, J. S. and D. Keltner (2000). Beyond Valence: Toward a Model of Emotion-Specific Influences on Judgement and Choice. *Cognition & Emotion 14*(4), 473–493.

Lerner, J. S. and D. Keltner (2001). Fear, Anger and Risk. *Journal of Personality and Social Psychology 81*(1), 146–159.

Locke, E. A. and G. P. Latham (1990). *A Theory of Goal Setting & Task Performance*. Prentice-Hall, Inc.

Loewenstein, G. (2000). Emotions in Economic Theory and Economic Behavior. *Preferences, Behavior and Welfare 90*(2), 426–432.

Loewenstein, G. and T. O'Donoghue (2007). The heat of the moment: Modeling interactions between affect and deliberation. *Unpublished manuscript*, 1–69.

Loewenstein, G. F. and J. S. Lerner (2003). The Role of Affect in Decision Making. In *Handbook of Affective Sciences*. Oxford University Press.

Lyubomirsky, S., L. King, and E. Diener (2005). The Benefits of Frequent Positive Affect: Does Happiness Lead to Success? *Psychological Bulletin 131*(6), 803–55.

Mackinnon, D. P., A. J. Fairchild, and M. S. Fritz (2007). Mediation Analysis. *Annual Review of Psychology 58*, 593–614.

Manucia, G. K., D. J. Baumann, and R. B. Cialdini (1984). Mood Influences on Helping: Direct Effects or Side Effects? *Journal of Personality and Social Psychology 46*(2), 357–364.

Marston, A., J. Hart, C. Hileman, and W. Faunce (1984). Toward the Laboratory Study of Sadness and Crying. *The American Journal of Psychology 97*(1), 127–131.

Martin, M. (1990). On the Induction of Mood. *Clinical Psychology Review 10*(6), 669–697.

Matthews, Gerald, M. Zeidner, and R. D. Roberts (2004). *Emotional Intelligence: Science and Myth*. MIT press.

Mellers, B. A., A. Schwartz, K. Ho, and I. Ritov (1997). Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options. *Psychological Science 8*(6), 423–429.

Melton, R. J. (1995). The Role of Positive Affect in Syllogism Performance. *Personality and Social Psychology Bulletin 21*(8), 788–794.

Oswald, A. J., E. Proto, and D. Sgroi (2015). Happiness and Productivity. *Journal of Labor Economics 33*(4), 789–822.

Petrides, K. V. (2009). Psychometric properties of the Trait Emotional Intelligence Questionnaire. In C. Stough, D. H. Saklofske, and J. D. Parker (Eds.), *Advances in the assessment of emotional intelligence*. New York: Springer.

Petrides, K. V., J. C. Pérez-González, and A. Furnham (2007). On the Criterion and Incremental Validity of Trait Emotional Intelligence. *Cognition and Emotion 21*(1), 26–55.

Pignatiello, M. F., C. J. Camp, and L. A. Rasar (1986). Musical Mood Induction: An Alternative to the Velten Technique. *Journal of Abnormal Psychology 95*(3), 295.

Proto, E., D. Sgroi, and M. Nazneen (2019). Happiness, Cooperation and Language. *Journal of Economic Behavior and Organization 168*, 209–228.

Risen, J. L. and T. Gilovich (2007). Another Look at Why People Are Reluctant to Exchange Lottery Tickets. *Journal of Personality and Social Psychology 93*(1), 12–22.

Russell, J. A., A. Weiss, and G. A. Mendelsohn (1989). Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology 57*(3), 493–502.

Salovey, P., M. Kokkonen, P. N. Lopes, and J. D. Mayer (2004). Emotional Intelligence: What do we know? In *Feelings and Emotions: The Amsterdam Symposium, Jun, 2001, Amsterdam, Netherlands*. Cambridge University Press.

Salovey, P. and J. D. Mayer (1990). Emotional Intelligence. *Imagination, Cognition, and Personality 9*(3), 185–211.

Schaefer, S. M., D. C. Jackson, R. J. Davidson, G. K. Aguirre, D. Y. Kimberg, and S. L. Thompson-Schill (2002). Modulation of Amygdalar Activity by the Conscious Regulation of Negative Emotion. *Journal of Cognitive Neuroscience 14*(6), 913–921.

Schwarz, N. and G. L. Clore (1983). Mood, Misattribution, and Judgments of Weil-Being: Informative and Directive Functions of Affective States. *Journal of Personality and Social Psychology 45*(3), 513–523.

Shepperd, J. A., C. Findley-Klein, K. D. Kwavnick, D. Walker, and S. Perez (2000). Bracing for Loss. *Journal of Personality and Social Psychology 78*(4), 620–634.

Sterling, P. (2015). Principles of allostasis: Optimal design, predictive regulation, pathophysiology, and rational therapeutics. In *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*, pp. 17–64.

Stigler, G. J. and G. S. Becker (1977). De Gustibus Non Est Disputandum. *The American Economic Review 67*(2), 76–90.

Tett, R. P., K. E. Fox, and A. Wang (2005). Development and Validation of a Self-Report Measure of Emotional Intelligence as a Multidimensional Trait Domain. *Personality and Social Psychology Bulletin 31*(7), 859–888.

Västfjäll, D. (2001). Emotion Induction Through Music: A Review of the Musical Mood Induction Procedure. *Musicae Scientiae 5*(1_suppl), 173–211.

Velten, E. J. (1968). A Laboratory Task for Induction of Mood States. *Behaviour Research and Therapy 6*(4), 473–482.

Vosgerau, J. (2010). How Prevalent Is Wishful Thinking? Misattribution of Arousal Causes Optimism and Pessimism in Subjective Probabilities. *Journal of Experimental Psychology: General 139*(1), 32–48.

Watson, D., L. A. Clark, and A. Tellegen (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology 54*(6), 1063–1070.

Weber, E. U. (1994). From Subjective Probabilities to Decision Weights: The Effect of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events. *Psychological Bulletin 115*(2), 228–242.

Westermann, R., K. Spies, G. Stahl, and F. W. Hesse (1996). Relative Effectiveness and Validity of Mood Induction Procedures: A Meta-Analysis. *European Journal of Social Psychology 26*(4), 557–580.

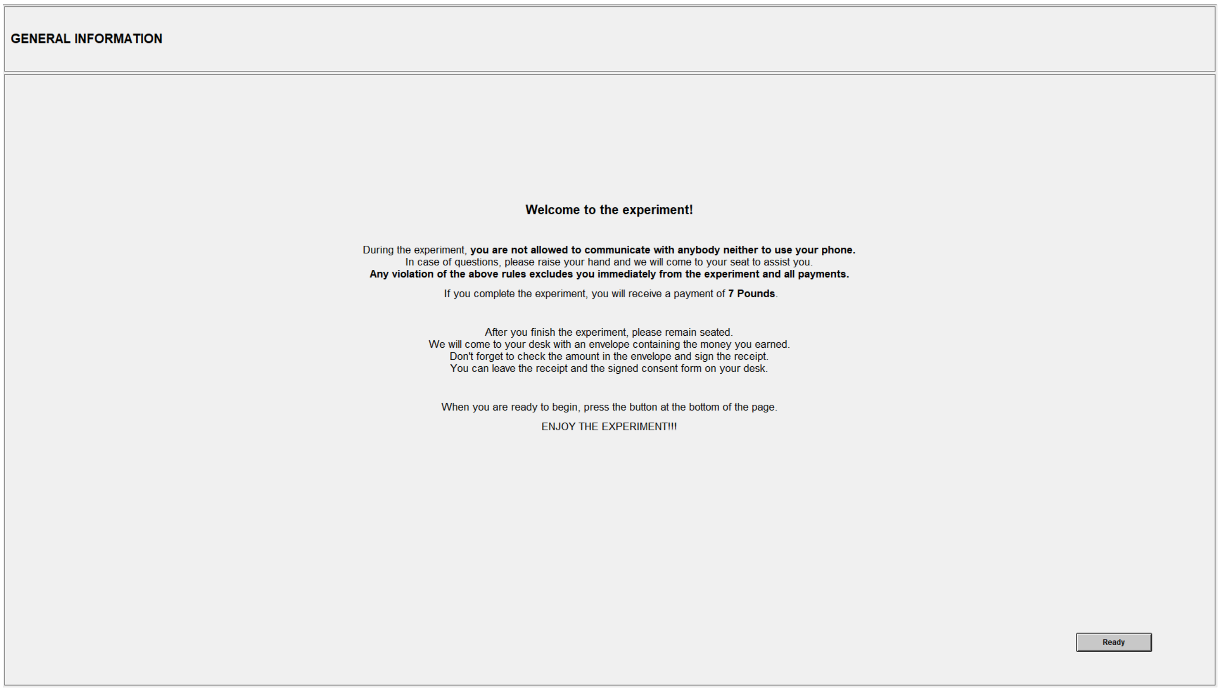
Weyant, J. M. (1978). Effects of Mood States, Costs, and Benefits on Helping. *Journal of Personality and Social Psychology 36*(10), 1169–1176.

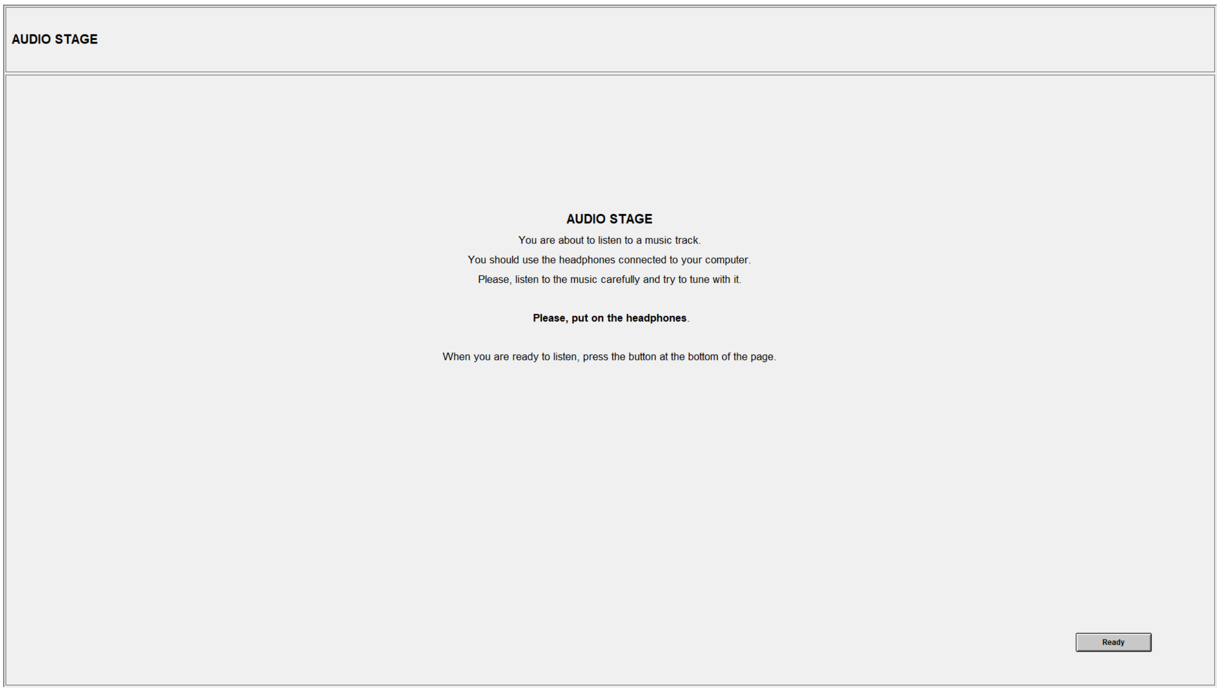
Yip, J. A. and S. Côté (2013). The Emotionally Intelligent Decision Maker: Emotion-Understanding Ability Reduces the Effect of Incidental Anxiety on Risk Taking. *Psychological Science 24*(1), 48–55.

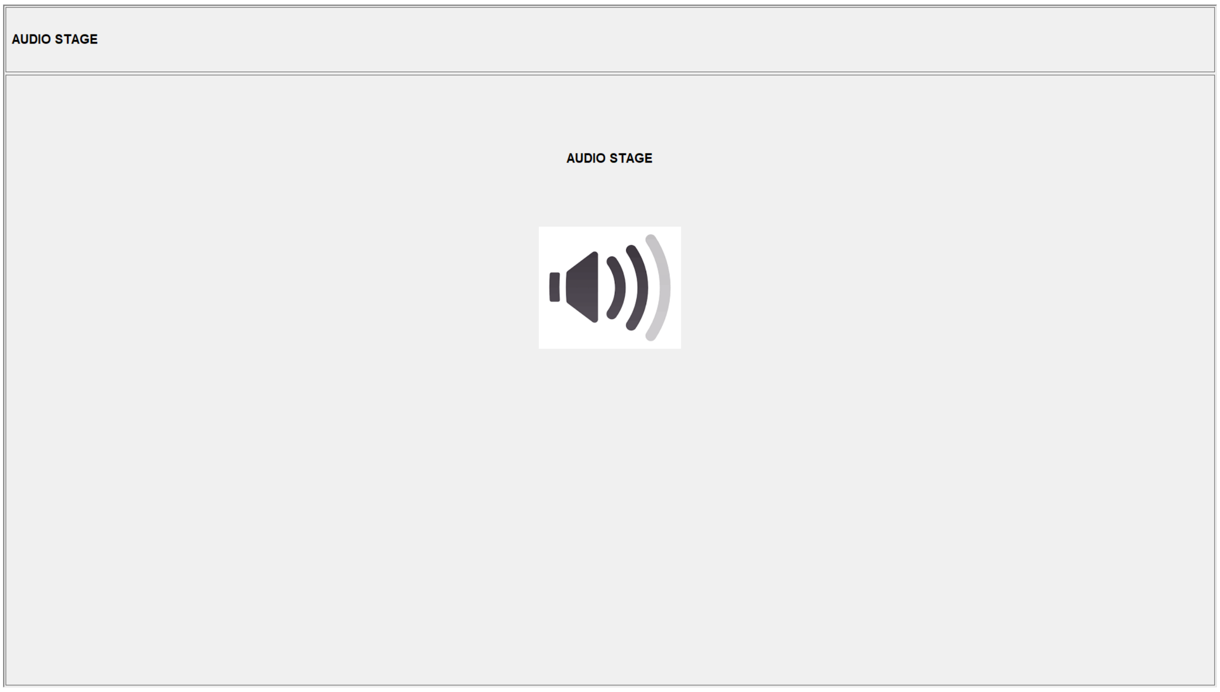
Online Appendix

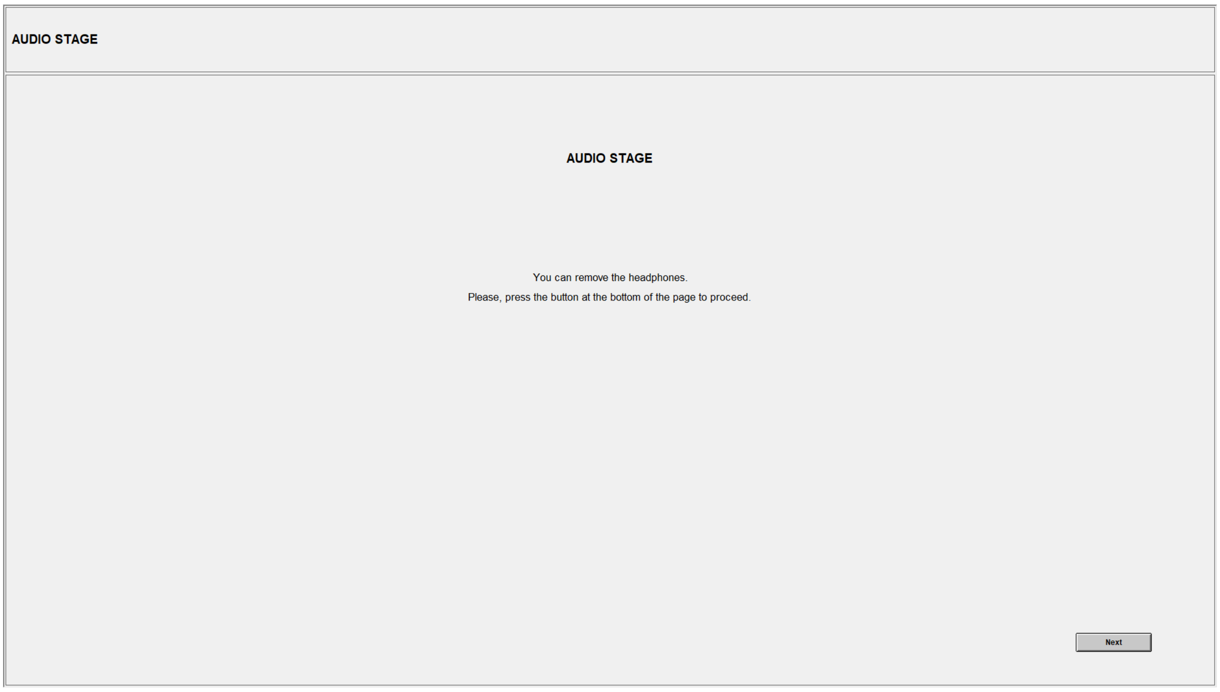
A Appendix: Pilot sessions - Experimental instructions

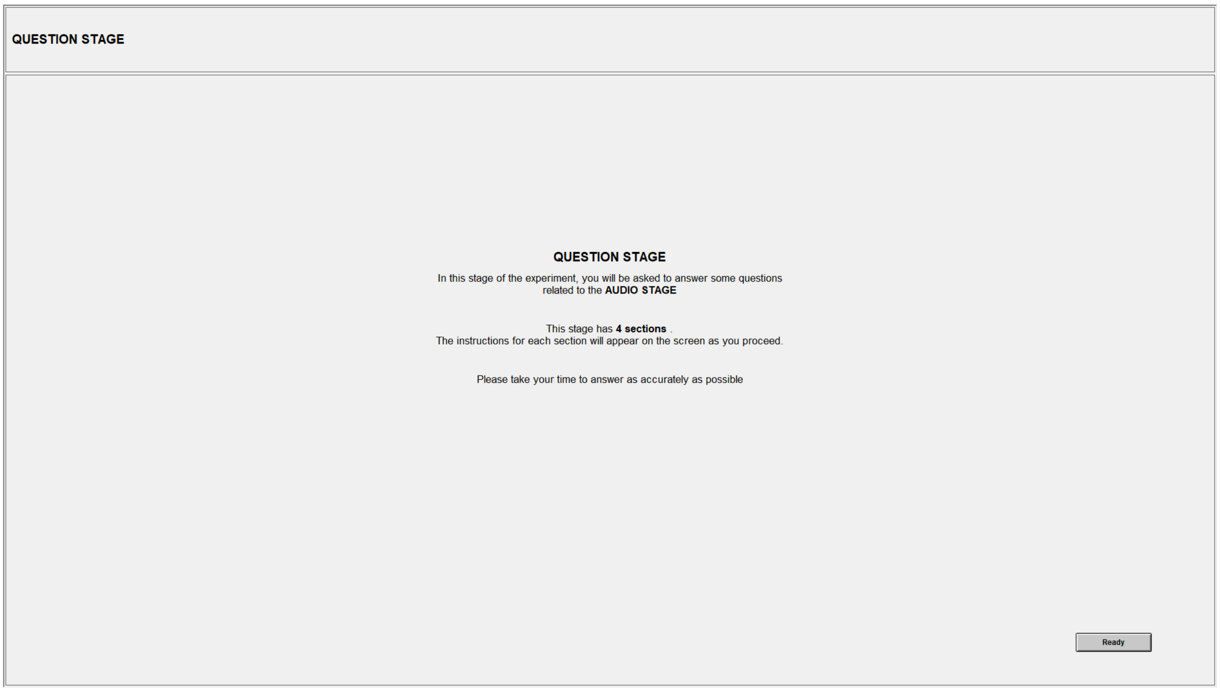
All the information concerning the experiment appeared directly on the screen. Participants were not given any written instructions. Here we report the screenshots for the test of the positive valence music tracks. Participants testing the negative valence music tracks received exactly the same instructions but listened to different audio recordings. Participants in the neutral condition answered the same question but did not listen to any audio recording.

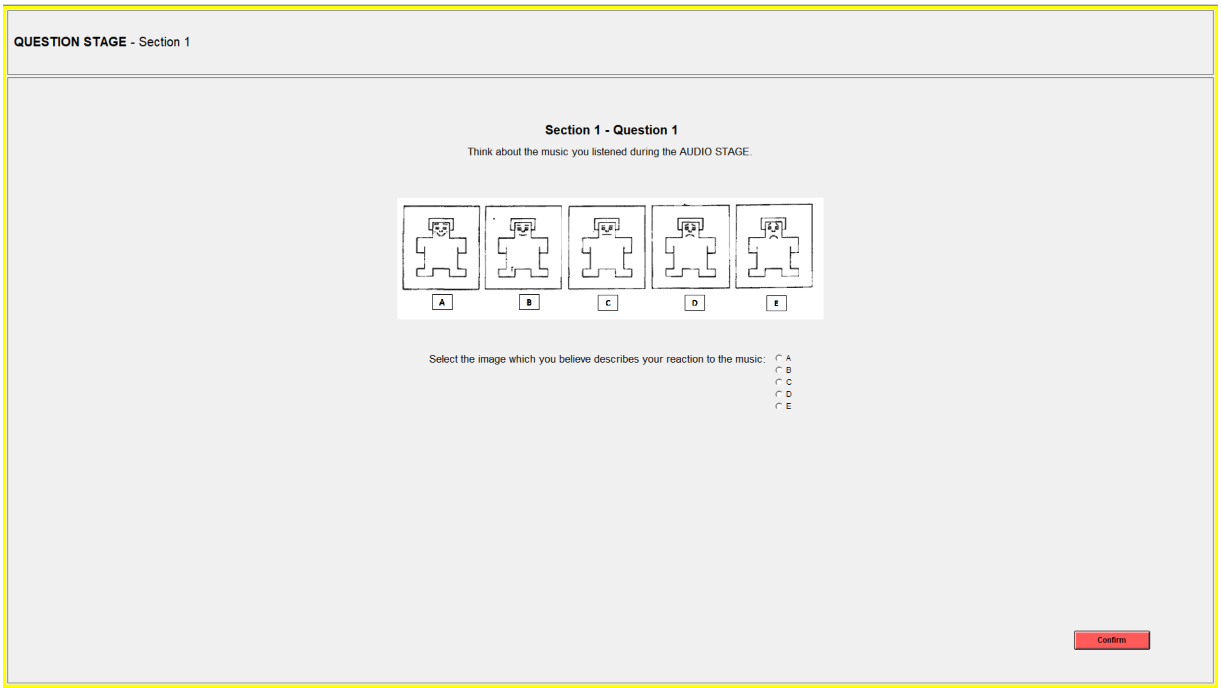


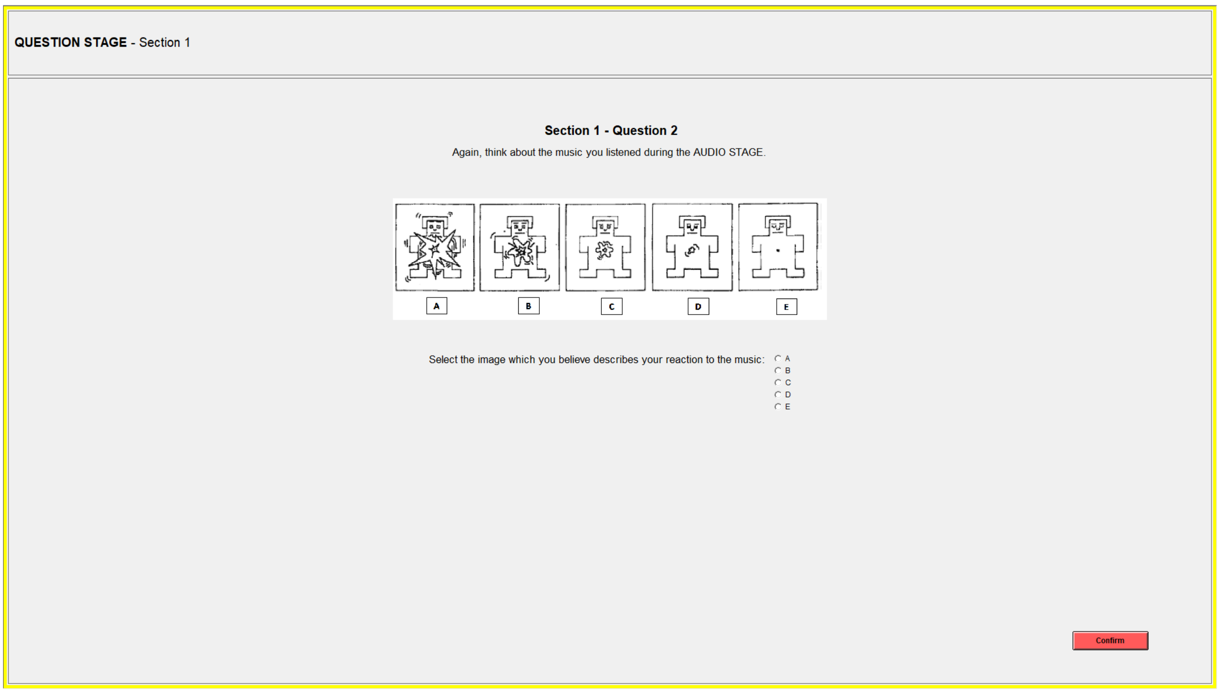

AUDIO STAGE
AUDIO STAGE
You are about to listen to a music track.
You should use the headphones connected to your computer.
Please, listen to the music carefully and try to tune with it.
Please, put on the headphones.
When you are ready to listen, press the button at the bottom of the page.
Ready

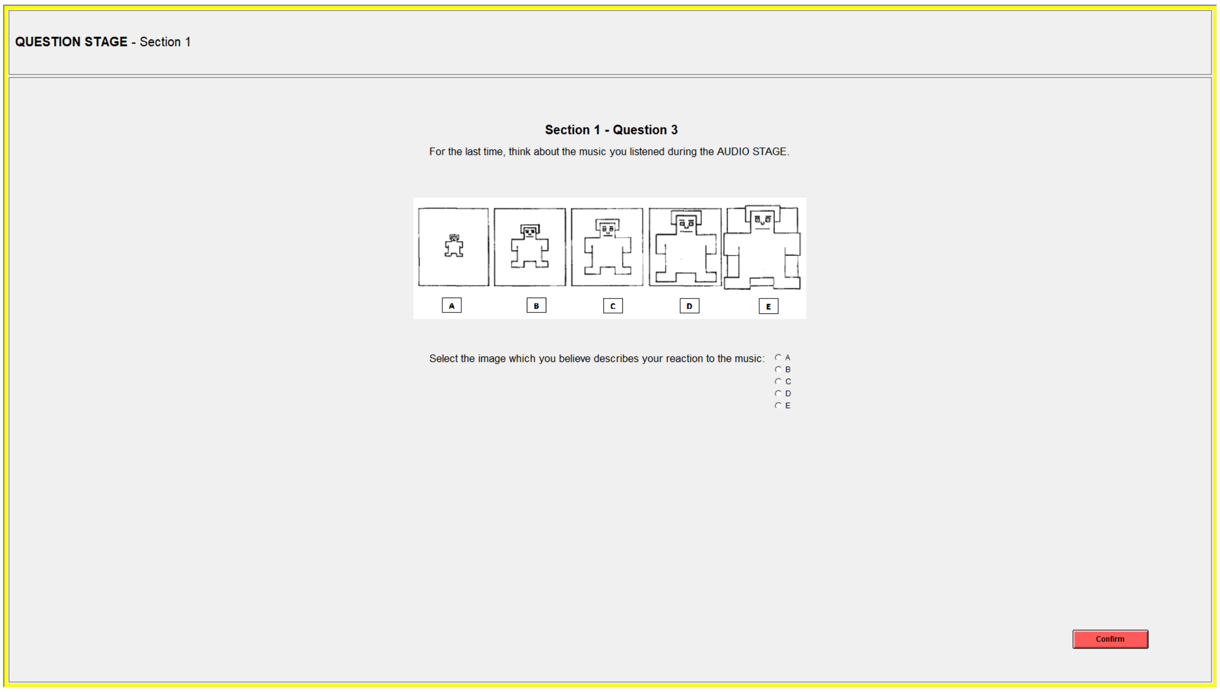

AUDIO STAGE
AUDIO STAGE


AUDIO STAGE
AUDIO STAGE
You can remove the headphones.
Please, press the button at the bottom of the page to proceed.
Next

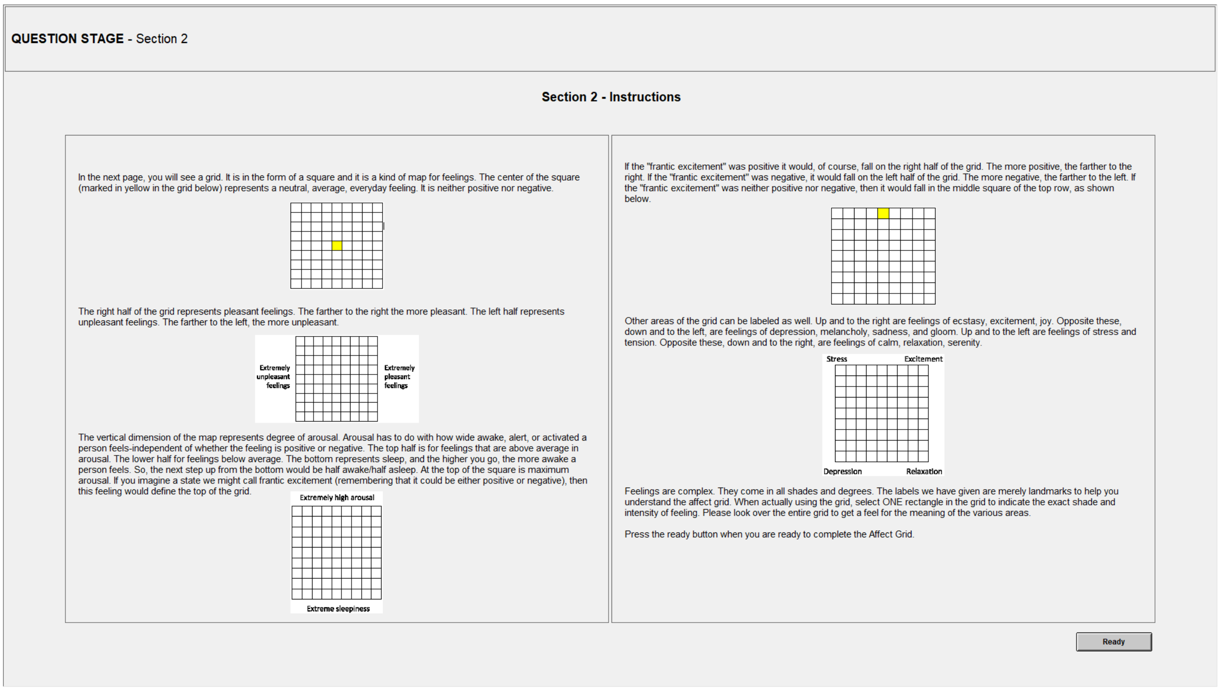

QUESTION STAGE
QUESTION STAGE
In this stage of the experiment, you will be asked to answer some questions related to the AUDIO STAGE
This stage has 4 sections .
The instructions for each section will appear on the screen as you proceed.
Please take your time to answer as accurately as possible
Ready


QUESTION STAGE - Section 1
Section 1 - Question 1
Think about the music you listened during the AUDIO STAGE.
A
B
C
D
E
Select the image which you believe describes your reaction to the music:
A
B
C
D
E
Confirm

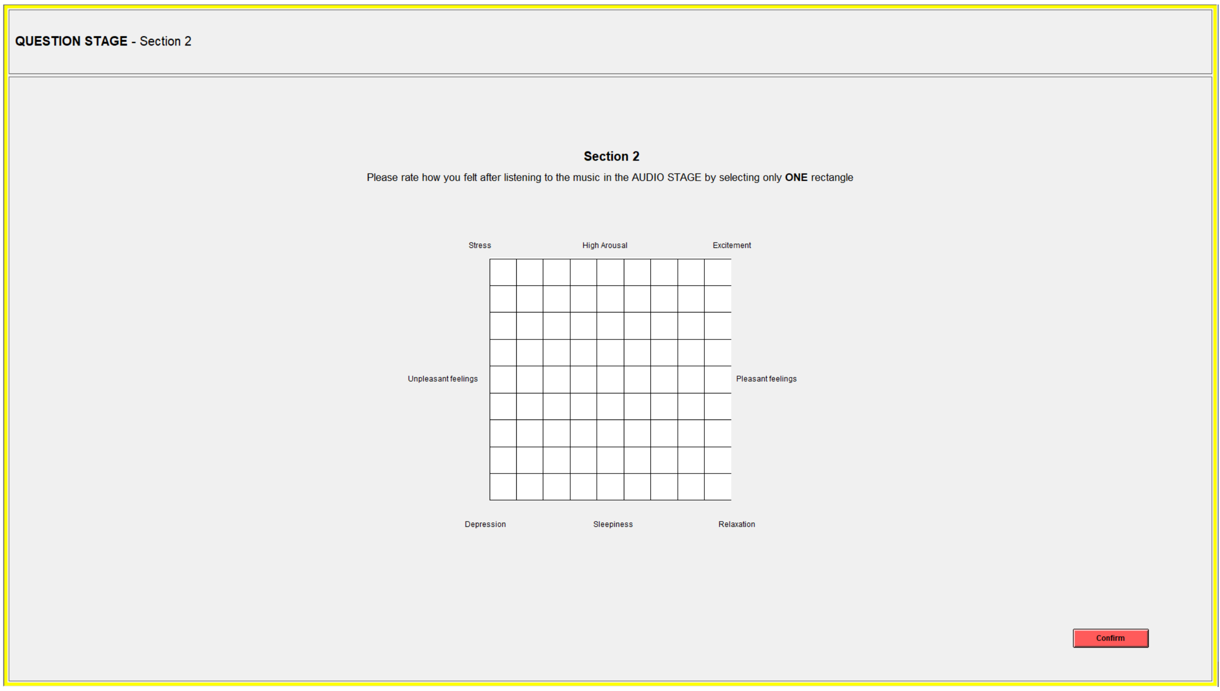

QUESTION STAGE - Section 1
Section 1 - Question 2
Again, think about the music you listened during the AUDIO STAGE.
A
B
C
D
E
Select the image which you believe describes your reaction to the music:
A
B
C
D
E
Confirm



QUESTION STAGE - Section 1
Section 1 - Question 3
For the last time, think about the music you listened during the AUDIO STAGE.
A
B
C
D
E
Select the image which you believe describes your reaction to the music:
A
B
C
D
E
Confirm



QUESTION STAGE - Section 2
Section 2 - Instructions
In the next page, you will see a grid. It is in the form of a square and it is a kind of map for feelings. The center of the square (marked in yellow in the grid below) represents a neutral, average, everyday feeling. It is neither positive nor negative.
The right half of the grid represents pleasant feelings. The farther to the right the more pleasant. The left half represents unpleasant feelings. The farther to the left, the more unpleasant.
Extremely unpleasant feelings
Extremely pleasant feelings
The vertical dimension of the map represents degree of arousal. Arousal has to do with how wide awake, alert, or activated a person feels-independent of whether the feeling is positive or negative. The top half is for feelings that are above average in arousal. The lower half for feelings below average. The bottom represents sleep, and the higher you go, the more awake a person feels. So, the next step up from the bottom would be half awake/half asleep. At the top of the square is maximum arousal. If you imagine a state we might call frantic excitement (remembering that it could be either positive or negative), then this feeling would define the top of the grid.
Extremely high arousal
Extreme sleepiness
If the "frantic excitement" was positive it would, of course, fall on the right half of the grid. The more positive, the farther to the right. If the "frantic excitement" was negative, it would fall on the left half of the grid. The more negative, the farther to the left. If the "frantic excitement" was neither positive nor negative, then it would fall in the middle square of the top row, as shown below.
Other areas of the grid can be labeled as well. Up and to the right are feelings of ecstasy, excitement, joy. Opposite these, down and to the left, are feelings of depression, melancholy, sadness, and gloom. Up and to the left are feelings of stress and tension. Opposite these, down and to the right, are feelings of calm, relaxation, serenity.
Stress
Excitement
Depression
Relaxation
Feelings are complex. They come in all shades and degrees. The labels we have given are merely landmarks to help you understand the affect grid. When actually using the grid, select ONE rectangle in the grid to indicate the exact shade and intensity of feeling. Please look over the entire grid to get a feel for the meaning of the various areas.
Press the ready button when you are ready to complete the Affect Grid.
Ready



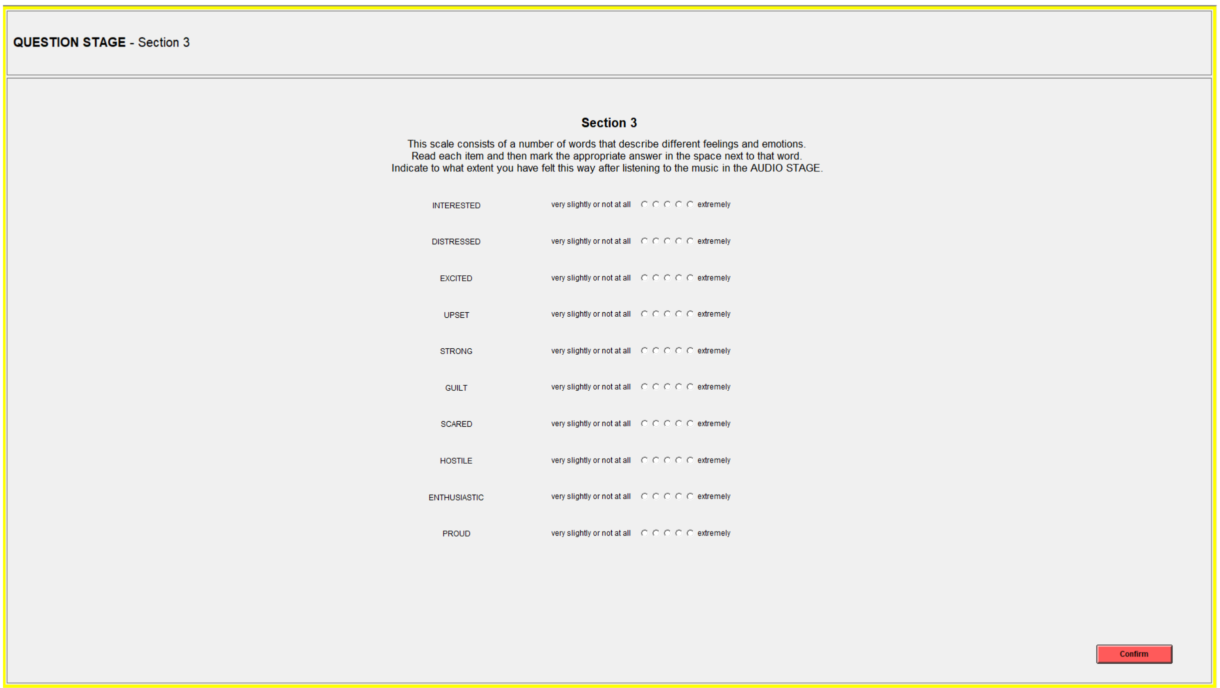
QUESTION STAGE - Section 2

Section 2

Please rate how you felt after listening to the music in the AUDIO STAGE by selecting only **ONE** rectangle

Stress
High Arousal
Excitement
Unpleasant feelings
Pleasant feelings
Depression
Sleepiness
Relaxation

Confirm



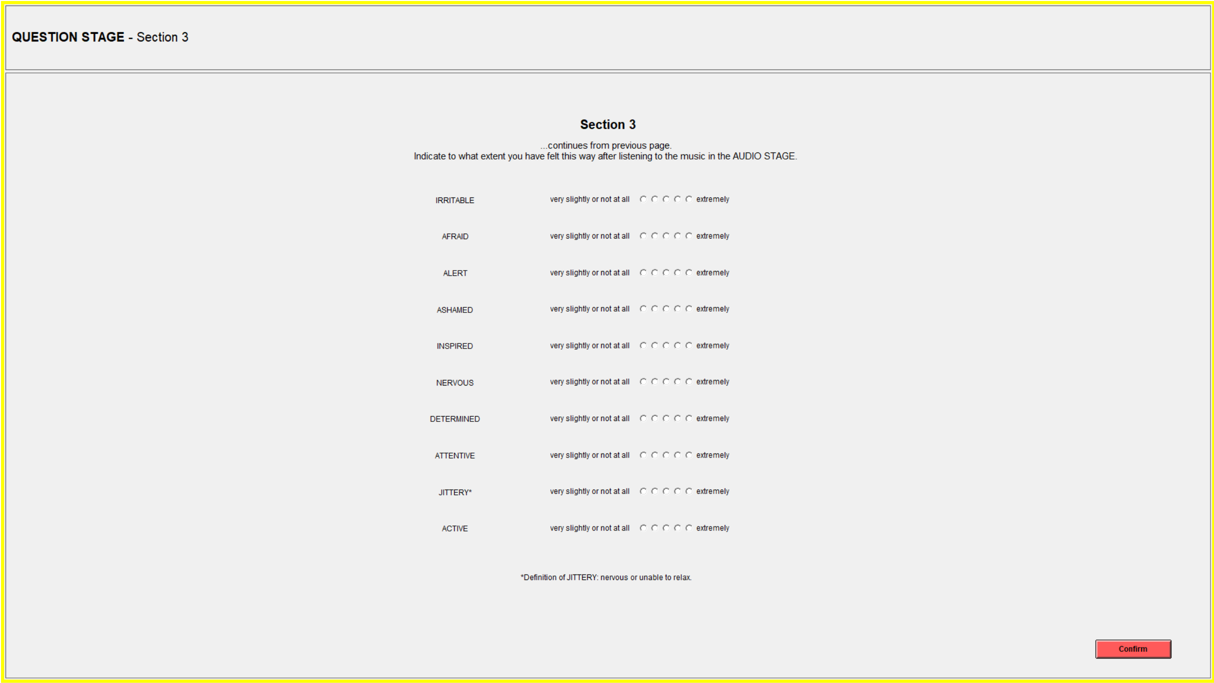
QUESTION STAGE - Section 3

Section 3

This scale consists of a number of words that describe different feelings and emotions.
Read each item and then mark the appropriate answer in the space next to that word.
Indicate to what extent you have felt this way after listening to the music in the AUDIO STAGE.

INTERESTED	very slightly or not at all	extremely
DISTRESSED	very slightly or not at all	extremely
EXCITED	very slightly or not at all	extremely
UPSET	very slightly or not at all	extremely
STRONG	very slightly or not at all	extremely
GUILT	very slightly or not at all	extremely
SCARED	very slightly or not at all	extremely
HOSTILE	very slightly or not at all	extremely
ENTHUSIASTIC	very slightly or not at all	extremely
PROUD	very slightly or not at all	extremely

Confirm



QUESTION STAGE - Section 3

Section 3

...continues from previous page.
Indicate to what extent you have felt this way after listening to the music in the AUDIO STAGE.

IRRITABLE very slightly or not at all extremely

AFRAID very slightly or not at all extremely

ALERT very slightly or not at all extremely

ASHAMED very slightly or not at all extremely

INSPIRED very slightly or not at all extremely

NERVOUS very slightly or not at all extremely

DETERMINED very slightly or not at all extremely

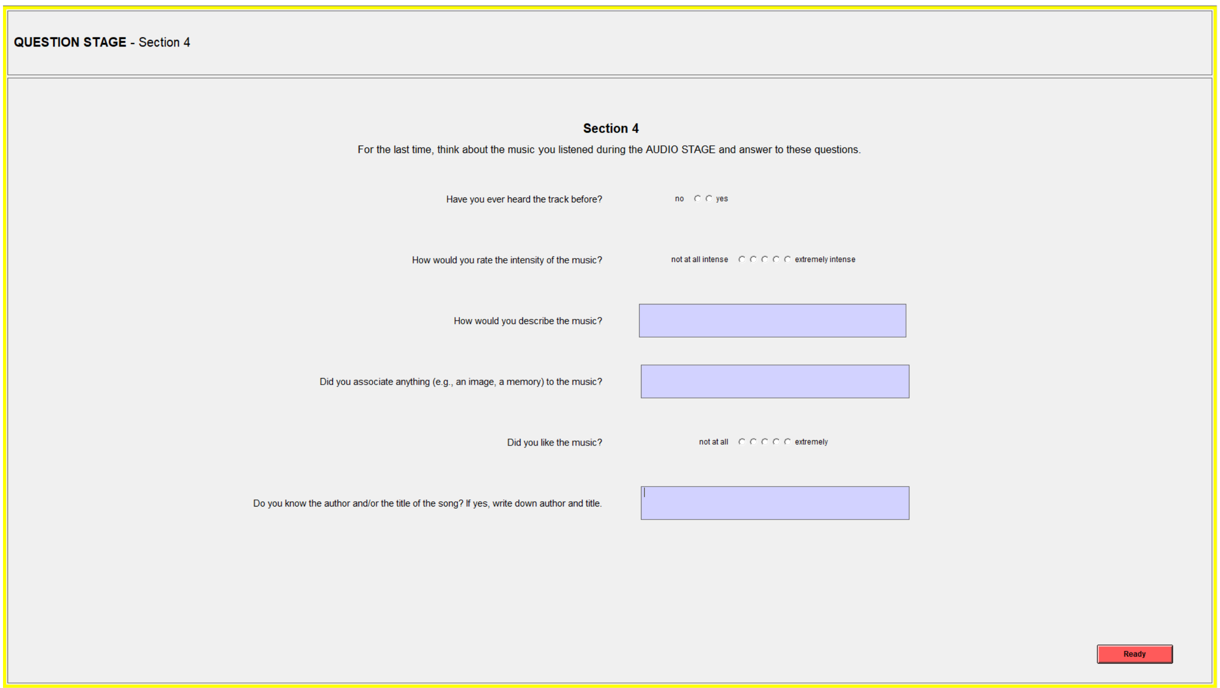
ATTENTIVE very slightly or not at all extremely

JITTERY* very slightly or not at all extremely

ACTIVE very slightly or not at all extremely

*Definition of JITTERY: nervous or unable to relax.

Confirm



QUESTION STAGE - Section 4

Section 4

For the last time, think about the music you listened during the AUDIO STAGE and answer to these questions.

Have you ever heard the track before? no yes

How would you rate the intensity of the music? not at all intense extremely intense

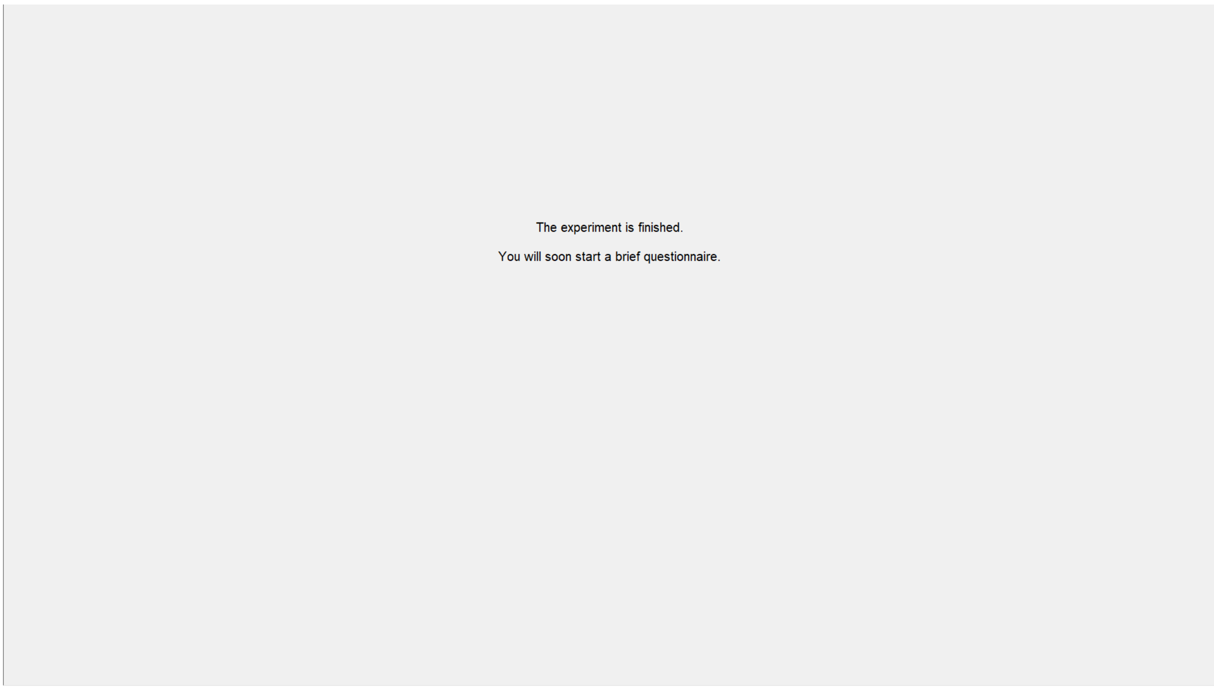
How would you describe the music?

Did you associate anything (e.g., an image, a memory) to the music?

Did you like the music? not at all extremely

Do you know the author and/or the title of the song? If yes, write down author and title.

Ready


The experiment is finished.
You will soon start a brief questionnaire.

B Appendix: Music appreciation and valence

According to our pilot sessions, two of the music tracks that were supposed to increase negative affects - Holst and Moussorgosky - provoked an increase in *positive* affects. We analysed the answers to the question 'Did you like the music?' in the post-experimental questionnaire as we hypothezise an interaction between music appreciation and valence. To answer the question, participants state how much they liked the music on a 5-Likert scale, from 'Not at all' to 'Extremely'. Table B1 report Pearson correlation values between music taste and SAM, Affect grid, and PANAS answers.

	Music appreciation
Arousal - SAM	0.0351
Arousal - Affect Grid	0.0047
Valence - SAM	0.7724***
Valence - Affect Grid	0.8638***
Positive emotions PANAS	0.7591***
Negative emotions PANAS	-0.6118***

Table B1: Correlation valence and music appreciation. Pearson correlation; *$p<0.1$, **$p<0.05$, ***$p<0.01$. Sample size: 46.

We observe a strong positive (or negative) correlation between how much participants like the music and how much they report positive (or negative) affects. There is no correlation between arousal and music appreciation. Figure B1 show the level of appreciation by music track. Holst and Moussorgosky have high levels of appreciation which explain the high values of positive valence reported after listening to those tracks.

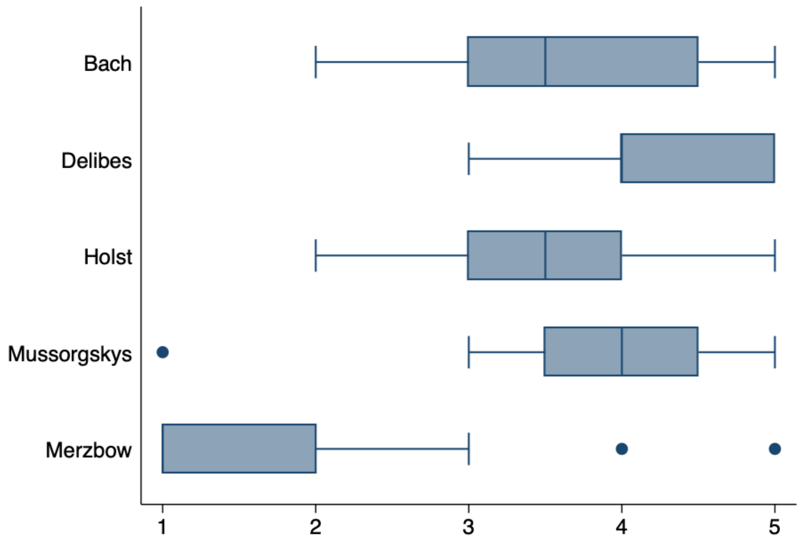
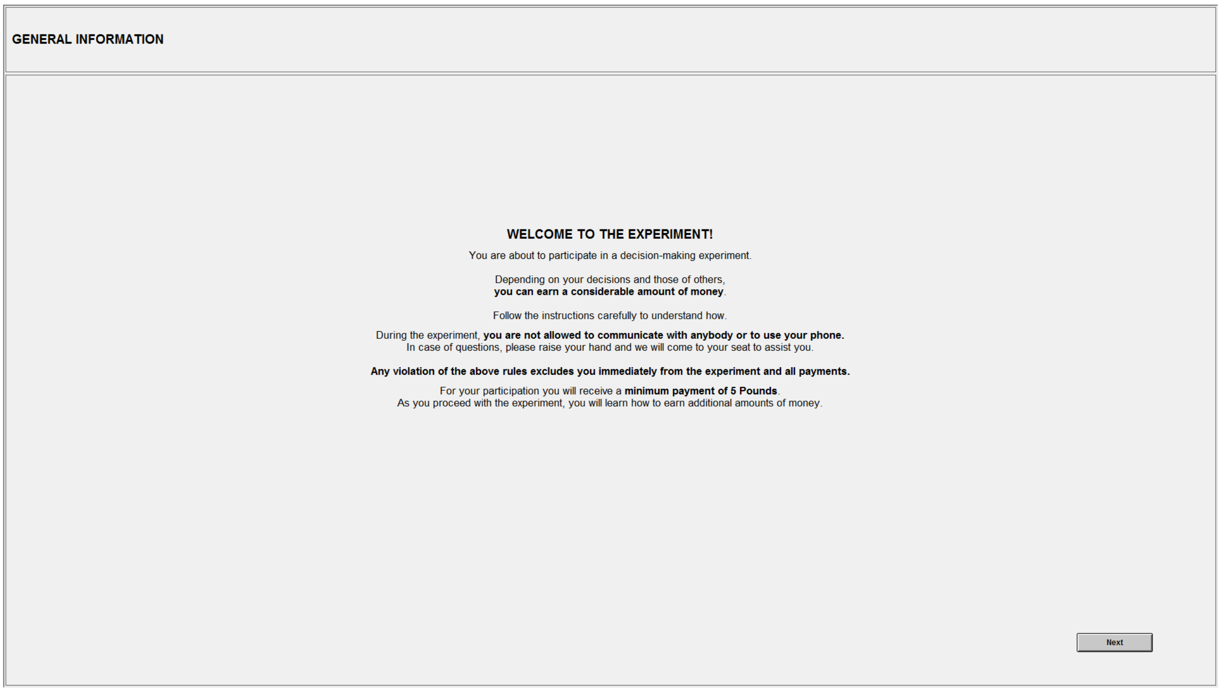


Figure B1: Average music appreciation (median, 25th and 75th percentiles, dots are outside values)

C Appendix: Experimental instructions

All the information concerning the experiment appeared directly on the screen. Participants were not given any written instructions. Here we report the screenshots from a session using music with negative valence at Timing 1. The only variation in the instructions across treatments is the timing of the AIP.



EXPERIMENTAL INSTRUCTIONS

EXPERIMENTAL INSTRUCTIONS

The experiment consists of **4 stages**, and some surveys.

In the next page you will find the instruction for STAGE 1.
The rest of the stages will be explained on the computer screen as you proceed with the experiment.

Please read the instructions carefully as real money can be gained in the experiment!

At the end of the experiment, **2 stages out of 4 will be randomly selected**, and **you will be paid the payoffs you earned during the 2 stages selected plus the 4 Pounds show-up fee**.

After you finish the experiment, please **remain seated**.
We will come to your desk with an **envelope containing the money** you earned.
Don't forget to check the amount in the envelope and sign the receipt.
You can leave the receipt and the signed consent form on your desk.

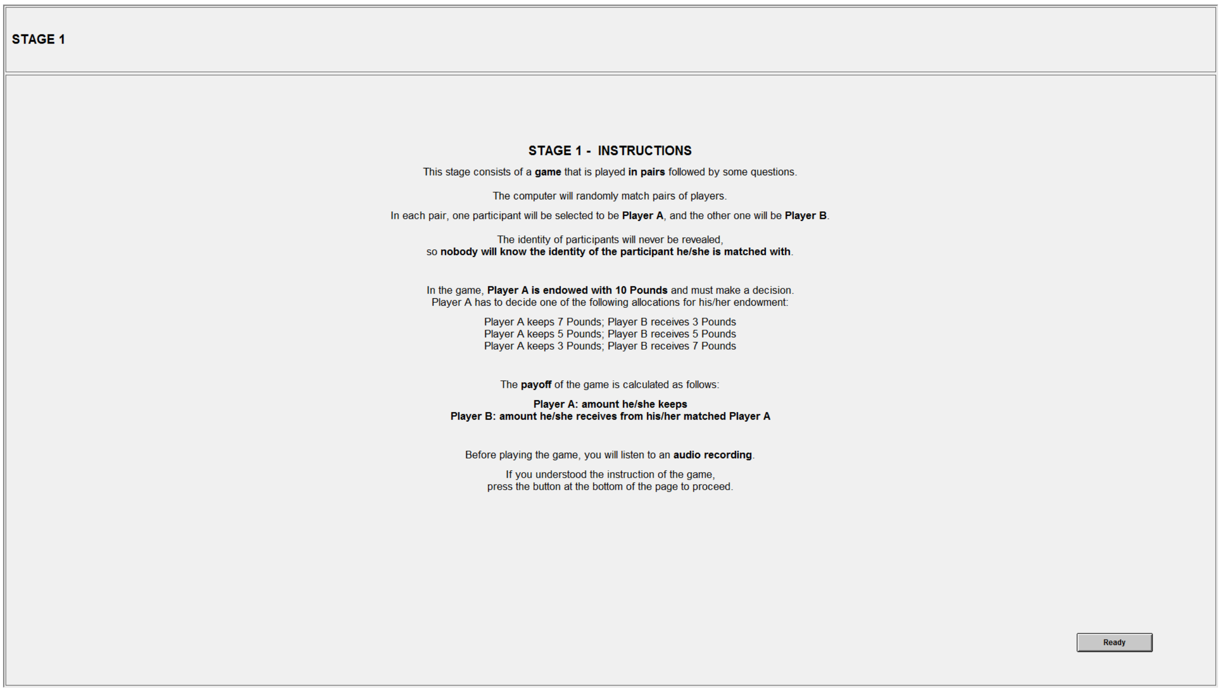
When you are ready to begin, press the button at the bottom of the page.

ENJOY THE EXPERIMENT!!!

Ready

C.1 Dictator's Game

All participants



STAGE 1

STAGE 1 - INSTRUCTIONS

This stage consists of a **game** that is played **in pairs** followed by some questions.

The computer will randomly match pairs of players.

In each pair, one participant will be selected to be **Player A**, and the other one will be **Player B**.

The identity of participants will never be revealed,
so **nobody will know the identity of the participant he/she is matched with**.

In the game, **Player A is endowed with 10 Pounds** and must make a decision.
Player A has to decide one of the following allocations for his/her endowment:

Player A keeps 7 Pounds; Player B receives 3 Pounds
Player A keeps 5 Pounds; Player B receives 5 Pounds
Player A keeps 3 Pounds; Player B receives 7 Pounds

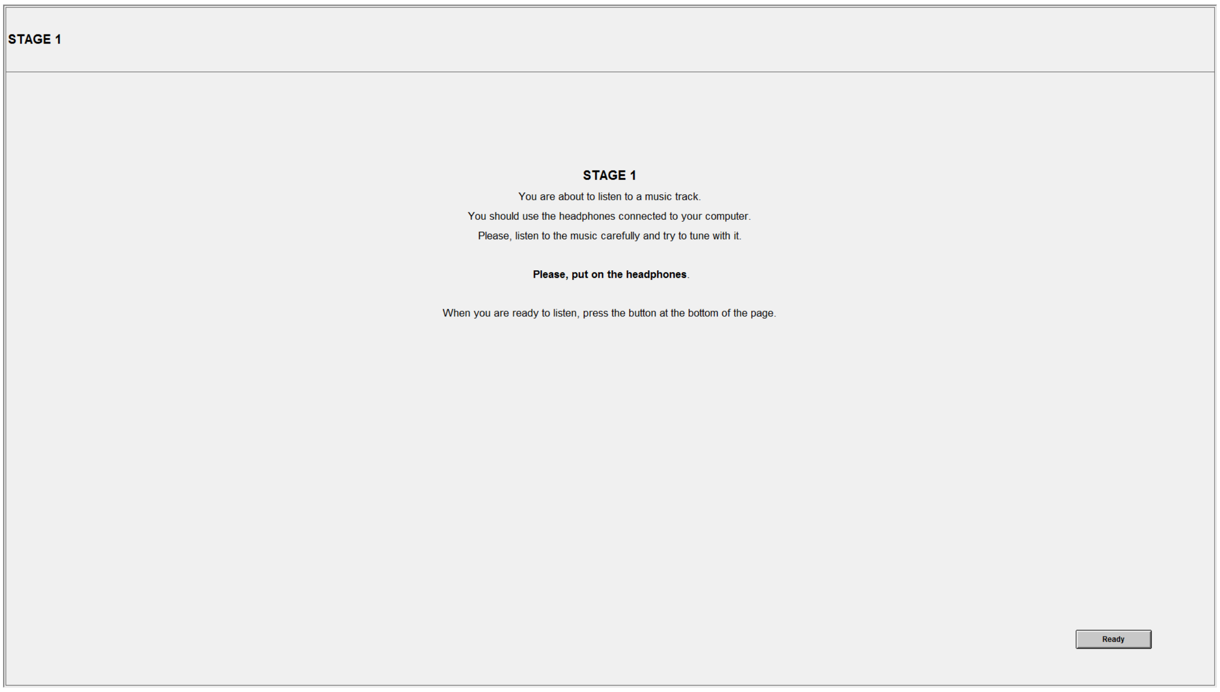
The **payoff** of the game is calculated as follows:

Player A: amount he/she keeps
Player B: amount he/she receives from his/her matched Player A

Before playing the game, you will listen to an **audio recording**.

If you understood the instruction of the game,
press the button at the bottom of the page to proceed.

Ready



STAGE 1

STAGE 1

You are about to listen to a music track.

You should use the headphones connected to your computer.

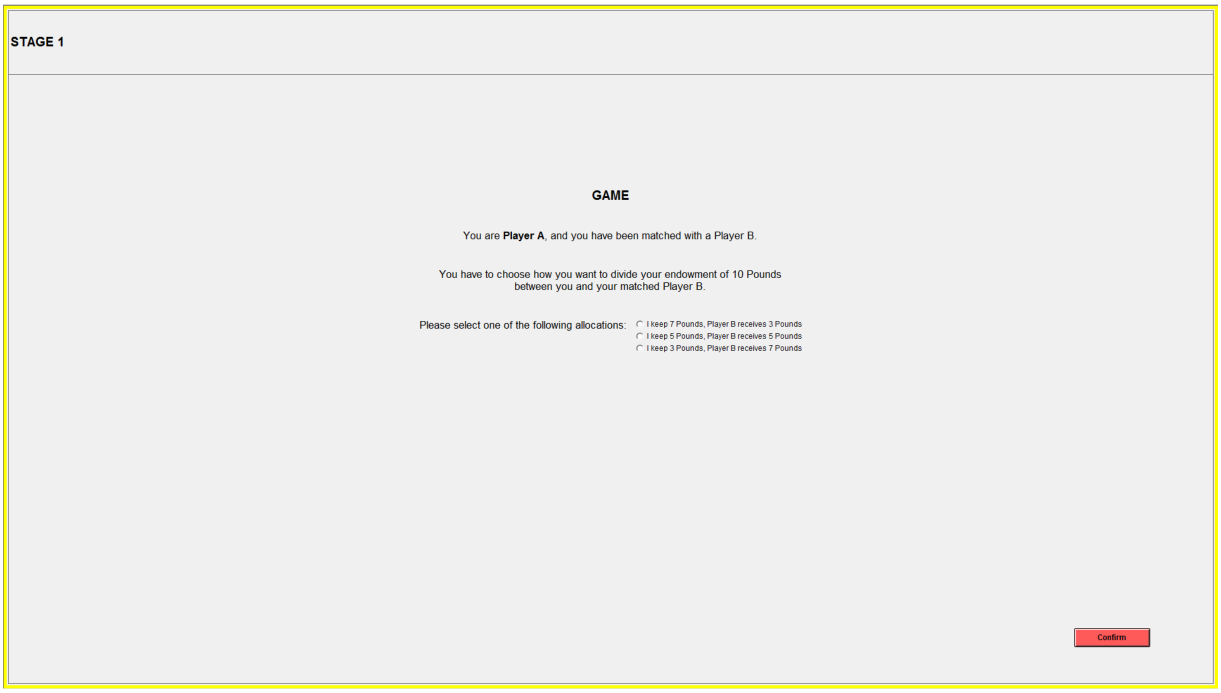
Please, listen to the music carefully and try to tune with it.

Please, put on the headphones.

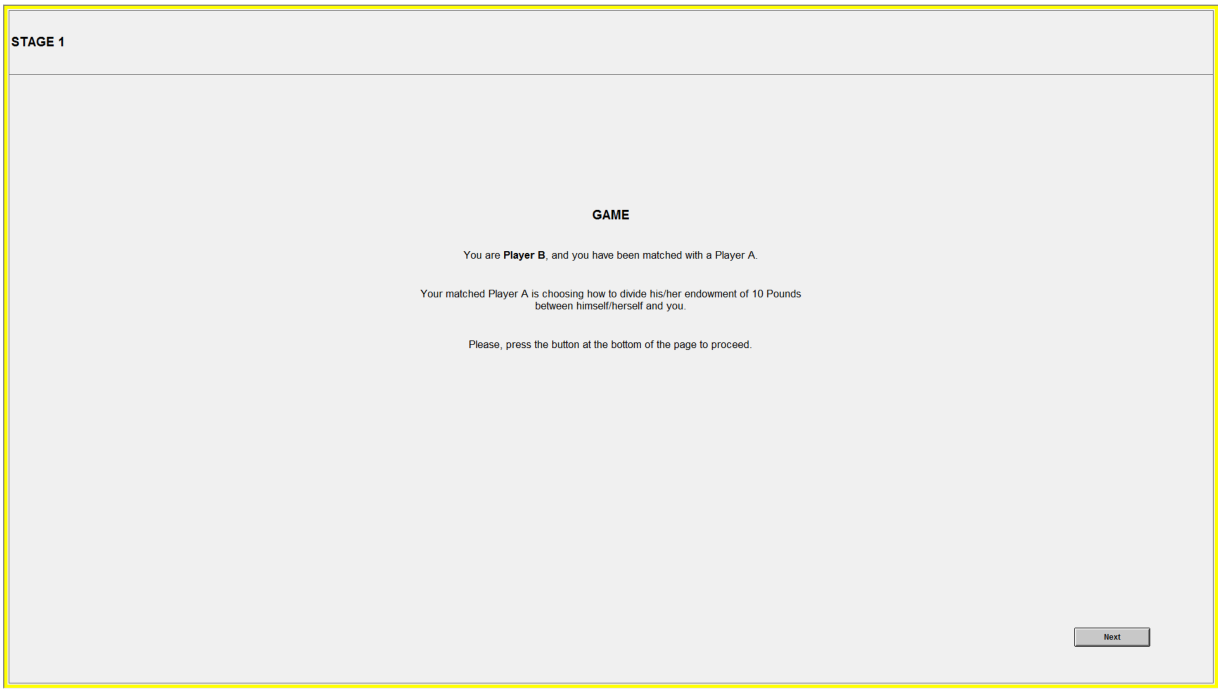
When you are ready to listen, press the button at the bottom of the page.

Ready

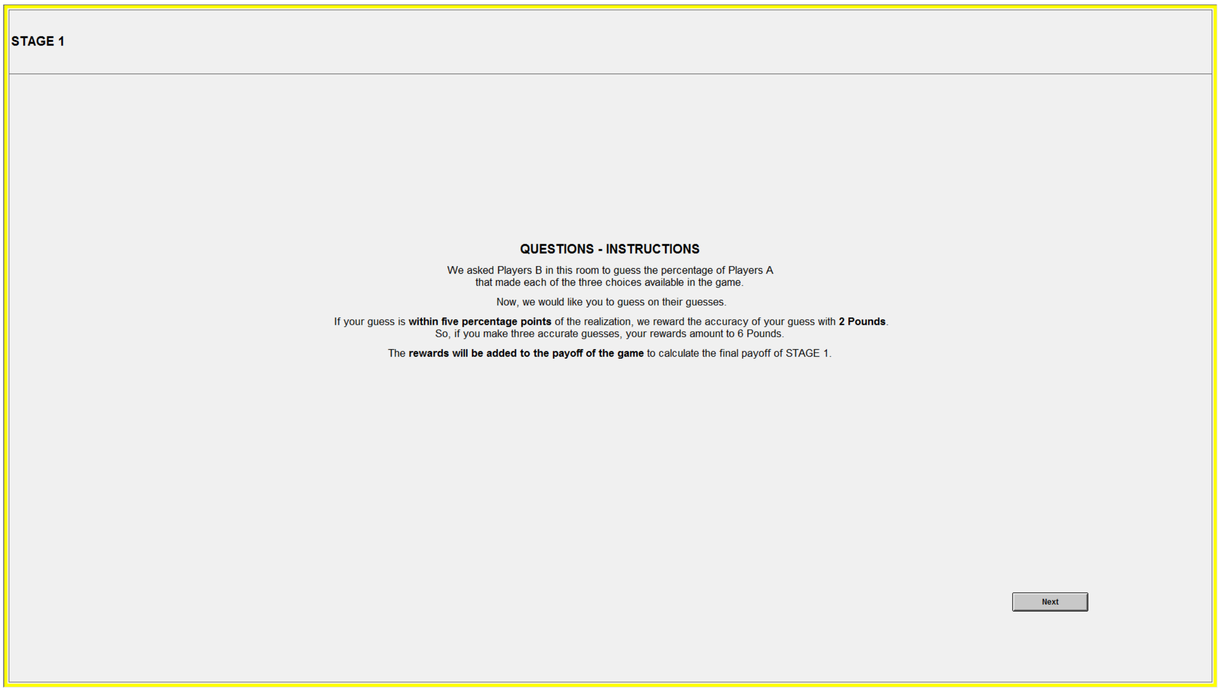
Dictators



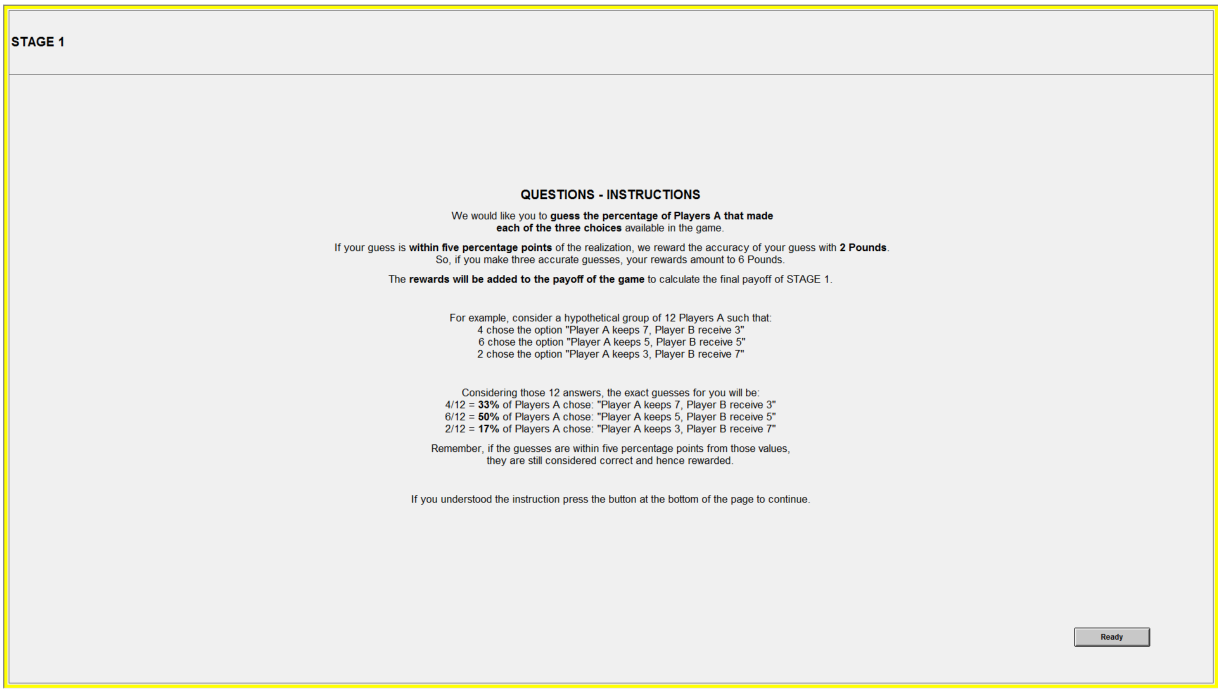
Recipients



Dictators



Recipients



Dictators

STAGE 1

For example, consider 3 hypothetical Players B making the following guesses:

Guess of first hypothetical Player B:
55% of Players A chose: "Player A keeps 7, Player B receives 3"
35% of Players A chose: "Player A keeps 5, Player B receives 5"
10% of Players A chose: "Player A keeps 3, Player B receives 7"

Guess of second hypothetical Player B:
20% of Players A chose: "Player A keeps 7, Player B receives 3"
70% of Players A chose: "Player A keeps 5, Player B receives 5"
10% of Players A chose: "Player A keeps 3, Player B receives 7"

Guess of third hypothetical Player B:
25% of Players A chose: "Player A keeps 7, Player B receives 3"
35% of Players A chose: "Player A keeps 5, Player B receives 5"
40% of Players A chose: "Player A keeps 3, Player B receives 7"

Back Next

STAGE 1

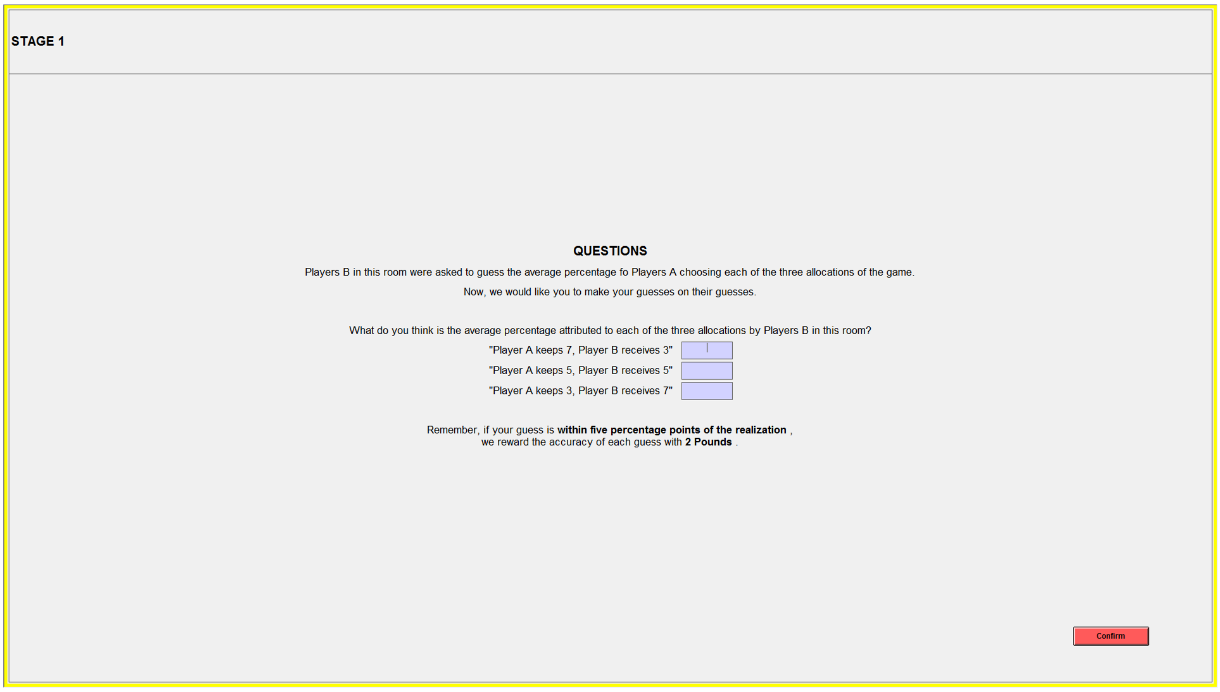
Considering the 3 answers in the previous page, the exact guesses for you will be:

(55+20+25)/3 = **33%** of Players A chose: "Player A keeps 7, Player B receives 3"
(35+70+35)/3 = **47%** of Players A chose: "Player A keeps 5, Player B receives 5"
(10+10+40)/3 = **20%** of Players A chose: "Player A keeps 3, Player B receives 7"

Still if the guesses are within five percentage points from those values,
they are still considered correct and hence rewarded.

If you understood the instruction press the button at the bottom of the page to proceed.

Back Ready



STAGE 1

QUESTIONS

Players B in this room were asked to guess the average percentage fo Players A choosing each of the three allocations of the game.

Now, we would like you to make your guesses on their guesses.

What do you think is the average percentage attributed to each of the three allocations by Players B in this room?

"Player A keeps 7, Player B receives 3"

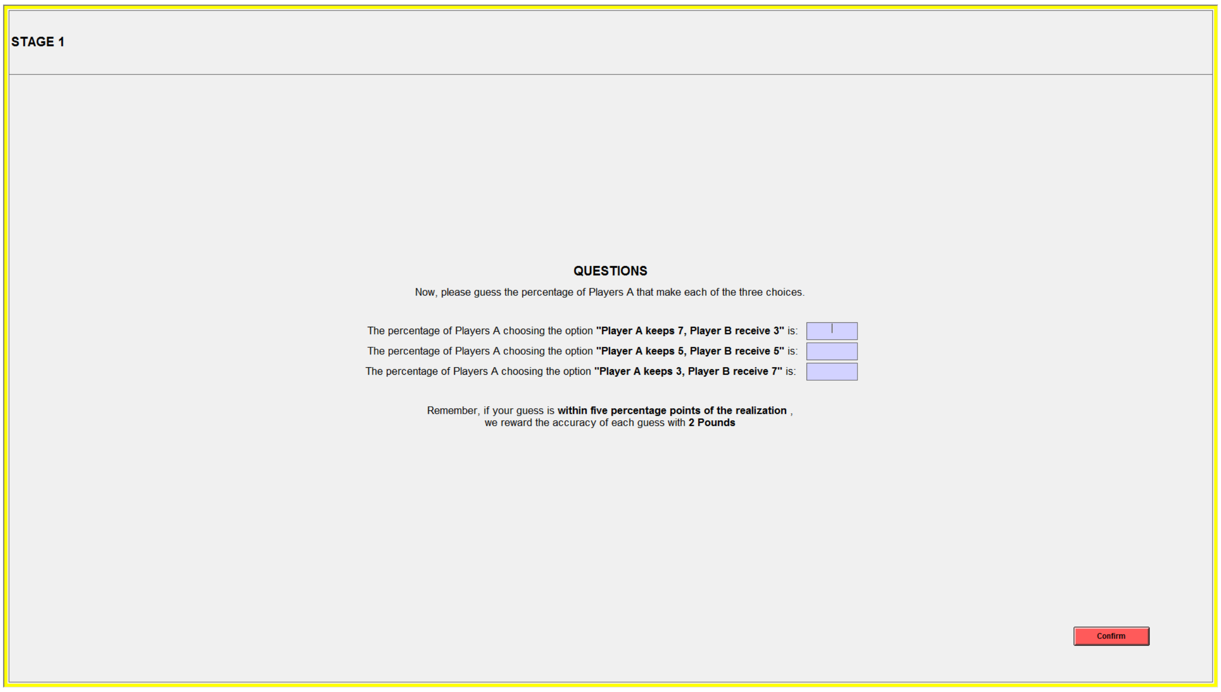
"Player A keeps 5, Player B receives 5"

"Player A keeps 3, Player B receives 7"

Remember, if your guess is **within five percentage points of the realization** ,
we reward the accuracy of each guess with **2 Pounds** .

Confirm

Recipients



STAGE 1

QUESTIONS

Now, please guess the percentage of Players A that make each of the three choices.

The percentage of Players A choosing the option **"Player A keeps 7, Player B receive 3"** is:

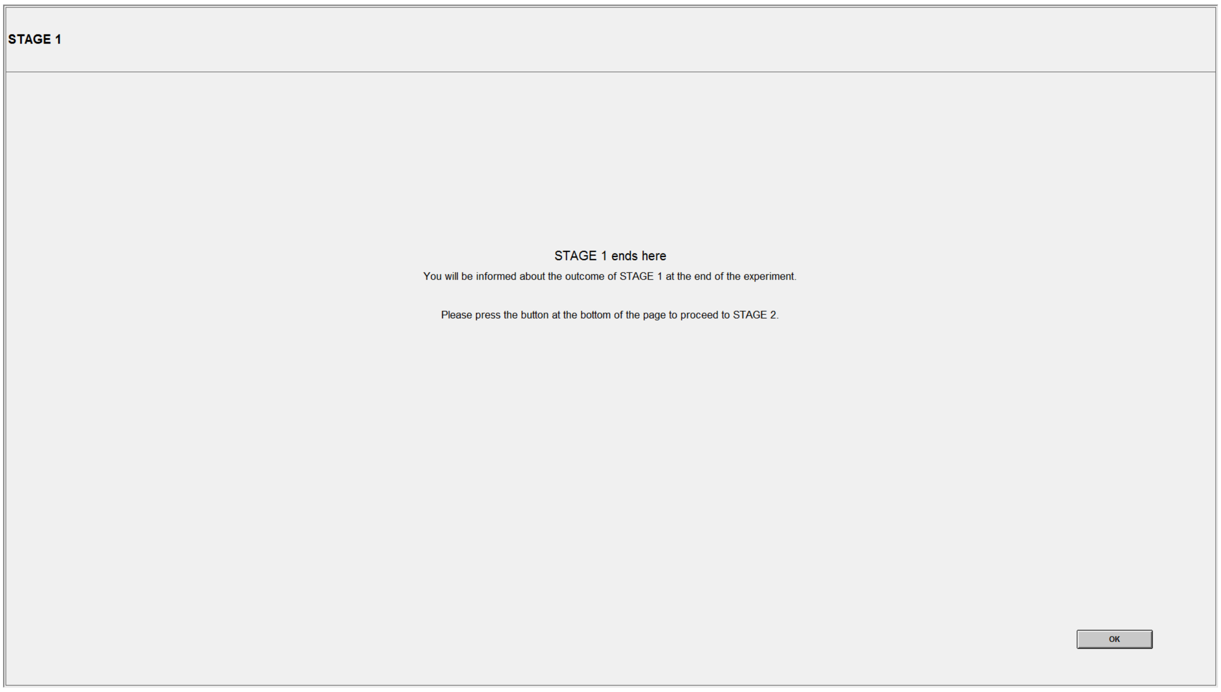
The percentage of Players A choosing the option **"Player A keeps 5, Player B receive 5"** is:

The percentage of Players A choosing the option **"Player A keeps 3, Player B receive 7"** is:

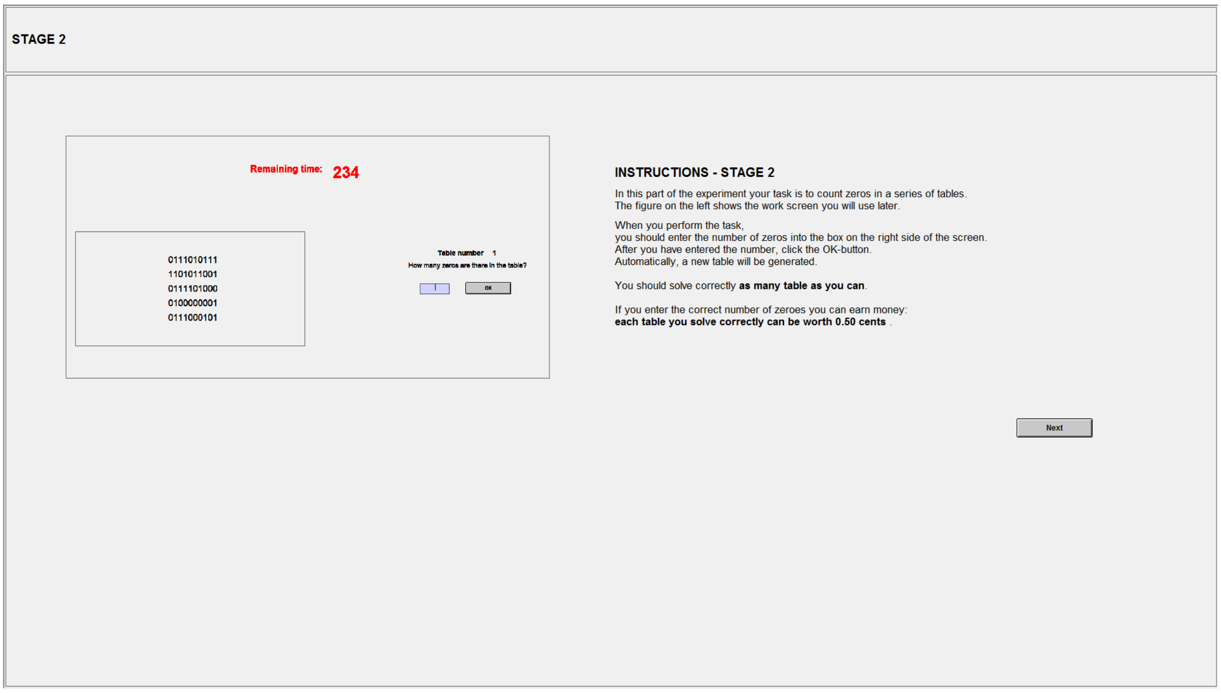
Remember, if your guess is **within five percentage points of the realization** ,
we reward the accuracy of each guess with **2 Pounds**

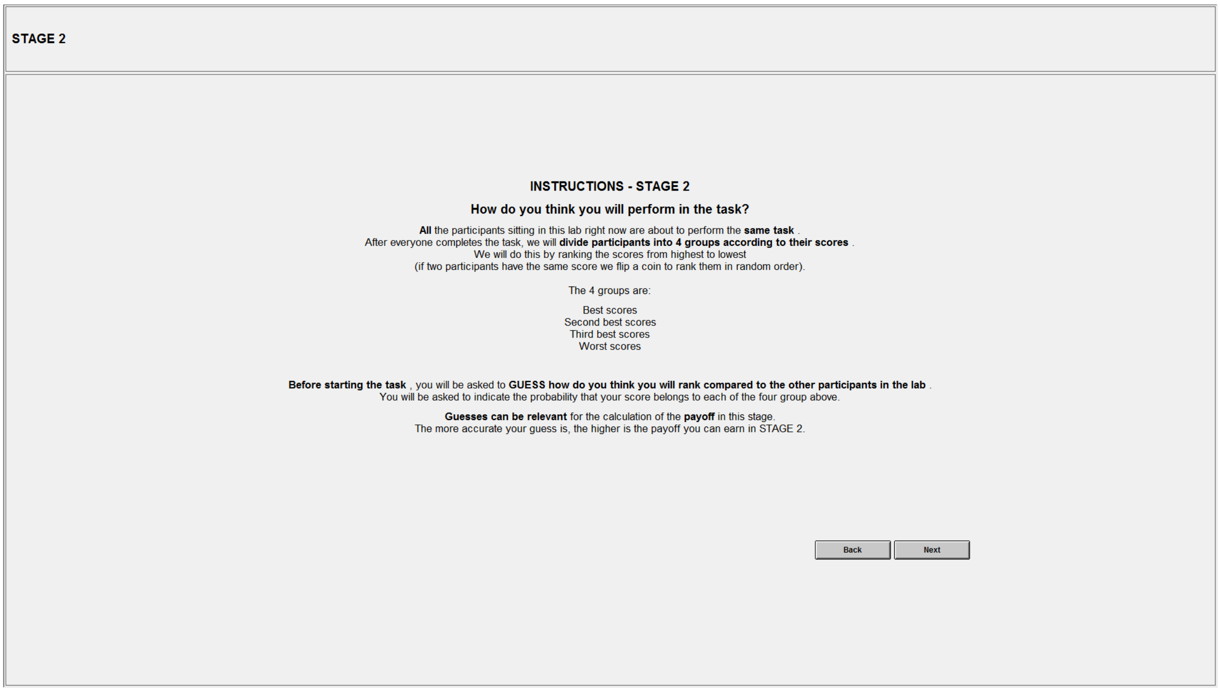
Confirm

All participants



C.2 Effort task





STAGE 2

INSTRUCTIONS - STAGE 2

How is the accuracy of my guess rewarded?

As a general rule, the more precise you are with your guesses the higher your reward is going to be.

More specifically, the accuracy of your guess is rewarded according to the formula: **12+12 log(p)**
where p is the probability you assigned to the group that contains your actual score from the task.
This means that **the reward is proportional to the probability assigned to the correct group** .

Consider the following example:
if your score is the best and you assign probability 100% to be in the "Best scores" group
you earn the maximum amount possible. Your guess is rewarded with 12 Pounds.

However, if your score is the best but your guess is inaccurate we penalize your payoff from the guess.
So, if your score is the best and your guess on "Best scores" group is 80%, your payment is given by 12+12 log(0.8)=12 + 12 (-0.1).
Your payoff from the guess is 12 - 1.2 = 10.8 Pounds.

One last example: if your score is in the "Second best scores" and you assigned 50% probability of belonging to that group,
your reward is given by 12+12 log(0.5)=12 + 12(-0.3).
Your payoff from the guess is 12 - 12*0.3 =8.4 Pounds.

This means that the more accurate is your guess the higher is your reward.

Back Next

STAGE 2

INSTRUCTIONS - STAGE 2

How is the accuracy of my guess rewarded?

As a general rule, the more precise you are with your guesses the higher your reward is going to be.

More specifically, the accuracy of your guess is rewarded according to the formula: **12+12 log(p)**
where p is the probability you assigned to the group that contains your actual score from the task.
This means that **the reward is proportional to the probability assigned to the correct group** .

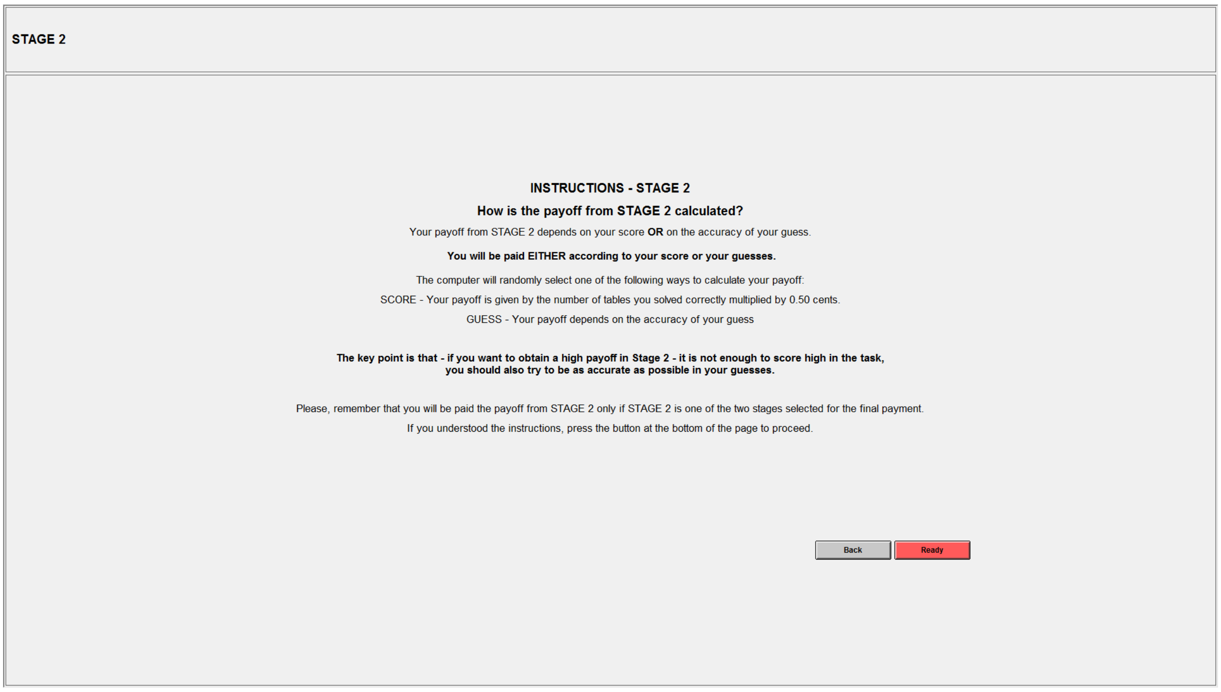
Consider the following example:
if your score is the best and you assign probability 100% to be in the "Best scores" group
you earn the maximum amount possible. Your guess is rewarded with 12 Pounds.

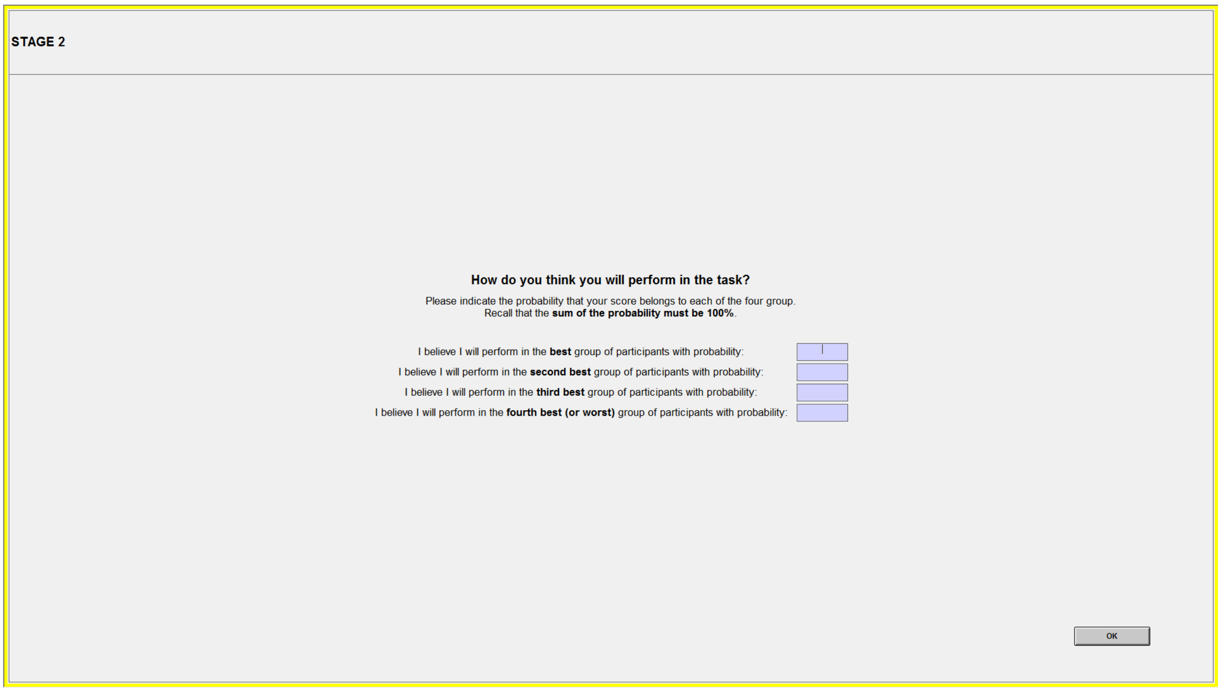
However, if your score is the best but your guess is inaccurate we penalize your payoff from the guess.
So, if your score is the best and your guess on "Best scores" group is 80%, your payment is given by 12+12 log(0.8)=12 + 12 (-0.1).
Your payoff from the guess is 12 - 1.2 = 10.8 Pounds.

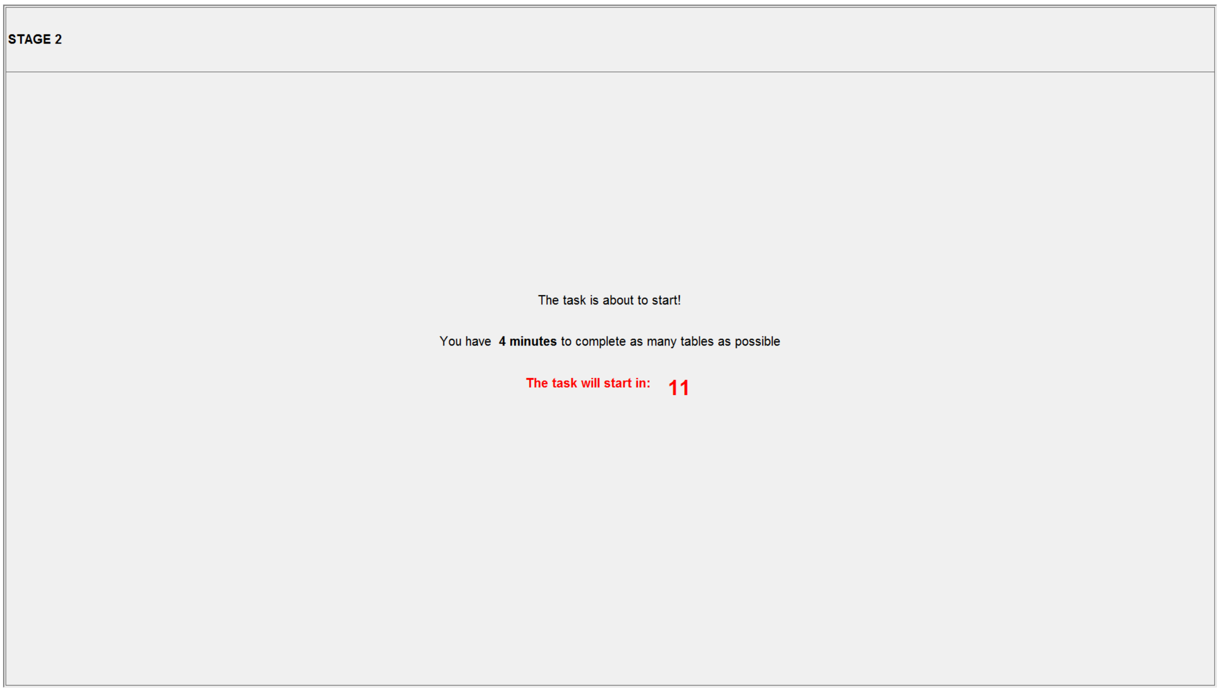
One last example: if your score is in the "Second best scores" and you assigned 50% probability of belonging to that group,
your reward is given by 12+12 log(0.5)=12 + 12(-0.3).
Your payoff from the guess is 12 - 12*0.3 =8.4 Pounds.

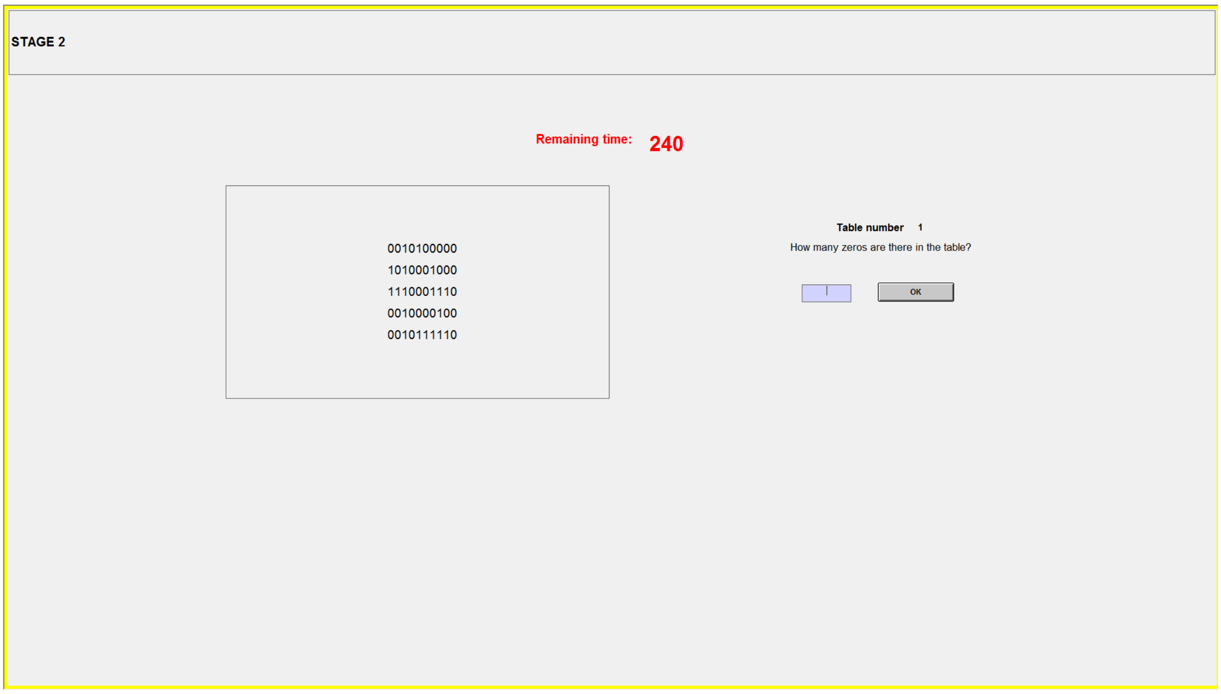
This means that the more accurate is your guess the higher is your reward.

Back Next


STAGE 2
INSTRUCTIONS - STAGE 2
How is the payoff from STAGE 2 calculated?
Your payoff from STAGE 2 depends on your score OR on the accuracy of your guess.
You will be paid EITHER according to your score or your guesses.
The computer will randomly select one of the following ways to calculate your payoff:
SCORE - Your payoff is given by the number of tables you solved correctly multiplied by 0.50 cents.
GUESS - Your payoff depends on the accuracy of your guess
The key point is that - if you want to obtain a high payoff in Stage 2 - it is not enough to score high in the task, you should also try to be as accurate as possible in your guesses.
Please, remember that you will be paid the payoff from STAGE 2 only if STAGE 2 is one of the two stages selected for the final payment.
If you understood the instructions, press the button at the bottom of the page to proceed.
Back
Ready


STAGE 2
How do you think you will perform in the task?
Please indicate the probability that your score belongs to each of the four group.
Recall that the sum of the probability must be 100%.
I believe I will perform in the best group of participants with probability:
I believe I will perform in the second best group of participants with probability:
I believe I will perform in the third best group of participants with probability:
I believe I will perform in the fourth best (or worst) group of participants with probability:
OK


STAGE 2
The task is about to start!
You have 4 minutes to complete as many tables as possible
The task will start in: 11


STAGE 2
Remaining time: 240
0010100000
1010001000
1110001110
0010000100
0010111110
Table number 1
How many zeros are there in the table?
OK

STAGE 2

STAGE 2 ends here

You will be informed about the outcome of STAGE 2 at the end of the experiment.

Please press the button at the bottom of the page to proceed to STAGE 3.

OK

STAGE 3

INSTRUCTIONS STAGE 3

In this part of the experiment, you will be asked to participate in two games.

The payoff of STAGE 3 corresponds to the sum of the payoffs obtained in the two GAMES (1 and 2) that you are about to play in STAGE 3

Please, remember that you will be paid the payoff from STAGE 3 only if STAGE 3 is one of the two stages selected for the final payment.

When you are ready to proceed press the button at the bottom of the page.

Ready

STAGE 3 - GAME 1

INSTRUCTIONS STAGE 3 - GAME 1

In the next page you will find a **list of lotteries**
and you will be asked to **chose the one you would like to play** .

Each lottery reports two possible outcomes (A and B) that are equally likely to occur.
Hence, **every lottery will return each of the outcome with 50% probability** .

Before beginning the game, please consider the following lottery:

Outcome A	Outcome B
40	80

Please, answer the following questions based on the lottery above.
Those are trial questions and they are only meant to check if you understand how a lottery works.
I will not use your answer in this page to calculate your payoff in STAGE 3.

What is the probability (in %) that the outcome of the lottery will be 40?

What is the probability (in %) that the outcome of the lottery will be 60?

Confirm

STAGE 3 - GAME 1

STAGE 3 - GAME 1

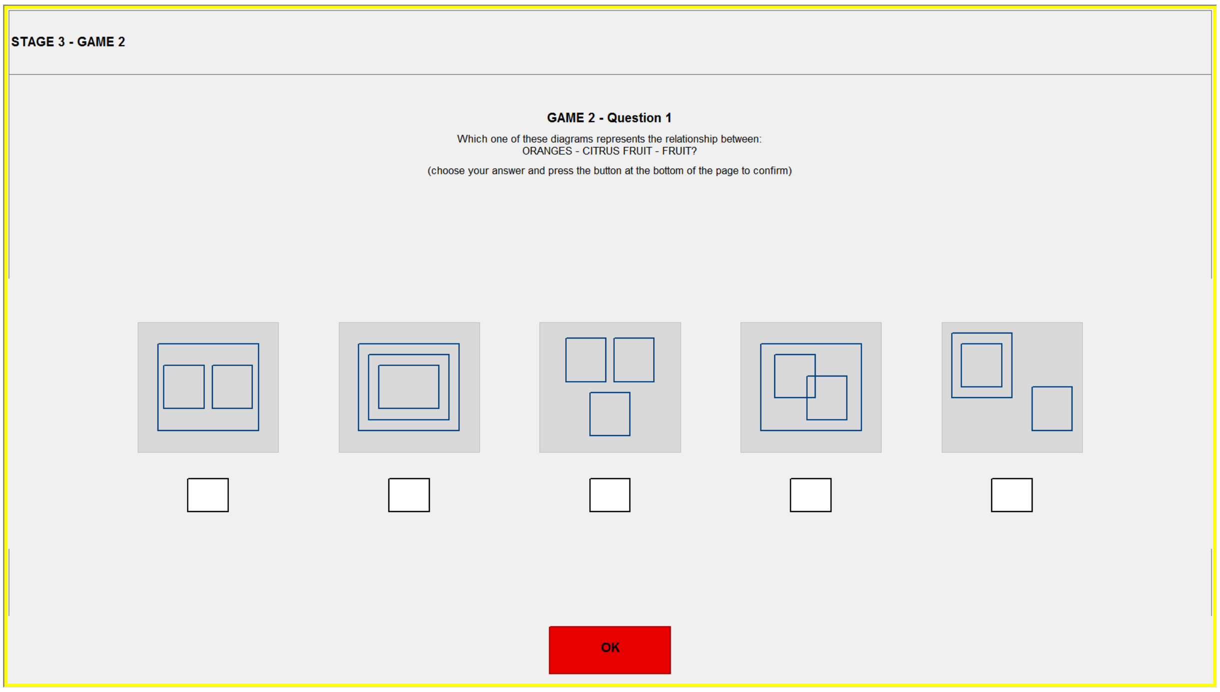
Please, select your preferred lottery from the list below.

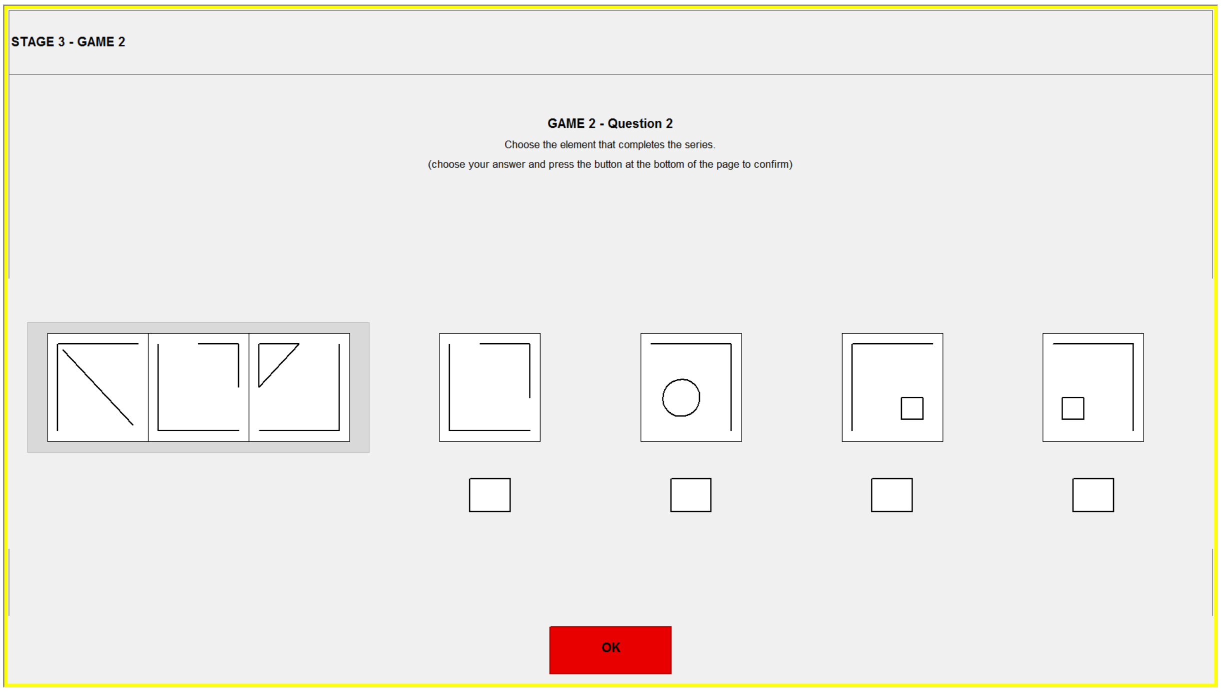
Once you confirm your choice, the computer will randomly (i.e., with 50% probability)
draw either Outcome A or Outcome B for the lottery you chose.
The outcome drawn by the computer will be your payoff in Pounds from GAME 1 of STAGE 3

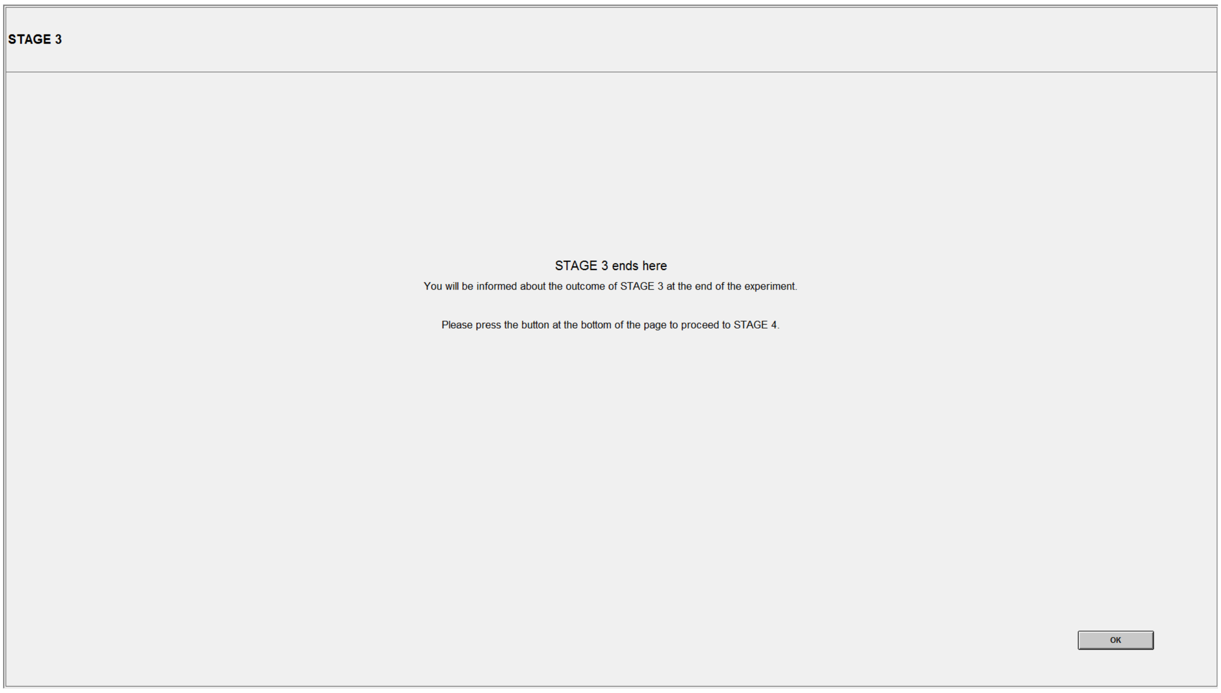
Lottery number	Outcome A	Outcome B
1	3.5 Pounds	3.5 Pounds
2	3 Pounds	4.5 Pounds
3	2.5 Pounds	5.5 Pounds
4	2 Pounds	6.5 Pounds
5	1.5 Pounds	7.5 Pounds
6	0.25 Pounds	8.75 Pounds

Please select your preferred lottery:
- Lottery 1
- Lottery 2
- Lottery 3
- Lottery 4
- Lottery 5
- Lottery 6

Confirm

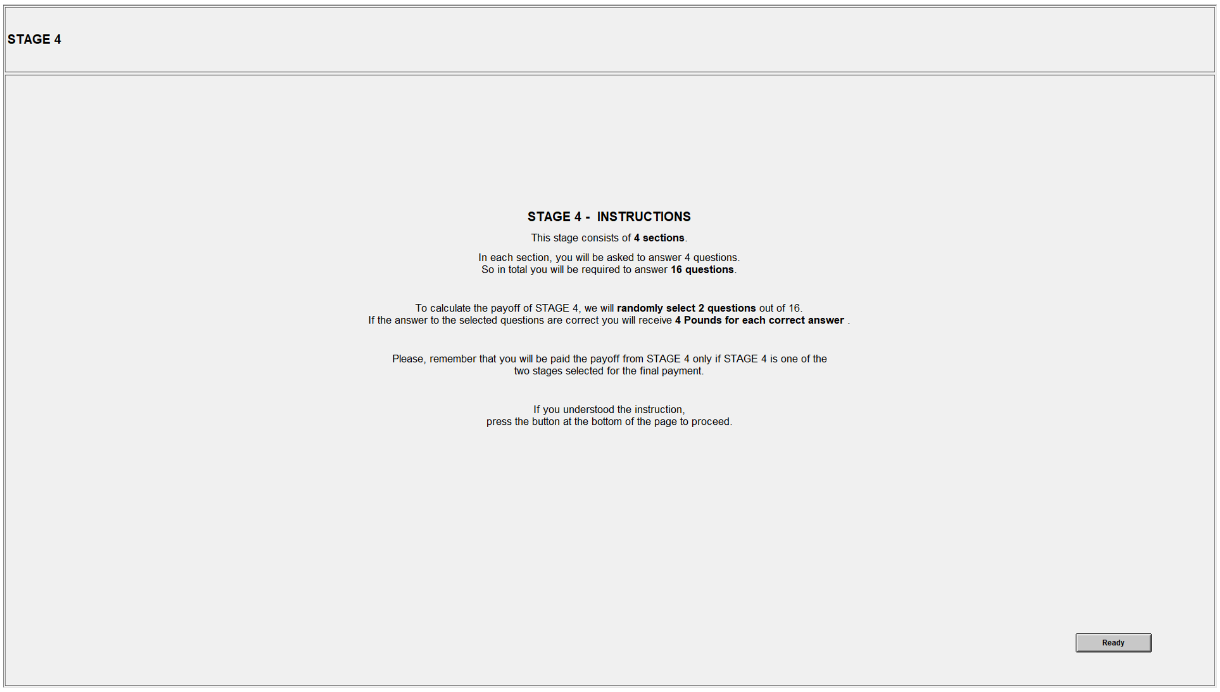

STAGE 3 - GAME 2
GAME 2 - Question 1
Which one of these diagrams represents the relationship between:
ORANGES - CITRUS FRUIT - FRUIT?
(choose your answer and press the button at the bottom of the page to confirm)
OK


STAGE 3 - GAME 2
GAME 2 - Question 2
Choose the element that completes the series.
(choose your answer and press the button at the bottom of the page to confirm)
OK


STAGE 3
STAGE 3 ends here
You will be informed about the outcome of STAGE 3 at the end of the experiment.
Please press the button at the bottom of the page to proceed to STAGE 4.
OK

C.3 Emotional ability

All the information concerning the task appeared directly on the screen. All participants faced the same items. Before the task, we distributed a handout with a definition of the emotions which appeared in the multiple choices.



Read the mind in the eyes

We selected 4 items of the 'Read the mind in the eyes' test revised version (Baron-Cohen et al., 2001). The items are reported below.

STAGE 4 - Section 1

Section 1

In this task you will watch pictures of eyes.
For each set of eyes, you should choose the word that best describes what the person in the picture is thinking or feeling.
Before making your choice,
make sure that you have read all 4 words that are given as options.

You may feel that more than one word is applicable but please choose just one word,
the word which you consider to be most suitable.
If you really don't know the meaning of a word you can look it up in the definition handout.

If you understood the instructions, press the button at the bottom of the page to proceed.

Ready

STAGE 4 - Section 1

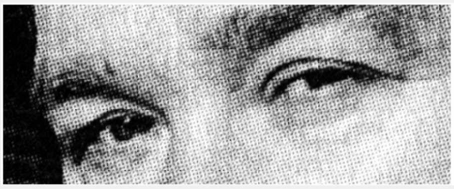
Section 1 - Question 1

Which word best describes what the person in the picture is thinking or feeling?
- grateful
- flirtatious
- hostile
- disappointed

Ready

STAGE 4 - Section 1

Section 1 - Question 2



Which word best describes what the person in the picture is thinking or feeling?

- indifferent
- embarassed
- skeptical
- dispirited

Ready

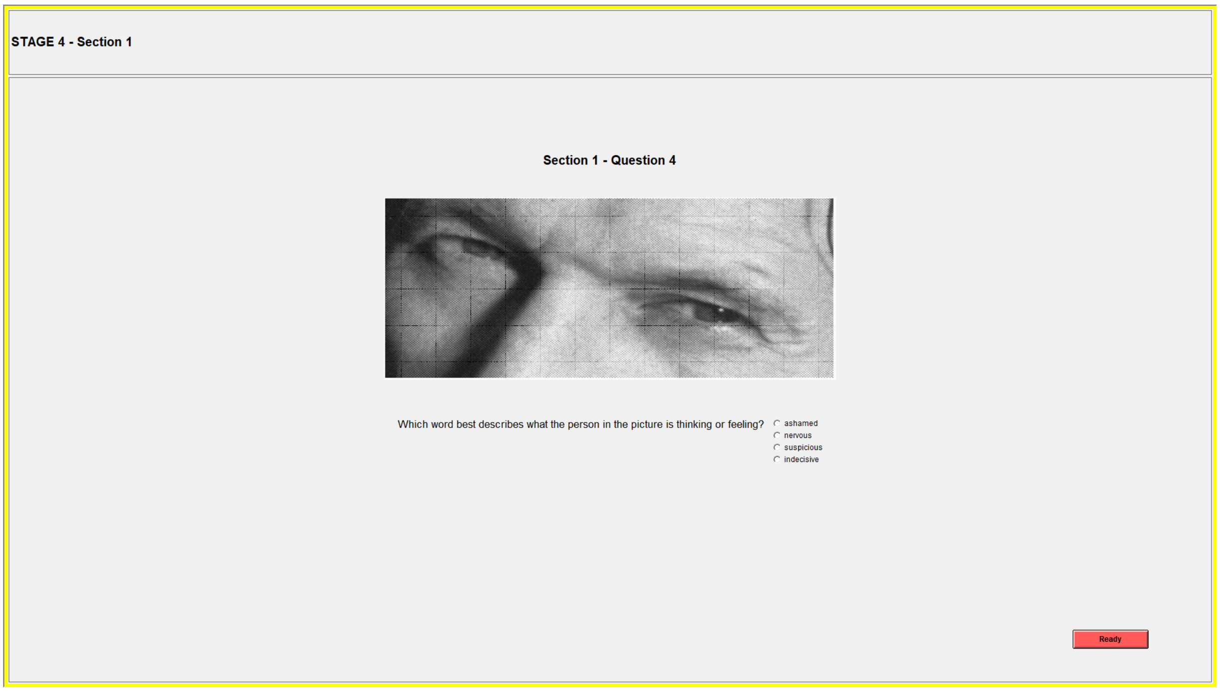
STAGE 4 - Section 1

Section 1 - Question 3

Which word best describes what the person in the picture is thinking or feeling?

- dominant
- friendly
- guilty
- horrified

Ready



The selection was based on the level of difficulty associated with each item. The difficulty is calculated as the percentage of people that correctly identified the emotion expressed in the eyes (the higher the percentage the lower the difficulty level). We selected two easy and two hard items. For more details, we refer to the paper by Baron-Cohen et al. (2001).

Read the mind in the voice

We selected 4 items of the ‘Read the mind in the voice’ test revised (Golan et al., 2007). Participants listened three times to the audio before being able to give their answer and proceed. The screenshots reported below show the spoken phrases and the choices available to answer.

STAGE 4 - Section 2

Section 2

You are about to listen to a series of voice recordings.
You should listen carefully not only to what the speaker is saying but also **HOW it is saying it** .
Then **you have to choose the word (out of 4 options) that best describes the way the speaker is feeling**.
You may feel that more than one word is applicable but please choose just one word,
the word which you consider to be most suitable.
Before making your choice, make sure that you have read all 4 words.
If you really don't know what a word means you can look it up in the definition handout.

You will listen to **4 short recordings, each one of them for three times** .

PLEASE PUT ON YOUR HEADPHONES

If you understood the instructions, press the button at the bottom of the page to proceed.

Ready

STAGE 4 - Section 2

Section 2 - Question 1

The speaker says: "Collie said you were up here"

Which word best describes what the speaker is feeling?
- surprised
- grateful
- friendly
- interested

STAGE 4 - Section 2

Section 2 - Question 2

The speaker says: "Please! We must go."

Which word best describes what the speaker is feeling?

- worried
- shy
- upset
- insulted

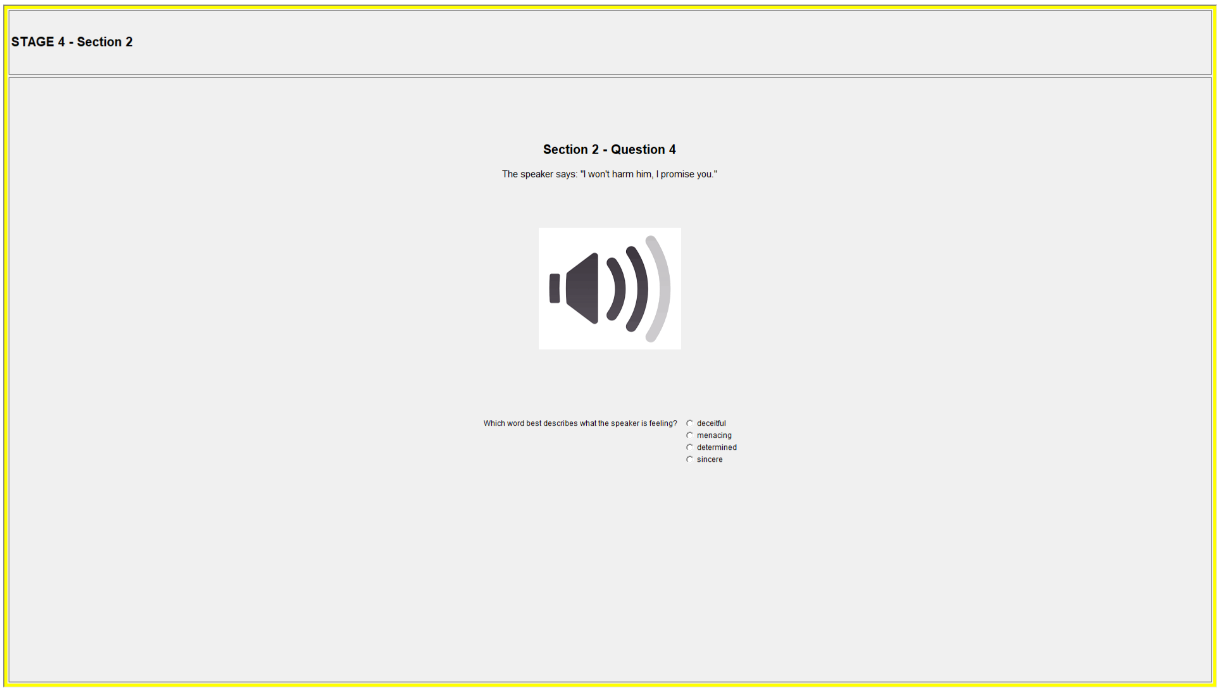
STAGE 4 - Section 2

Section 2 - Question 3

The speaker says: "Oh, my god!"

Which word best describes what the speaker is feeling?

- terrified
- broken
- frustrated
- angry



The selection was based on the level of difficulty associated with each item. As before, the difficulty is calculated as the percentage of people that correctly identified the emotion expressed in the eyes (the higher the percentage the lower the difficulty level). We selected two easy and two hard items. For more details, we refer to the paper by Golan et al. (2006).

Read the mind in films

We selected 4 items of the ‘Read the mind in films’ task Golan et al. (2007). Participants watched each video three times before being able to give their answer and proceed. The screenshots reported below show the instructions and the choices available to answer.

STAGE 4 - Section 3

Section 3

In this task you will watch scenes from films.

You will be asked to tell how one of the characters is feeling at the end of the scene .

We will tell you which character to focus your attention on and give you 4 possibile feelings.
This way you will have the chance to focus on the right character,
and to make sure you are familiar with all the presented answers before watching the scene.

After watching the scene, **choose the right word that best describes how this character is feeling**.

If you feel that more than one word is applicable, please choose the word which you consider to be the most suitable.

If you really don't know the meaning of a word you can look it up in the definition handout.

You will watch 4 short movies, each one of them for three times.

PLEASE PUT ON YOUR HEADPHONES

If you understood the instructions, press the button at the bottom of the page to proceed.

Ready

STAGE 4 - Section 3

Section 3 - Question 1

Focus on the **man** character

At the end of the scene, how is the man feeling?

- composed
- belittled
- humble
- uneasy

STAGE 4 - Section 3

Section 3 - Question 2

Focus on the **old lady** character

At the end of the scene, how is the old lady feeling?

- ○ wanting
- ○ complaining
- ○ enjoying
- ○ grateful

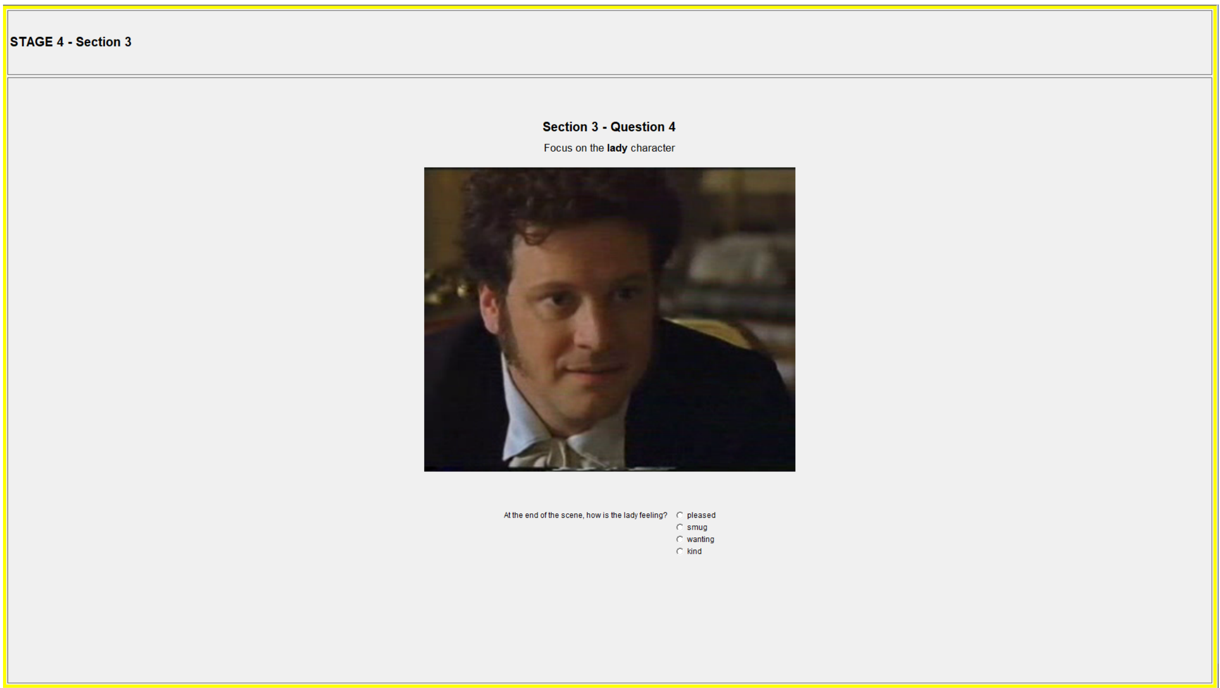
STAGE 4 - Section 3

Section 3 - Question 3

Focus on the **old man** character

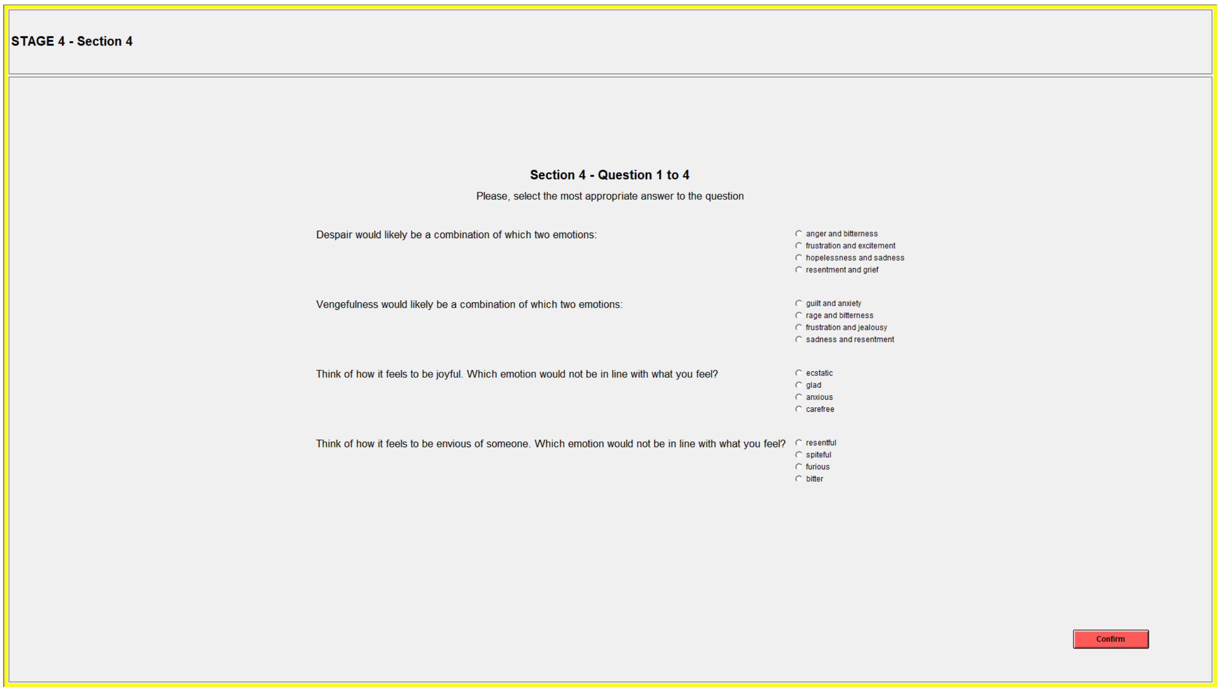
How is the old man feeling?

- ○ assertive
- ○ appalled
- ○ exasperated
- ○ tense



Verbal

The verbal recognition of emotions is a self-designed short test intended to measure the ability to describe emotions in words.



C.4 Trait Emotional Intelligence Questionnaire - Short Form

Instructions: Please answer each statement below by putting a circle around the number that best reflects your degree of agreement or disagreement with that statement. Do not think too long about the exact meaning of the statements. Work quickly and try to answer as accurately as possible. There are no right or wrong answers. There are seven possible responses to each statement ranging from ‘Completely Disagree’ (number 1) to ‘Completely Agree’ (number 7).

- Expressing my emotions with words is not a problem for me
- I often find it difficult to see things from another person’s viewpoint.
- On the whole, I’m a highly motivated person.
- I usually find it difficult to regulate my emotions.
- I generally don’t find life enjoyable.
- I can deal effectively with people.
- I tend to change my mind frequently.

- Many times, I can’t figure out what emotion I’m feeling.
- I feel that I have a number of good qualities.
- I often find it difficult to stand up for my rights.
- I’m usually able to influence the way other people feel.
- On the whole, I have a gloomy perspective on most things.
- Those close to me often complain that I don’t treat them right.
- I often find it difficult to adjust my life according to the circumstances.
- On the whole, I’m able to deal with stress.
- I often find it difficult to show my affection to those close to me.
- I’m normally able to ‘get into someone’s shoes’ and experience their emotions.
- I normally find it difficult to keep myself motivated.
- I’m usually able to find ways to control my emotions when I want to.
- On the whole, I’m pleased with my life.
- I would describe myself as a good negotiator.
- I tend to get involved in things I later wish I could get out of.
- I often pause and think about my feelings.
- I believe I’m full of personal strengths.
- I tend to ‘back down’ even if I know I’m right.
- I don’t seem to have any power at all over other people’s feelings.
- I generally believe that things will work out fine in my life.
- I find it difficult to bond well even with those close to me.
- Generally, I’m able to adapt to new environments.
- Others admire me for being relaxed.

D Appendix: Details on experimental sessions

Table D1 reports details on the experimental sessions. All the sessions have been run at the EssexLab of the University of Essex.

	Treatment	Participants	Date	Starting time	Duration
Session 1	N1	28	12/07/2018	14.00	90 min
Session 2	P1	28	12/07/2018	16.30	85 min
Session 3	N2	24	16/07/2018	14.00	95 min
Session 4	P2	20	16/07/2018	16.30	80 min
Session 5	P3	28	11/10/2018	14.00	80 min
Session 6	N3	24	11/10/2018	16.30	85 min
Session 7	P2	28	12/10/2018	14.00	85 min
Session 8	Baseline	28	12/10/2018	16.30	90 min
Session 9	P2	20	25/10/2018	13.05	80 min
Session 10	N1	24	26/11/2018	11.45	90 min
Session 11	P1	32	26/11/2018	14.00	85 min
Session 12	N2	28	26/11/2018	16.30	90 min
Session 13	P4	32	18/11/2019	14.00	80 min
Session 14	N4	24	18/11/2019	16.30	85 min

Table D1: Schedule of experimental sessions

Four observations were dropped from the analysis. Three participants had to leave before the end of the sessions (Session 5, and Session 9 twice), producing incomplete data. One participant was unable to understand any probability concept and could not complete the belief elicitation task without the help of the lab assistant (Session 12). Table D2 show the sample size by treatment.

HP treatment	Participants	HN treatment	Participants
HP-T1	60	HN-T1	52
HP-T2	66	HN-T2	51
HP-T3	27	HN-T3	24
HP-T4	32	HN-T4	24
Baseline	28		

Table D2: Sample size by treatment

E Appendix: Random assignment check

Table E1 and E2 report regressions of observables on treatments dummies. The percentage of female in the neutral condition (78%) is significantly higher than that in most of the treatments. None of the other observables is significantly predicted by the treatment.

	(1) Age	(2) Female	(3) Education	(4) Students	(5) Siblings	(6) Music
Negative - Timing 1	1.3104	-0.1896	0.0907	-0.0055	-0.3352	0.0616
	(2.3613)	(0.1161)	(0.1516)	(0.0721)	(0.3776)	(0.1178)
Positive - Timing 1	0.2310	-0.3024***	0.1214	-0.0119	-0.7429**	-0.0619
	(2.3056)	(0.1134)	(0.1480)	(0.0704)	(0.3687)	(0.1147)
Negative - Timing 2	4.1751*	-0.2171*	0.0371	-0.1246*	-0.1232	0.0616
	(2.3694)	(0.1165)	(0.1521)	(0.0724)	(0.3789)	(0.1178)
Positive - Timing 2	1.0476	-0.2706**	0.0942	-0.0346	-0.1580	0.0260
	(2.2720)	(0.1117)	(0.1458)	(0.0694)	(0.3633)	(0.1130)
Negative - Timing 3	-1.9524	-0.3274**	0.0714	0.0714	-0.7679*	0.0714
	(2.8022)	(0.1378)	(0.1799)	(0.0856)	(0.4481)	(0.1394)
Positive - Timing 3	0.6587	-0.2302*	-0.0675	-0.1138	-0.4392	0.1270
	(2.7171)	(0.1336)	(0.1744)	(0.0830)	(0.4345)	(0.1351)
Negative - Timing 4	-3.5774	-0.3690***	-0.0536	0.0714	-0.7679*	0.1964
	(2.6002)	(0.1381)	(0.1705)	(0.0788)	(0.4360)	(0.1390)
Positive - Timing 4	-3.8482	-0.2232*	0.0089	0.0714	-0.3616	-0.0536
	(2.4188)	(0.1285)	(0.1586)	(0.0733)	(0.4056)	(0.1293)
Observations	364	364	364	364	364	363
R-squared	0.0576	0.0295	0.0096	0.0501	0.0289	0.0201

Table E1: Random assignment check (1). Standard deviation in parenthesis. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Excluded category: neutral condition. Music indicates whether the participants has experience in playing an instrument.

	(1) IQ	(2) Risk aversion	(3) TEIQue-SF	(4) Emot. ability
Negative - Timing 1	0.1621	0.7225*	-0.0090	-0.0824
	(0.1746)	(0.3673)	(0.1790)	(0.5481)
Positive - Timing 1	0.2262	0.1905	-0.1872	-0.2619
	(0.1705)	(0.3586)	(0.1748)	(0.5352)
Negative - Timing 2	0.1870	0.2591	0.0956	-0.4090
	(0.1752)	(0.3685)	(0.1796)	(0.5500)
Positive - Timing 2	0.0747	0.5390	-0.0482	0.1017
	(0.1680)	(0.3534)	(0.1722)	(0.5274)
Negative - Timing 3	0.1845	-0.0595	-0.0847	-0.9286
	(0.2072)	(0.4359)	(0.2124)	(0.6505)
Positive - Timing 3	0.0410	0.0238	-0.1478	0.2751
	(0.2009)	(0.4226)	(0.2060)	(0.6307)
Negative - Timing 4	0.3929*	0.6071	-0.2861	-0.1786
	(0.2042)	(0.4333)	(0.2088)	(0.6520)
Positive - Timing 4	0.0491	0.6071	-0.2073	-0.8036
	(0.1899)	(0.4031)	(0.1942)	(0.6065)
Observations	364	364	364	364
R-squared	0.0180	0.0288	0.0217	0.0204

Table E2: Random assignment check (2). Standard deviation in parenthesis. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Excluded category: neutral condition.

F Appendix: Emotional intelligence randomization check

F.1 Emotional ability

Figure F1 shows the range of values and means of the emotional ability task scores across treatments (left figure) and 95% confidence interval (right figure). The task scores seem to be balanced across treatments. We perform a pairwise comparison across treatments from a linear regression model where emotional ability is regressed on treatment dummies. The results indicate that only a few pairwise differences are marginally significant (Positive - Timing 2 and Positive - Timing 4, Positive - Timing 3 and Positive - Timing 4, Positive -Timing 3 and Negative - Timing 3).

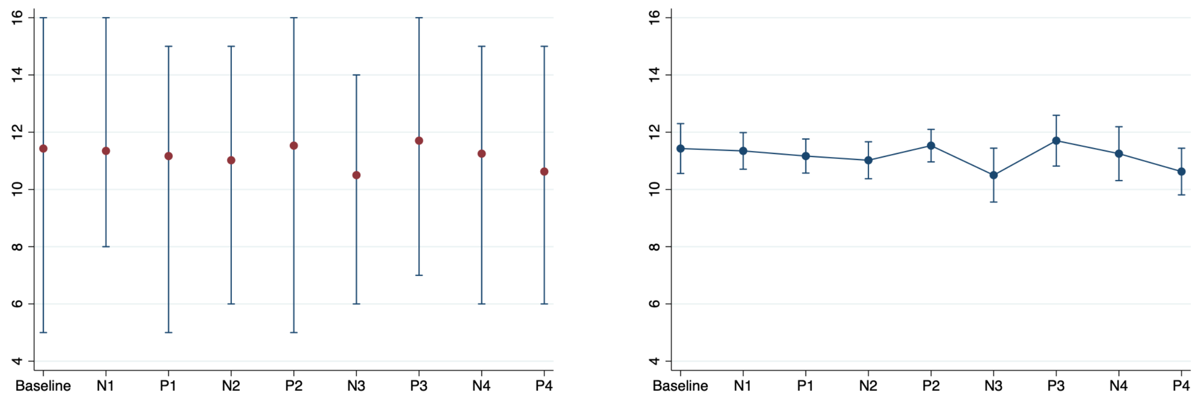


Figure F1: Left: range of values and means of the TEIQue-SF by treatment. Right: means of the TEIQue-SF by treatment with 95% confidence intervals

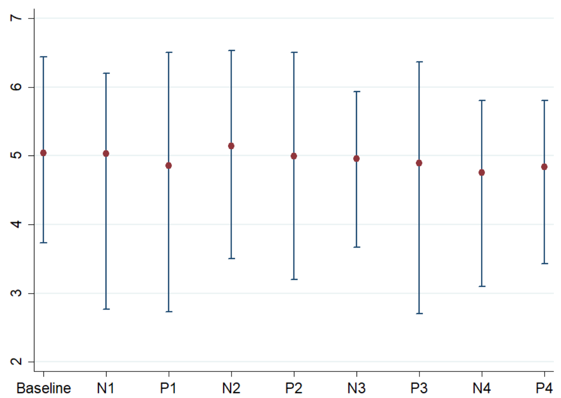
Another concern might be that the AIP influences the response to the emotional ability task. We can rule out this concern for two reasons. First, the mean values in the baseline condition are not significantly different from the rest of the treatments. Second, if we regress our measure of emotional ability on dummies indicating the timing of treatment (table F1), none of the coefficient is significant.

	Emotional ability
Timing 1	-0.1786
	(0.4965)
Timing 2	-0.1209
	(0.4944)
Timing 3	-0.2913
	(0.5527)
Timing 4	-0.5357
	(0.5439)
Observations	364
R-squared	0.0043

Table F1: Treatment timing and emotional ability. Standard deviation in parenthesis. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Excluded category: neutral condition.

F.2 Trait Emotional Intelligence

Figure F2 shows the range of values and means of the TEIQue-SF scores across treatments (left figure) and 95% confidence interval (right figure). The questionnaire scores seem to be balanced across treatments. We perform a pairwise comparison across treatments from a linear regression model where TEIQue-SF is regressed on treatment dummies. The results indicate that the only significant differences in means are between Negative - Timing 2 and: Positive - Timing 1 (difference=0.2828, p-value=0.049), Negative - Timing 4 (difference=-0.3817, p-value=0.041), and Positive - Timing 4 (difference=-0.3029, p-value=0.074).



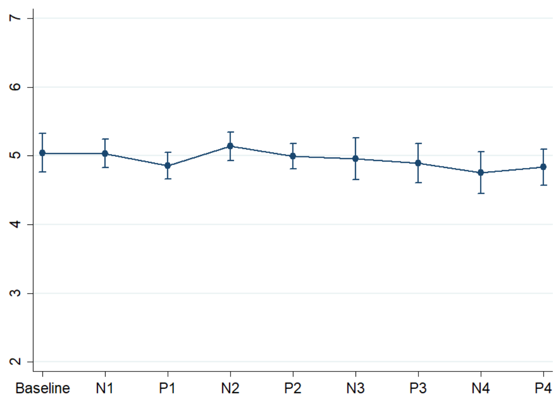


Figure F2: Left: range of values and means of the TEIQue-SF by treatment. Right: means of the TEIQue-SF by treatment with 95% confidence intervals

Another concern might be that the AIP influences the response to the questionnaire. We can rule out this concern for three reasons. First, it should be noted that there is a minimum interval of 30 minutes (in Timing 4 treatments) between the AIP and the questionnaire. During the interval, participants complete a series of tasks (risk task, IQ type questions, and emotion ability questions). Second, the mean values in the baseline condition are not significantly different from the rest of the treatments. Third, if we regress TEIQue-SF on dummies indicating the timing of treatment (table F2), none of the coefficient is significant nor there is a temporal pattern in the dimensions or sign of the coefficients.

	TEIQue-SF
Timing 1	-0.1045
	(0.1583)
Timing 2	0.0145
	(0.1576)
Timing 3	-0.1181
	(0.1762)
Timing 4	-0.2411
	(0.1734)
Observations	364
R-squared	0.0138

Table F2: Treatment timing and TEIQue-SF

G Appendix: Descriptives

Variable name	Value Range	Mean (SD)	Min Value	Max Value
Dictator's choice (discrete)	1-3	1.55 (0.5)	1	3
Recipient's first order beliefs on (7,3)	0-1	0.49 (0.22)	0.05	1
Recipient's first order beliefs on (5,5)	0-1	0.40 (0.19)	0	0.90
Recipient's first order beliefs on (3,7)	0-1	0.11 (0.11)	0	0.9
Dictator's second order beliefs on (7,3)	0-1	0.49 (0.19)	0	1
Dictator's second order beliefs on (5,5)	0-1	0.37 (0.15)	0	0.75
Dictator's second order beliefs on (3,7)	0-1	0.14 (0.12)	0	0.75

Table G1: DG - Summary statistics. Dictator's choice coding is as follow: 1 corresponds to (7, 3), 2 to (5, 5), and 3 to (3, 7).

Variable name	Value Range	Mean (SD)	Min Value	Max Value
Beliefs on best quartile	0-1	0.45 (0.28)	0	1
Beliefs on second best quartile	0-1	0.33 (0.19)	0	1
Beliefs on second worst quartile	0-1	0.15 (0.15)	0	1
Beliefs on worst quartile	0-1	0.68 (0.89)	0	0.7
Number of correct tables completed	0-26	13.66 (4.09)	1	26
Number of attempted tables	1-40	16.97 (3.96)	3	29
Accuracy in counting tables	0-1	0.80 (0.13)	0.17	0.96

Table G2: ET - Summary statistics

Variable name	Value Range	Mean (SD)	Min Value	Max Value
TEIQue-SF	0-7	4.96 (0.75)	2.7	6.53
TEI Well Being	0-7	5.36 (1.08)	1.33	7
TEI Emotionality	0-7	5.02 (0.88)	2.88	7
TEI Self-control	0-7	4.47 (1.05)	1.17	6.67
TEI Sociability	0-7	4.99 (0.97)	2.17	7
TEI Emotion Regulation	0-7	4.65 (1.01)	1.25	7
Ability to recognize emotions	0-16	11.21 (2.34)	5	16
Eyes	0-4	3.19 (0.85)	0	4
Voices	0-4	2.57 (1.01)	0	4
Video	0-4	2.71 (0.96)	0	4
Words	0-4	2.73 (1.08)	0	4

Table G3: Emotional intelligence - Summary statistics

H Appendix: Emotional intelligence and demographics

Table H1 reports pairwise correlation coefficients between the two measures of emotional intelligence and demographics. Emotional 'ability' does not correlate significantly with any variable. Trait emotional intelligence is positively correlated with age and marginally with risk aversion but not with other variables. The two measures of emotional intelligence are not correlated between them (Pearson's correlation coefficient=-0.0746, p-value=0.1554, sample size=364).

	Emotional 'ability'	TEIQue-SF
Age	-0.0362	0.1215**
Female	0.0399	0.0176
Education	-0.0498	0.0057
Student	0.0449	-0.0354
Number of siblings	0.0018	-0.0097
Music	-0.0086	-0.0117
IQ	-0.0009	-0.0648
Risk Aversion	0.0322	0.0887*

Table H1: Pairwise correlation. Significance level $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Music indicates whether the participants has experience in playing an instrument. Sample size: 364.

I Appendix: Analysis with emotional ability

We estimate the following equation:

$$Y_i = \alpha + \beta_1 \text{Treatment}_i + \beta_2 \text{Treatment}_i \times \text{EA}_i + \beta_3 \text{EA}_i + \gamma_1 X_{i,1} + \ldots + \gamma_M X_{i,M} + \epsilon_i \tag{8}$$

where EA_i is the measure of emotional intelligence presented in section 3 as belonging to the 'ability' approach. The rest of the equation is as described in section 4.3. Emotional ability is centred around the mean.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Selfish choice	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25	Correct tables
Timing of treatment:	Timing 1	Timing 2	Timing 2	Timing 3	Timing 4
Negative × EA	-0.0179	0.0127	-0.0145	0.0096	-0.4961
	(0.0580)	(0.0224)	(0.0270)	(0.0244)	(0.4582)
Positive × EA	-0.0193	-0.0118	0.0050	0.0051	-0.3446
	(0.0459)	(0.0177)	(0.0161)	(0.0185)	(0.4254)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	181	181	183	364	364
R-squared	0.0658	0.0583	0.1214	0.0733	0.1034

Table I1: OLS Regressions - Average Treatment Effects. Robust standard errors in parentheses. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes.

	Bottom EA percentile			Top EA percentile		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25
Timing 1	0.0754	0.0064	-0.0365	0.0195	-0.0292	0.1712**
	(0.0558)	(0.0796)	(0.0658)	(0.0591)	(0.0863)	(0.0804)
Timing 2	0.0129	0.0246	-0.0662	0.0328	-0.0456	0.1420*
	(0.0654)	(0.0704)	(0.0695)	(0.0605)	(0.0768)	(0.0721)
Timing 3			-0.0141			0.0612
			(0.0777)			(0.0831)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	66	68	134	56	57	113
R-squared	0.2372	0.0977	0.0806	0.1874	0.1845	0.1199

Table I2: OLS Regressions, split sample. Robust standard errors in parentheses. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes.

J Appendix: Robustness checks

J.1 Analysis using Trait Emotional Intelligence quartiles

	Bottom TEIQue-SF quartile			Top TEIQue-SF quartile		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25
Timing 1	-0.0241	-0.0618	0.0791	-0.0636	0.0128	-0.1775*
	(0.0764)	(0.0791)	(0.0805)	(0.0673)	(0.0717)	(0.0930)
Timing 2	0.0390	-0.0532	0.0383	-0.1909***	0.1493*	-0.1697*
	(0.0798)	(0.0783)	(0.0782)	(0.0652)	(0.0875)	(0.0985)
Timing 3			0.0533			-0.3324***
			(0.1006)			(0.1011)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	43	53	96	42	42	84
R-squared	0.1658	0.2511	0.1540	0.3323	0.2226	0.1584

Table J1: OLS Regressions, split sample. Robust standard errors in parentheses. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes.

J.2 Regression analysis on treated

Exploiting the correlation between music appreciation and AIP efficacy (see B), we construct a sub-sample of 'treated' participants by dropping form participants who liked at least 'Moderately' (i.e., score above 3) Merzbow and 'Not at all' or 'A little' (i.e., score below 3) Delibes. We then repeat the the analysis of section 4.2 and 4.3 on the sub-sample of 'treated'.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Selfish choice	Dictators' $SOB_{(7,3)}$	Recipients' $FOB_{(7,3)}$	Beliefs Top25	Correct tables
Timing of treatment:	Timing 1	Timing 2	Timing 2	Timing 3	Timing 4
Negative	0.0734	0.0125	-0.0661	-0.0022	-1.2852
	(0.1253)	(0.0598)	(0.0559)	(0.0725)	(1.1607)
Positive	0.2629**	0.0700	-0.0857*	-0.0652	-0.3609
	(0.1210)	(0.0488)	(0.0494)	(0.0588)	(1.0542)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	147	147	157	304	304
R-squared	0.0786	0.0542	0.0871	0.0680	0.0859

Table J2: OLS Regressions - Average treatment effect on treated. Robust standard errors in parentheses. Significance level $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes.

Dependent variable:	(1) Selfish choice	(2) Dictators' $SOB_{(7,3)}$	(3) Recipients' $FOB_{(7,3)}$	(4) Beliefs Top25	(5) Correct tables
Timing of treatment:	Timing 1	Timing 2	Timing 2	Timing 3	Timing 4
Negative × TEI	-0.0127	-0.0635	0.1230**	-0.2530**	0.9232
	(0.1510)	(0.0506)	(0.0622)	(0.1054)	(1.4995)
Positive × TEI	0.3348***	-0.1279**	-0.0042	-0.1226	-1.5885
	(0.1133)	(0.0594)	(0.0734)	(0.0756)	(1.7724)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	147	147	157	304	304
R-squared	0.1182	0.1221	0.1308	0.1019	0.1177

Table J3: OLS Regressions - Treated subsample. Robust standard errors in parentheses. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes.

	Bottom TEIQue-SF percentile			Top TEIQue-SF percentile		
Dependent variable:	(1) Dictators' $SOB_{(7,3)}$	(2) Recipients' $FOB_{(7,3)}$	(3) Beliefs Top25	(4) Dictators' $SOB_{(7,3)}$	(5) Recipients' $FOB_{(7,3)}$	(6) Beliefs Top25
Timing 1	-0.0591	-0.0482	0.0505	-0.0443	0.0409	-0.1524**
	(0.0782)	(0.0743)	(0.0752)	(0.0574)	(0.0580)	(0.0751)
Timing 2	0.0734	-0.1311*	0.0329	-0.1669**	0.1759**	-0.1747**
	(0.0647)	(0.0769)	(0.0756)	(0.0659)	(0.0733)	(0.0749)
Timing 3			0.1714*			-0.3328***
			(0.0908)			(0.0751)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	46	58	104	46	53	99
R-squared	0.1621	0.1726	0.1470	0.2598	0.2863	0.2235

Table J4: OLS Regressions, split sample - Treated subsample. Robust standard errors in parentheses. Significance level *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Controls: age, gender, number of siblings, education, cognitive skills, risk attitudes.