Social reference points: Pre-analysis plan*

Jan Schmitz, Julien Senn & Christian Zehnder

August 6, 2018

*A previous version of this PAP was uploaded on June 8, 2017 (AEARCTR-0002260). The main hypotheses that we want to test are the same as the ones outlined in our previous PAP. The following details have changed: 1) We initially planned to use z-Tree for the data collection. Software tests (conducted after uploading the first PAP) have revealed unforeseen technical issues with some of the high-frequency features of our setup. For this reason, we decided to switch to an online software. 2) To increase power, the number of possible reference players has been reduced from four to three. 3) For ease of implementation we replaced the slider task with the "pushing-button" task recently used in DellaVigna and Pope (2017). 4) We added new treatments to study how social reference points compare to and interact with incentives.

Contents

1 Introduction

This document describes the design and the analysis plan for an experiment aimed at evaluating the effects of social reference points on effort provision. In the context of our study, social reference points are understood as peers with which individuals compare themselves. Broadly speaking, our experiment is aimed at better understanding a) whether exogenously assigned social reference points affect performance, b) whether endogenously chosen social reference points affect performance differently than exogenously assigned social reference points, c) how people choose their social reference points, and d) how social reference points compare to and interact with incentives. We also plan to study heterogeneity in treatment effects along several dimensions.

Our experiment is centered around the real-effort task used in DellaVigna and Pope (2017). The task involves alternating presses of 'a' and 'b', achieving a point for each a-b alternation. There are two production phases with a duration of 5 minutes. Subjects are instructed to score as many points as possible. They are paid a flat show-up fee of USD 1.5.

Our experiment involves two different sets of participants: the baseline participants and the main participants. The baseline participants receive information about their own performance only. Their performance can therefore not be influenced by any sort of comparison with other players. Our main participants, in contrast, will be exposed to various types of social reference points. In most treatments participants do not face monetary incentives to perform, but we also implement a set of incentive treatments in which participants receive a piece rate.

This document is organized as follows. Section 2 describes the design used with our baseline participants. Section 3 describes the experiment involving the main participants. Section 4 outlines our main behavioral predictions. Section 5 describes the setting of the experiment (sample size, power, etc.). Section 6 details our empirical strategy.

2 Baseline participants

We collected data on 60 baseline participants. At the beginning of the experiment, subjects were informed that the experiment consists of several parts. Instructions were displayed on participants' screens part by part, e.g. they received the instructions for Part 2) only after having completed Part 1) of the experiment. Thus, the participants were unaware of what would come next as they were completing the experiment. In the following, we describe the different parts faced by the baseline participants.

- At the very beginning of the experiment, we collected questionnaire data on sociodemographics (See Appendix C.1)
- In Part 1) of the experiment, we measured subjects' performance in the a-b task (effort_1). Subjects were informed about the task (including a short trial example) and the time at their disposal. While working on the task participants were constantly updated on their current score (graphically represented as a growing

vertical bar) and the remaining time. A screenshot is provided in Appendix (see figure B1). Upon completion of the task, these baseline participants learned their total score. They were then invited to answer 3 questions aimed at assessing whether they are a) satisfied with their performance, b) stressed, and c) found the task difficult.

- In Part 2) of the experiment, we measured subjects' performance in the same a-b task again (effort_2). The conditions remained exactly the same as in part 1.

- At the end of the experiment, we collected additional questionnaire data (see section 3.3.2). We also elicited subjects' coefficient of loss aversion. Following Abeler et al. (2011), we look for a participant's switching point between accepting and refusing a 50-50 (hypothetical) lottery yielding a loss of USD X or a gain of USD 6, where $X = \{2, 3, 4, 5, 6, 7\}$.

The data on the baseline participants has already been collected. These sessions were run prior to submitting the pre-analysis plan because baseline participants are not at the center of our interest, i.e. they are not assigned to any treatment. The performance of the baseline participants will serve as (potential) social reference points for all the main participants who will take part to future sessions. Using an independent and isolated set of participants to create the social reference points has several advantages:

- There are no strategic interactions between the main participants and the baseline participants. This avoids simultaneity problems that could emerge if the main participants and the baseline participants could observe each other.

- All main participants in our various treatments will face the exact same social reference points.

Descriptive statistics on the baseline participants can be found in Appendix A.

3 Main participants: treatment variations

We now come to the description of our main experiment.

In all the treatments, Part 1) (first performance measure in a-b task, effort_1) is identical to the corresponding parts detailed for the baseline participants above. Part 2) varies from one treatment to another.

3.1 Rank-only treatment (RANKONLY)

In the beginning of Part 2), participants in the RANKONLY treatment

- are informed that their performance will now be compared to the performance of 60 other participants (i.e. the baseline participants) who have completed the exact same task at an earlier point in time.

- learn their rank (a number between 1 and 61) within the specified comparison sample. (See screenshot B2 in Appendix for details).

We then measure their performance at the a-b task in Part 2) ($effort_2$). Note that in this treatment, participants only see information about their own performance as they are completing Part 2).[1]

3.2 Social reference point variations

For all the treatments in which we vary the social reference point, we initialize this reference point in the beginning of Part 2) using the following procedure:

- Participants are informed that their performance will now be compared to the performance of 60 other participants (i.e. the baseline participants) who have completed the exact same task at an earlier point in time.[2]
- Participants learn their rank (a number between 1 and 61) within the specified comparison sample. (See screen B2 in Appendix for details).[3]
- Participants are informed that they will have to complete the task again and that they *may* get real-time information about the performance of another player who has completed the exact same task before (comparison to a reference player). They are informed that these three other participants represent different levels of performance. More specifically, they are told that
  - The participant at rank 4 achieved a high performance in Round 1.[4]
  - The participant at rank 26 achieved a medium performance in Round 1.
  - The participant at rank 49 achieved a low performance in Round 1.
- If a reference player is assigned to a participant, the participant can see the development of his own current performance and the real-time development of the performance of the reference player while completing Part 2) (both performances are visually represented by growing vertical bars, see figure B7 in Appendix).
- The way in which social reference points are assigned to participants depends on the treatment (as explained in detail below).

Importantly, the 3 potential reference players will *always* be the same for all the participants of the main study. Details on the 3 baseline participants that are used as "potential reference players" can be found in Figure A.2 in Appendix.

3.2.1 Exogenous treatments (EXO)

In the exogenous treatment, subjects are *exogenously* assigned either one of the 3 potential reference players or no reference player at all (see Figure B3 to B5 in Appendix).

[1] Hence, this treatment is almost identical to the baseline condition, with the exception that participants get to see their ranking within the set of baseline participants before starting Part 2).

[2] Same as in section 3.1.

[3] Same as in section 3.1.

[4] By ranked, we mean the distribution of 60 baseline participants (based on their performance in round 1, i.e. based on effort1). Note that, depending on own performance, the rank of some (all) of these three reference players might shift up by one unit.

- Participants who are not assigned a reference player will complete Part 2) of the experiment in the exact same conditions as in Part 1), i.e. without receiving any information about the performance of another player.

- If they are assigned a social reference point (in the form of a reference player) in Part 2), participants will constantly see how the score of their reference player evolved as he or she was completing Part 2) of the experiment (see figure B7 in Appendix).

In what follows, we use the following terminology

- EXO-NO: Exogenously assignment to NO reference player
- EXO-LO: Exogenously assignment to the LOW reference player (ranked 49)
- EXO-MID: Exogenously assignment to the MEDIUM reference player (ranked 26)
- EXO-HI: Exogenously assignment to the HIGH reference player (ranked 4)

3.2.2 Optimal exogenous treatment (EXO-BEST)

If exogenously assigned social reference points do affect players' performance, random assignment of participants to different reference players (as described in the exogenous treatment in Section 3.2.1 above) may not maximize performance.[5] To create a valid benchmark for the best possible performance with exogenous assignment of social reference points, we plan to run the so-called EXO-BEST treatment. In this treatment, we will use the data collected in the EXO treatment to predict the optimal social reference point for each participant (based on his or her gender and performance in round 1).[6] We will then assign the best reference player to each participant, i.e. each participant in EXO-BEST will complete Part 2) while seeing (if applicable) how the score of their optimal reference player evolved as he/she was completing the experiment. For more details on the procedure used to predict optimal reference players, see section 6.4.2.

3.2.3 Endogenous treatment (ENDO)

In the endogenous treatment, each participant has the possibility to select one of the potential reference players (or no other player) as his or her reference player for the second part of the experiment (see Figure B6 in Appendix).

- If participants choose no reference player, they will complete Part 2) of the experiment in the exact same conditions as in Part 1), i.e. without receiving any information about the performance another player.

- If participants choose to see a reference player, they will be able to constantly see how the performance of their reference player evolved as he or she was completing Part 2) of the experiment (see figure B7 in Appendix).

[5] The reason is that some players may be assigned suboptimal social reference points (too weak or too strong).

[6] Note that it might be optimal for some participant to be assigned no reference player.

3.2.4 Monetary incentives treatments

In order to assess the magnitude of the effects of social reference points, we compare them to a PIECERATE treatment in which participants are incentivized for their production in round 2. Finally, we investigate the interaction between social reference points and monetary incentives by implementing two treatments

- EXOBEST & PIECERATE
- ENDO & PIECERATE

We describe these 3 treatments in detail in Section 6.5.4

3.3 Questionnaire measures

3.3.1 Beliefs about own/other player's performance

The baseline participants and the main participants are asked about their beliefs regarding their own performance (and the other player's performance, if applicable) both at the beginning of Part 1) and of Part 2). More specifically:

- Just before starting Part 1), we measure participants' beliefs about their own performance by asking them "*What do you think, how many points will you be able to score in the given 5 minutes?*".
- Just before starting Part 2), each participant is asked the same question again. Moreover, participants who can compare themselves to a reference player (either endogenously picked or exogenously assigned) are asked "*What do you think, how many points will the other participant be able to score in the given 5 minutes?*"

We do not incentivize beliefs.

3.3.2 Additional measures

- At the very beginning of the experiment, we collect questionnaire data on sociodemographics (See Appendix C.1)
- At the end of Part 1) and Part 2), we ask subjects to assess how satisfied they are with their performance, how difficult they found the task and how stressed they are (See exact questions in Appendix C.2).
- Finally, at the very end of the experiment participants will be asked a set of questions which, depending on the treatment, is aimed at getting a better understanding of the effects reference players have (see Appendix C.3).
- The experiment ends with the elicitation of subjects' coefficient of loss aversion. Following Abeler et al. (2011), we look for a participant's switching point between accepting and refusing a 50-50 (hypothetical) lottery yielding a loss of USD X or a gain of USD 6, where $X = \{2, 3, 4, 5, 6, 7\}$.

4 Hypotheses

The experiment is designed to understand whether and how social reference points affect performance. First, we focus on investigating whether and how different exogenously assigned reference points affect the performance of participants. In a second step we assess whether endogenously chosen reference points have stronger motivating effects than exogenously assigned reference points. We then study how social reference points interact with monetary incentives.

4.1 Exogenous reference points

We first study *whether* and *how* different exogenously assigned social reference points affect performance. We hypothesize that the possibility to compare oneself with a salient reference player may lead to psychological effects that change performance. However, the effect of a reference player might depend on the difference between the two players' ability. Hypotheses 1a to 1c detail how different types of reference players are predicted to affect a player's performance.

Compared to RANKONLY:

Hypothesis 1a (motivating effect): *Being exposed to a reference player who is equally good or better, increases performance.*

Hypothesis 1b (discouragement effect): *The motivating effect is not monotonic in the ability difference: being exposed to a reference player who is much better reduces the motivating effect or even offsets it completely.*

Hypothesis 1c (relaxation effect): *Being exposed to a reference player who is worse does not increase performance or may even lower it.*

Hypotheses 1a-1c imply that being exposed to a *completely random* reference player (as in the EXO treatment) might – on average – not improve performance. However, being assigned to the "best reference player" (as in the EXOBEST treatment) should increase performance.

Hence, compared to RANKONLY:

Hypothesis 2a *Random assignment to reference players (EXO treatment) might not increase performance.*

Hypothesis 2b *Assignment to the "best reference player" (EXO-BEST treatment) increases performance.*

4.2 Endogenous reference points

With regard to the endogenous choice of social reference points two main points are of interest. The first one is the choice of the reference player itself. We hypothesize that

most players know what motivates them and pick their reference player in a close to optimal manner.

Hypothesis 3 *Participants tend to choose the reference player (if any) who is predicted to have the highest motivating effect for them.*

The second point concerns the effect of an endogenously chosen reference player on the player's performance. We predict that players value the choice option and feel empowered by it. As a consequence, we hypothesize that endogenously picked reference players have more powerful effects than exogenously assigned reference players.

Hypothesis 4 (empowerment effect) *The same reference player has a more positive effect on performance if the player chooses the reference player him- or herself (as compared to being exogenously assigned to the same reference player).*

Hypothesis 4 is most interesting for choices of reference players that have a positive effect on performance when exogenously assigned. Hypotheses 1a-1c and 3 imply that we do not expect that many players will choose a reference player whose performance is worse than their own. However, if such choices occur we would expect that the effect on performance is positive for those particular players or at least less negative than in the corresponding case with exogenous assignment.

Hypotheses 3 and 4 imply that performance in the ENDO treatment should be at least as high as in the EXOBEST treatment:

Hypothesis 5 *Overall, the ENDO treatment does not reduce performance compared to the EXOBEST treatment.*

4.3 Social reference points and monetary incentives

Finally, we will investigate how monetary incentives and social reference points interact. In particular, we are interested in knowing whether social reference points (either exogenously assigned or endogenously chosen) still matter when monetary incentives (a piece rate) are in place. We have no priors regarding the direction of the effect for these two hypotheses.

Hypothesis 6 *Do exogenous social reference points improve performance beyond the level attained under monetary incentives ?*

Hypothesis 7 *Do endogenous social reference points improve performance beyond the level attained under monetary incentives ?*

5 Setting[7]

5.1 Subject pool

We run the experiment on Amazon Mechanical Turk, an online platform that is widely used for economic experiments. We will recruit participants who a) currently live in the United States, b) have an approval rate of 80 percent at least and c) have at least 50 approved previous tasks.

5.2 Planned numer of observations and randomization to treatments

We plan to recruit 6500 participants in total (ideal). We want to reach a sample of

- 500 subjects per treatment for the treatments EXO-NO, EXO-LO, EXO-MID, EXO-HI, EXO-BEST, EXOBEST&PIECERATE, PIECERATE
- 1000 subjects per treatment for the treatments RANKONLY, ENDO, ENDO&PIECERATE

so as to attain sufficiently precise estimates of the productivity per treatment.[8] Participants will be randomly allocated to a treatment when they log on to the survey. The task will be kept open on Amazon Mechanical Turk until 6500 subjects have completed the study.

The experiment will be run in two waves.

- Wave 1: Collect data on 500 RANKONLY, 500 EXO-NO, 500 EXO-LO, 500 EXO-MID, 500 EXO-HI
- Wave 2: Collect data on 500 RANKONLY, 1000 ENDO, 500 EXOBEST, 500 PIECERATE, 1000 ENDO&PIECERATE, 500 EXOBEST&PIECERATE

The experiment is run in two waves in order to be able to implement the EXOBEST treatments (which can only be implemented after all the EXO treatments have been collected). The RANKONLY treatment is run twice (in wave 1 and 2) in order to have a clean comparison group in both waves.

5.3 Exclusion rules

The final sample will exclude subjects that (1) do not complete the MTurk task within 60 minutes of starting or (2) exit then re-enter the task (3) abandon the task or (4) score 2000 or more points (based on results from DellaVigna & Pope and on our baseline participants, it seems physically impossible to score more than 2000 points in 5 minutes, suggesting that such individuals are using bots).

[7] Parts of this section draw on the pre-registration by DellaVigna and Pope (2017).

[8] We double the sample size in the ENDO and the PIECERATE&ENDO treatment because different participants are likely to choose different reference players. We collect more data on RANKONLY because our data will be collected in two waves and we want a clean comparison group in both waves.

5.4 Power

In DellaVigna and Pope (2017), participants perform the same task for 10 consecutive minutes. They report an average number of clicks of 1936 and a standard deviation of 668 (for the treatment with no piece rate, the average is 1550 and the standard deviation is 720). In our experiment, task duration is set to 5 minutes. We therefore expect an average performance of about 900 and a standard deviation of approximately 350. Assuming that this is approximately the standard deviation of each treatment in the experiment and assuming a sample of 500 per treatment, we reach a power level of 80% to reject the null hypothesis of zero difference in average points between two treatments when the actual difference between the two treatments is 62 points (an effect size much smaller than what is documented in DellaVigna and Pope).

5.5 IRB

IRB approval was obtained at University of Lausanne on July 24, 2018.

6 Main analysis

6.1 Baseline balance

We will test for baseline balance for the following variables: gender and performance in round 1. We will regress each of these variables on treatment indicators to see if there are imbalances. We will account for multiple hypothesis testing by regressing the treatment indicator on all of the variables, and we will conduct a joint F-test, to see if the coefficients are jointly different from zero.

6.2 Main outcome variables

Our main dependent variable will be participants' effort at the a-b task in Part 2), i.e. effort_2. The variable effort_1 corresponds to participant's effort in Part 1). We will also estimate the effects of our treatments on the following two alternative dependent variable

- $\frac{\text{effort}_2 - \text{effort}_1}{\text{effort}_1}$
- $\text{zscored}(\text{effort}_2)$[9]

For ease of exposition, the specifications discussed in this pre-analysis plan do not include control variables. Nevertheless, each specification will also be estimated with controls. The controls that will we use are gender, age, and effort_1.[10]

6.3 Heterogeneity analysis

In this subsection, we describe the various dimensions along which heterogeneity in treatment effects will be investigated.

[9] Whenever effort_2 will be zscored, we will also zscore effort_1.

[10] We will run one specification with only effort_1 as a control, and another specification with all three variables as controls.

6.3.1 Gender

Previous work has documented that, in some cases, men perform better than women in competitive environments (Gneezy et al., 2003). In addition, experimental evidence suggests that women have a tendency to shy away from competition while men embrace it (Niederle and Vesterlund, 2007). This leads us to think that, if there are gender differences in preferences for competition, it is likely that a) social reference points affect men and women differently, and b) men and women choose different social reference points. In particular, women might benefit less from a better reference player and are therefore also less likely to choose such a reference player.

6.3.2 Ability

It can be expected that the effect of a given social reference point differs from one individual to another depending on the individual's ability at the task. For example, an individual performing very badly at the task might not feel motivated to the same extent by a slightly better reference point than an individual performing well. Therefore, we will analyze if the effects discussed above are mediated by the individual's ability (as proxied by effort_1, their performance in part 1).

6.3.3 Loss aversion

Finally, it can be expected that loss aversion around an expectation-based reference point correlates significantly with loss aversion around a social reference point. This relationship has, however, not been established. Hence, we will investigate if the treatment effects we document are stronger for people that are identified as more loss averse than the median in the Abeler et al. task, compared to those that are less loss averse.

Each equation specified below will be re-estimated to account for heterogeneity in treatment effects.[11] We prespecify the following variables for heterogeneity analysis:

- **Male**, indicated by a dummy taking the value 1 if the individual is a male.
- **Lossaverse**, indicated by a dummy variable taking the value 1 if the individual's estimated coefficient of loss aversion (obtained from the Abeler et al. task) is above the median
- **Ability**, indicated by a dummy variable taking the value 1 if the individual's ability, i.e. effort in part 2, is above the median.

[11] For example, we will re-estimate Equation 1 as follows:

$$\text{effort}_{i2} = \alpha + \delta_0 \text{interaction}_i + \delta_1 \text{EXO}_i + \delta_2 \text{EXO} \times \text{interaction}_i + \varepsilon_i$$

were interaction_i refers to one of the interaction variables prespecified. Interaction effects are identified by δ_2.

6.4 Estimating treatment effects (Part 1/2: data from wave 1)

We start with the analyzing of the data that will be collected in wave 1. In this subsection, we assess the effects of the RANKONLY and the EXO treatments.[12]

6.4.1 Assessing the average effect of exogenously assigned reference points

We test if exogenously assigned reference points improve average performance, compared to a situation in which participants only see their rank within the distribution of baseline participants (Hypothesis 2a). To that end, we test whether participants assigned to the EXO treatment perform indeed better than participants assigned to the RANK-ONLY treatment. Using participants from the EXO and the RANK-ONLY treatments, we estimate the following equation:

$$\text{effort}_{i2} = \alpha + \delta \text{EXO}_i + \varepsilon_i \tag{1}$$

where EXO_i is a dummy which is equal to one if the participant i has been assigned to the EXO treatment. The coefficient δ reveals whether being assigned an exogenous and random reference player has a motivating effect, as compared to participants who only see an intermediary "ranking information" before proceeding to part 2 without seeing any reference player.

6.4.2 Assessing the effects of different exogenously assigned reference points

In order to assess how different exogenous reference points affect performance, we will estimate the following model on the sample of participants assigned to the RANKONLY or the EXO treatments

$$\text{effort}_{i2} = \alpha + \delta_0 \text{NoRP}_i + \delta_1 \text{RP}_{1i} + \delta_2 \text{RP}_{2i} + \delta_3 \text{RP}_{3i} + \varepsilon_i \tag{2}$$

where NoRP_i is a dummy variable taking value 1 if player i has been assigned NO reference player and RP_{ji} is a dummy variable taking value 1 if player i has been assigned reference player j.[13] The constant (α) captures how participants in the RANKONLY treatment perform in round 2.

Estimating this equation for the full sample reveals the overall most effective exogenous reference player ($\max(\delta_1, \delta_2, \delta_3)$). In order to test if different reference players affect performance differently depending on own ability, we will estimate equation (2) for different ranges of ability levels. We will split up our sample in the following subsamples:

- Subsample 1: $\text{effort}_{i1} \leq \text{effort}_{RP_1 1}$
- Subsample 2: $\text{effort}_{RP_1 1} < \text{effort}_{i1} \leq \text{effort}_{RP_2 1}$
- Subsample 3: $\text{effort}_{RP_2 1} < \text{effort}_{i1} \leq \text{effort}_{RP_3 1}$

[12] I.e. for this part of the analysis we use the RANKONLY data that will be collected in wave 1.

[13] RP_1 corresponds to the worst reference player, i.e. the reference player that is ranked 49 in round 1. RP_2 corresponds to the medium reference player, i.e. the reference player that is ranked 26 in round 1. Finally, RP_3 corresponds to the best possible reference player, i.e. the baseline participant that is ranked 4 in round 1.

A graphical illustration of how the sample will be divided into subsamples (based on participants effort in part b) is provided in Figure 1 below.

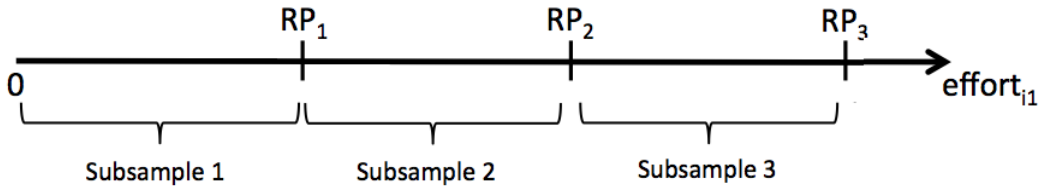


Figure 1: Partition of the EXO sample into different subsamples.

Ultimately, what we will estimate is the average response of different categories of participants (i.e. different subsamples) to seeing information on different reference players, as depicted in Figure 2. Note that this graph is just an illustration; The real response functions might take very different forms.

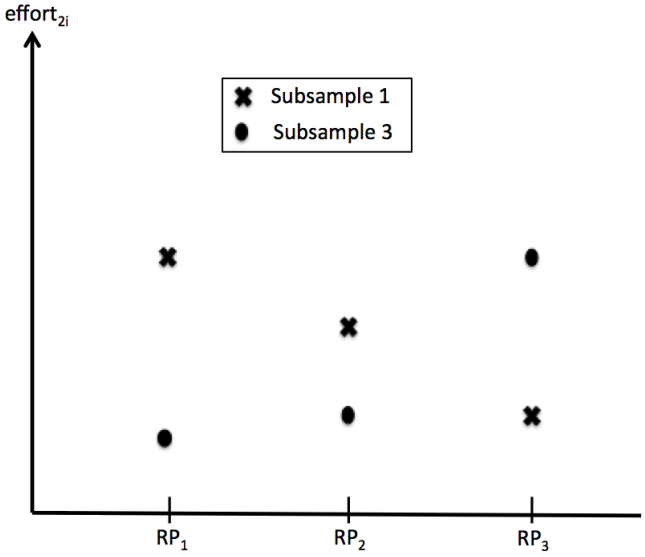


Figure 2: Example of a response function for participants in subsample 1 (crosses) and participants in subsample 3 (circles).

Tests:

- **Hypothesis 1a**
  Subsample 1: $\delta_1 > 0$, $\delta_2 > 0$, $\delta_3 \geq 0$
  Subsample 2: $\delta_2 > 0$, $\delta_3 > 0$
  Subsample 3: $\delta_3 > 0$

- **Hypothesis 1b**
  Subsample 1: $\delta_3 < \delta_2 < \delta_1$
  Subsample 2: $\delta_3 < \delta_2$

- **Hypothesis 1c**
  Subsample 2: $\delta_1 \leq 0$
  Subsample 3: $\delta_1 \leq 0$, $\delta_2 \leq 0$

6.5 Estimating treatment effects (Part 2/2: data from wave 2)

The remaining part of the analysis (including Section 6.5.4) will make use of data collected in wave 2.[14]

6.5.1 Assessing the effects of exogenously assigned "best" reference points

Using the estimated parameters in equation (1) for our subsamples allows us to predict, for each participant, which reference player would generate the largest motivating effect. More specifically, for each participant i we predict which reference player j would maximize his performance, given participant i's gender and performance in part 1). In the EXO-BEST treatment, we assign each participant to the reference player $j^{\max}$ that is predicted to maximize his performance in Part 2). That is, after having observed the effort in Part 1) of our EXO-BEST participants (and their gender), we exogenously assign them to their best reference player for Part 2).

We then test whether participants assigned to the EXO-BEST treatment perform better than participants assigned to the RANK-ONLY treatment (Hypothesis 2b). Using participants from the RANK-ONLY and the EXO-BEST treatment, we estimate the following equation:

$$\text{effort}_{i2} = \alpha + \delta \text{EXO-BEST}_i + \varepsilon_i$$

where EXO-BEST_i is a dummy which is equal to one if the participant i has been assigned to the EXO-BEST treatment. The coefficient δ reveals whether the motivating effect in the EXO-BEST treatment is larger than in the RANK-ONLY treatment.

6.5.2 Assessing the effects of endogenous reference points

We start by testing whether participants in the endogenous treatment choose their optimal reference player. We do so by comparing their actual choice with their predicted best choice player $j^{\max}$. More specifically, we test if the probability that a subject picks his optimal reference player is greater than the probability to choose it by chance (i.e. 25%).[15]

Tests:

- **Hypothesis 3**
  H_0: $\text{prob}(Y_i = RP_{j^{max}}) = 0.25$
  H_A: $\text{prob}(Y_i = RP_{j^{max}}) > 0.25$

were Y_i corresponds to participant i's chosen reference point and $RP_{j^{max}}$ corresponds to the reference player that is predicted to maximize his performance (as defined in section 6.5.1).[16]

[14] The relevant RANKONLY participants that we consider hereafter are those collected in wave 2.

[15] Given that there are 3 possible reference players and a possibility to chose "no reference player", the overall number of possibilities is 4.

[16] We will also test if the likelihood to choose ones' optimal reference player varies along these dimension of heterogeneity (male, ability, loss aversion). That is, we will test if

We then test if participants assigned to the ENDO treatment perform better than participants assigned to the RANK-ONLY treatment (Hypothesis 5). Using participants from the ENDO and the RANK-ONLY treatments, we estimate the following equation:

$$\text{effort}_{i2} = \alpha + \delta \text{ENDO}_i + \varepsilon_i$$

where ENDO_i is a dummy which is equal to one if the participant i has been assigned to the ENDO treatment. The coefficient δ reveals whether the motivating effect in the ENDO treatment is larger than in the RANK-ONLY treatment.

We test for the empowerment effect (Hypothesis 4) by analyzing whether participants assigned to the ENDO treatment perform better than participants assigned to the EXO-BEST treatment. Using participants from the ENDO and the EXO-BEST treatment, we estimate the following equation:

$$\text{effort}_{i2} = \alpha + \delta \text{ENDO}_i + \varepsilon_i$$

The coefficient δ reveals whether the motivating effect in the ENDO treatment is larger than in the EXO-BEST treatment.

6.5.3 Additional Analysis

To further understand how exactly exogenous and endogenous social reference points affect player's performance, we will carry out the following two-tiered strategy: First, we will analyze the treatment effects by relying on effort in round one as a common benchmark for comparison. We will test the effect different reference players on different deciles (quantiles) of the distribution of effort_1. To that end, we will rank all our participants (pooling all the treatments) depending on their performance in round 1, and investigate how participants within a certain decile are affected by the different treatments. This analysis will reveal how social reference points affect different parts of the ability distribution.

Second, we estimate the effect of endogenous choice on effort provision by comparing effort in round two of subjects who chose a reference player with the effort in round two of subjects who, in the exogenous treatments, state *they would have chosen the same reference player as the subjects in the endogenous treatment*[17]. This analysis provides further information on the difference between endogenous choice and exogenous assignment.

6.5.4 Monetary incentives and Social Reference Points

Piece rate only In order to calibrate the size of the different treatments described above, we plan to run an incentivized version of the RANK-ONLY treatment. Specifically, the participants in the PIECERATE treatment will receive a piece rate of 1 cent

H_0: $\text{prob}(Y_i = RP_{j^{max}} | \text{interaction}_i = 1) = \text{prob}(Y_i = RP_{j^{max}} | \text{interaction}_i = 0)$
H_A: $\text{prob}(Y_i = RP_{j^{max}} | \text{interaction}_i = 1) \neq \text{prob}(Y_i = RP_{j^{max}} | \text{interaction}_i = 0)$

for all the interaction variables prespecified in Section 6.3

[17] Participants in the EXO treatments are asked the quetsions *Could you have chosen a reference player, which one would you have chosen? [Participant ranked 4, Participant ranked 26, Participant ranked 49, None]*

per 100 points scored in round 2 (in addition to the flat fee).[18] Depending on their performance, this corresponds to a bonus of circa 10 to 20%. We assess the effects of the piece rate by estimating the following equation on the sample of participants in the RANKONLY and the PIECERATE treatments:

$$\text{effort}_{i2} = \alpha + \delta \text{PIECERATE}_i + \varepsilon_i$$

The effect size of the piece rate (δ) will be compared to the effect sizes of the different social reference point treatments using standard cross-specification tests (e.g. compared to the δ of the regression in section 6.5.1).

EXOBEST with piece rate A key question that remains open is whether exogenously assigned social reference point can improve performance beyond the performance that can be reached under monetary incentives (Hypothesis 6). In order to test this hypothesis, we run an EXOBEST treatment with incentives (piece rate of 1 cent per 100 points in round 2) and we conduct the following tests.

1. Using data from the PIECERATE and from the EXOBEST&PIECERATE treatments, we estimate:

$$\text{effort}_{i2} = \alpha + \delta \text{EXOBEST\&PIECERATE}_i + \varepsilon_i$$

where δ reveals if exogenous social reference points increase performance beyond the performance achieved under monetary incentives only (i.e. we test if $\delta > 0$).

2. Using data from the RANK-ONLY, the EXOBEST, the PIECERATE and the EXOBEST&PIECERATE treatments, we estimate

$$\text{effort}_{i2} = \alpha + \delta_1 \text{EXOBEST}_i + \delta_2 \text{PIECERATE}_i + \delta_3 \text{EXOBEST\&PIECERATE}_i + \varepsilon_i$$

where the variables indicate treatment dummies. Again, we test if an exogenously assigned reference player increases performance beyond the motivational effects of monetary incentives (i.e. we test if $\delta_3 > \delta_2$).[19]

ENDO with piece rate Finally, we investigate whether endogenously chosen social reference point can improve performance beyond the performance that can be reached under monetary incentives (Hypothesis 7). In order to test this hypothesis, we run an ENDO treatment with incentives (piece rate of 1 cent per 100 points in round 2) and we conduct the following tests.

[18] Note that there is still no piece rate in round 1.

[19] Alternatively, we might estimate an (equivalent) specification in which treatment dummies are replaced with indicators of treatment characteristic:

$$\text{effort}_{i2} = \alpha + \delta_1 \text{EXOBEST}_i + \delta_2 \text{PIECERATE}_i + \delta_3 (\text{EXOBEST x PIECERATE})_i + \varepsilon_i$$

where EXOBEST is a dummy for being assigned an EXOBEST reference player, PIECERATE is a dummy for receiving a piecerate, and (EXOBEST x PIECERATE) is an interatction variable. Under this specification, PIECERATE x EXOBEST improves performance beyond PIECERATE if $\delta_1 + \delta_3 > 0$.

1. Using data from the PIECERATE and from the ENDO&PIECERATE treatments, we estimate:

$$\text{effort}_{i2} = \alpha + \delta \text{ENDO\&PIECERATE}_i + \varepsilon_i$$

where δ if endogenous social reference points increase performance beyond the performance achieved under monetary incentives only (i.e. we test if $\delta > 0$).

2. Using data from the RANK-ONLY, the ENDO, the PIECERATE and the ENDO&PIECERATE treatments, we estimate[20]

$$\text{effort}_{i2} = \alpha + \delta_1 \text{ENDO}_i + \delta_2 \text{PIECERATE}_i + \delta_3 \text{ENDO\&PIECERATE}_i + \varepsilon_i$$

where the variables indicate treatment dummies. Again, we test if endogenously chosen reference players increases performance beyond the motivational effects of monetary incentives (i.e. we test if $\delta_3 > \delta_2$).

Finally, we will also analyze whether the presence of incentives in the ENDO treatment affects the way participants choose their reference player (as described in section 6.5.2).

7 Secondary analysis

7.1 Secondary outcome variable

We will also investigate whether the different treatments affect the probability of improving performance. To that end, we will replace the dependend variable in the analysis described above by a dummy variable taking value 1 if performance improves between period 1 and period 2, and 0 otherwise.

7.2 Additional specifications and tests

In this section, we describe some exploratory analysis that we will undertake. This part of the analysis is not motivated by any particular model. We do not have explicit hypotheses to formulate, nor do we have particular models to prespecify. However, our intuition suggests that interesting things might be discovered.

7.2.1 Choice of the reference point

In order to investigate if participants' chosen reference point depends on their performance in part 1) (or other individual characteristics), we will estimate a multinomial probit model on the sample of participants assigned to the endogenous treatment. More specifically, we will estimate the probability to choose a particular reference point $\Pr(Y_i{=}j|\mathbf{X})$, were Y_i corresponds to individual i's chosen reference point, $j = \{1, 2, 3\}$ is a categorical variable indicating which reference player j is chosen and $\mathbf{X}$ is a vector including gender and participant's performance in part 1).

Importantly, we will also conduct this analysis to determine whether men and women choose their reference point differently (see section 6.3.1 for a motivation).

[20] Same comments as in the previous footnote.

7.2.2 Real-time effect of reference players

Given that our data records the score of each participant "second-by-second", we will also explore how the score of one participant responds to the evolution of the score of his reference player. An graphical example of this kind of analysis is depicted in Figure 3.

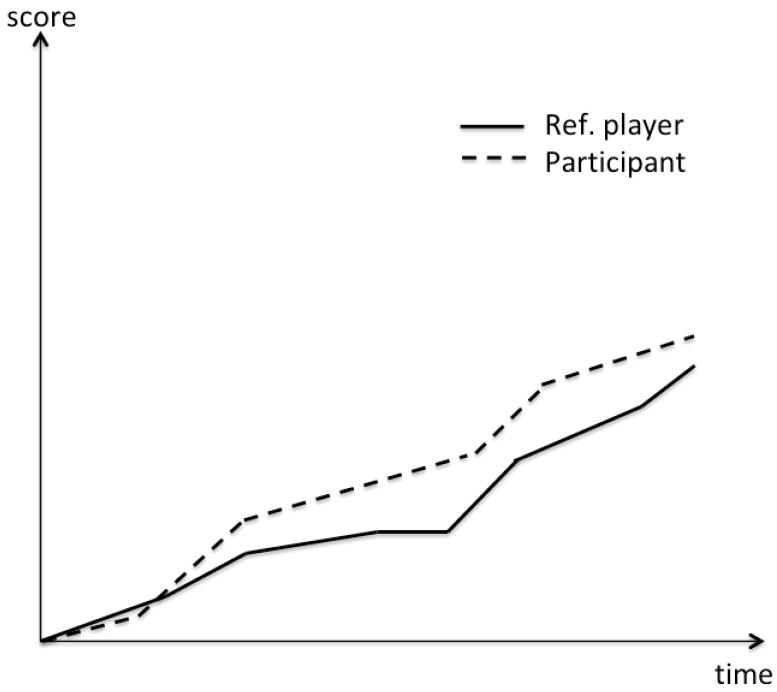


Figure 3: Example of evolution of the score of a participant (dashed line) and the score of its reference player (plain line).

7.2.3 Beliefs

The beliefs about the performance of the reference player will be used to assess the anticipated effect of a reference player.

$$\text{belief}(\text{effort}_{RPj})_i = \alpha + \delta_1 \text{RP}_{2i} + \delta_2 \text{RP}_{3i} + \delta_3 \text{RP}_{4i} + \varepsilon_i$$

were $\text{belief}(\text{effort}_{RPj})_i$ is player i's beliefs about the performance in part c) of his reference player j.

This specification allows to estimate if different reference players generate different expected performances. Note that the constant captures the anticipated effort of RP_1.

Importantly, we will check if exogenously assigned reference players generate different expectations that endogenously chosen ones. To do so, we will estimate this specification separately for the EXO-BEST and for the ENDO treatments, and perform simple cross specification Wald tests.

References

Abeler, J., A. Falk, L. Götte, and D. Huffman, "Reference points and effort provision," *The American Economic Review*, 2011, *101* (2), 470–492.

DellaVigna, Stefano and Devin Pope, "What motivates effort? Evidence and expert forecasts," *The Review of Economic Studies*, 2017, *85* (2), 1029–1069.

Gneezy, Uri, Muriel Niederle, Aldo Rustichini et al., "Performance in competitive environments: Gender differences," *The Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

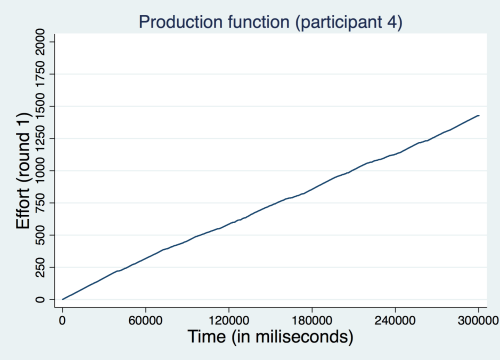
Niederle, Muriel and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Appendix A : Baseline participants

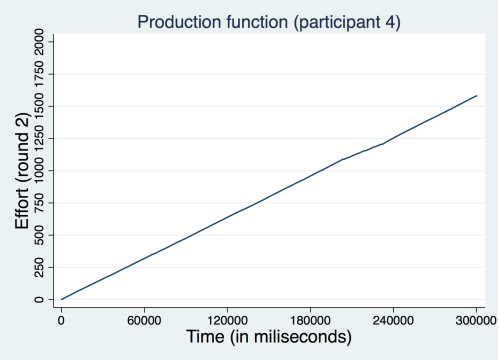
Table A1 depicts the ranking of the 60 baseline participants, sorted by their performance in round 1 (effort1). The three potential reference players are highlighted in yellow. **All our main participants will always be assigned to (or have the possibility to choose) one of these 3 reference players, unless they are assigned no reference player.** Figures A2 to A4 depict the production function in part 1 (panel a) and in part 2 (panel b) of each of these 3 potential reference players.

rank	subjectid	subjectnr	effort1	effort2
1	senn-ab-task-0000252	36	1553	1128
2	senn-ab-task-0000235	25	1488	1474
3	senn-ab-task-0000208	6	1446	1458
4	senn-ab-task-0000205	4	1428	1580
5	senn-ab-task-0000218	13	1415	1048
6	senn-ab-task-0000257	39	1409	1426
7	senn-ab-task-0000222	15	1366	544
8	senn-ab-task-0000226	19	1338	519
9	senn-ab-task-0000238	27	1325	1231
10	senn-ab-task-0000223	16	1307	1338
11	senn-ab-task-0000263	42	1301	1016
12	senn-ab-task-0000225	18	1299	1300
13	senn-ab-task-0000290	55	1284	1244
14	senn-ab-task-0000232	23	1259	861
15	senn-ab-task-0000228	20	1249	1226
16	senn-ab-task-0000203	3	1238	1081
17	senn-ab-task-0000278	51	1231	1326
18	senn-ab-task-0000288	54	1198	1310
19	senn-ab-task-0000272	47	1189	1109
20	senn-ab-task-0000210	8	1189	1133
21	senn-ab-task-0000255	38	1149	1258
22	senn-ab-task-0000229	21	1119	1297
23	senn-ab-task-0000292	56	1111	1402
24	senn-ab-task-0000273	48	1105	257
25	senn-ab-task-0000265	43	1077	1032
26	senn-ab-task-0000270	46	1073	1195
27	senn-ab-task-0000268	45	1062	1254
28	senn-ab-task-0000230	22	984	1139
29	senn-ab-task-0000219	14	968	1095
30	senn-ab-task-0000212	9	951	1126
31	senn-ab-task-0000243	30	950	891
32	senn-ab-task-0000202	2	929	1339
33	senn-ab-task-0000237	26	917	982
34	senn-ab-task-0000244	31	914	1058
35	senn-ab-task-0000209	7	897	1012
36	senn-ab-task-0000217	12	893	861
37	senn-ab-task-0000216	11	851	795
38	senn-ab-task-0000285	53	826	822
39	senn-ab-task-0000298	60	820	1069
40	senn-ab-task-0000253	37	809	1261
41	senn-ab-task-0000187	1	805	825
42	senn-ab-task-0000247	33	798	875
43	senn-ab-task-0000250	35	797	1246
44	senn-ab-task-0000297	59	778	888
45	senn-ab-task-0000293	57	739	853
46	senn-ab-task-0000277	50	707	900
47	senn-ab-task-0000248	34	694	714
48	senn-ab-task-0000295	58	589	528
49	senn-ab-task-0000241	29	584	678
50	senn-ab-task-0000262	41	337	171
51	senn-ab-task-0000240	28	336	333
52	senn-ab-task-0000234	24	250	179
53	senn-ab-task-0000282	52	229	302
54	senn-ab-task-0000224	17	205	174
55	senn-ab-task-0000214	10	139	111
56	senn-ab-task-0000275	49	118	126
57	senn-ab-task-0000246	32	101	0
58	senn-ab-task-0000267	44	2	995
59	senn-ab-task-0000260	40	0	812
60	senn-ab-task-0000206	5	0	944

Figure A1

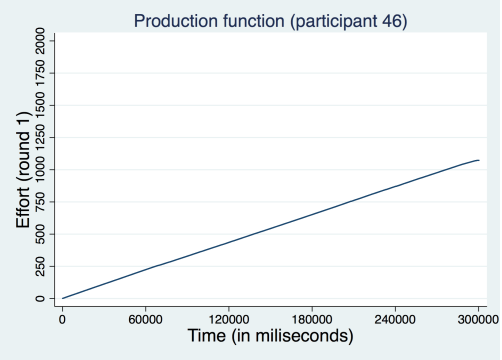


(a)

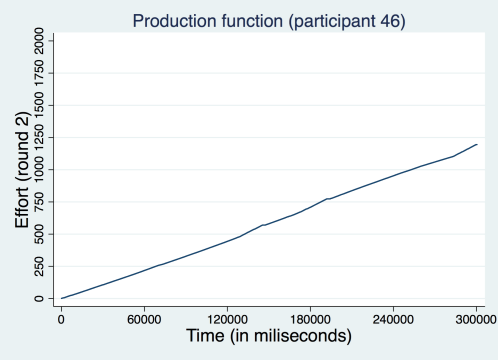


(b)

Figure A2: Top reference player (subjectnr=4, rank=4)

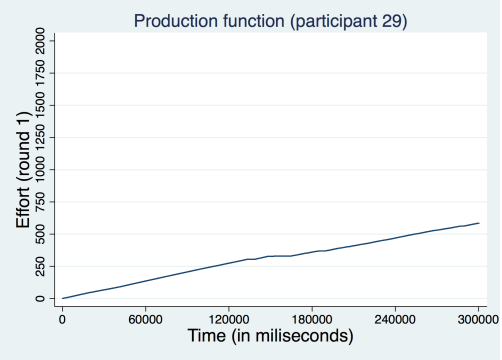


(a)

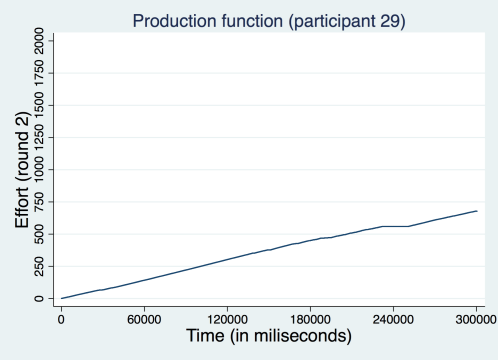


(b)

Figure A3: Mid reference player (subjectnr=46, rank=26)



(a)



(b)

Figure A4: Low reference player (subjectnr=29, rank=49)

Appendix B : Screenshots

Figure B1: Screenshot of round 1 (and round 2, if player does not see a reference player).

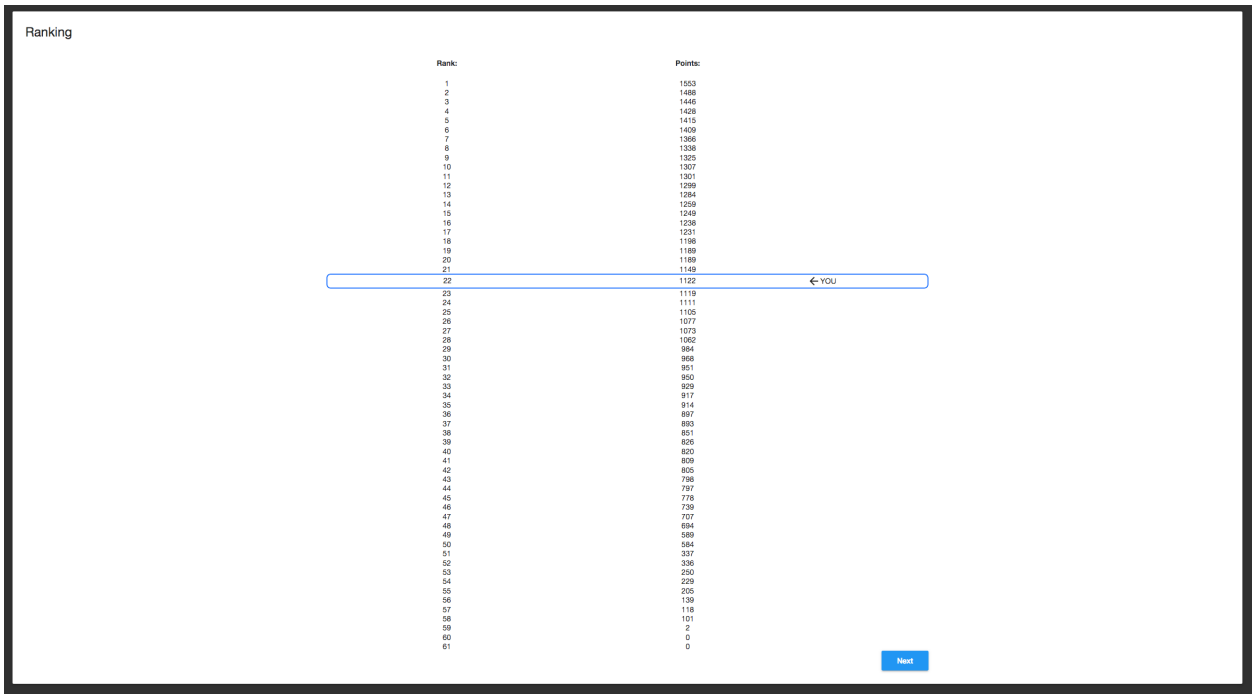


Figure B2: RANK-ONLY information screen. The rank and score of all the baseline participants is depicted on the right side of the screen. The participant's rank and score are highlighted (blue box). This ranking is also displayed to participants in the EXO, EXO-BEST and ENDO treatments.

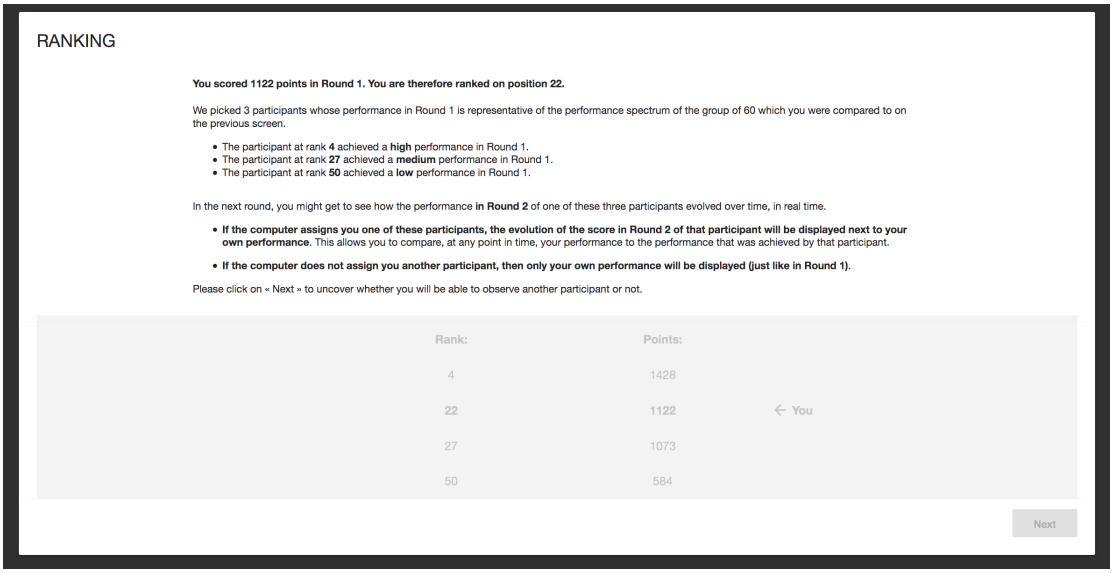


Figure B3: EXO (and EXO-BEST). After getting information about the distribution of performance of the baseline players (Figure B2), partcipants in the EXO treatments learn about the 3 potential reference players.

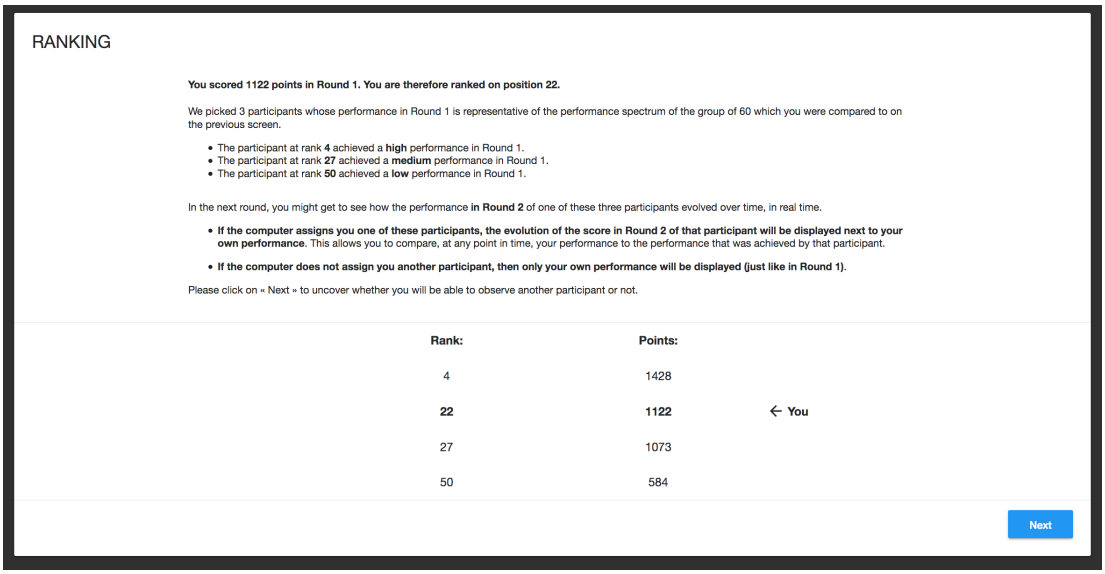


Figure B4: EXO (and EXO-BEST).

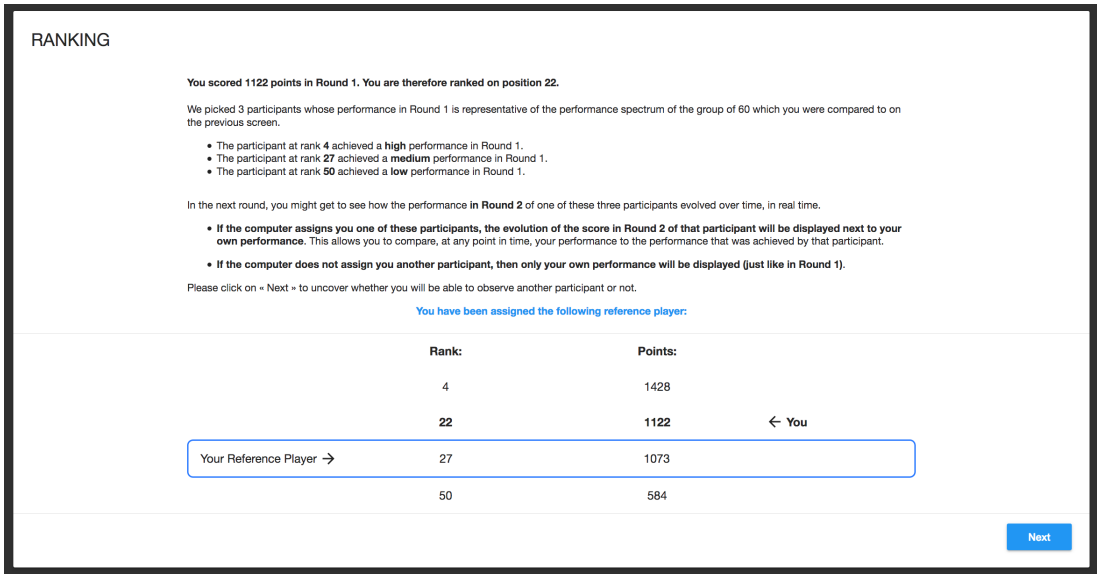


Figure B5: EXO (and EXO-BEST). They are then randomly assigned to one of the reference players.

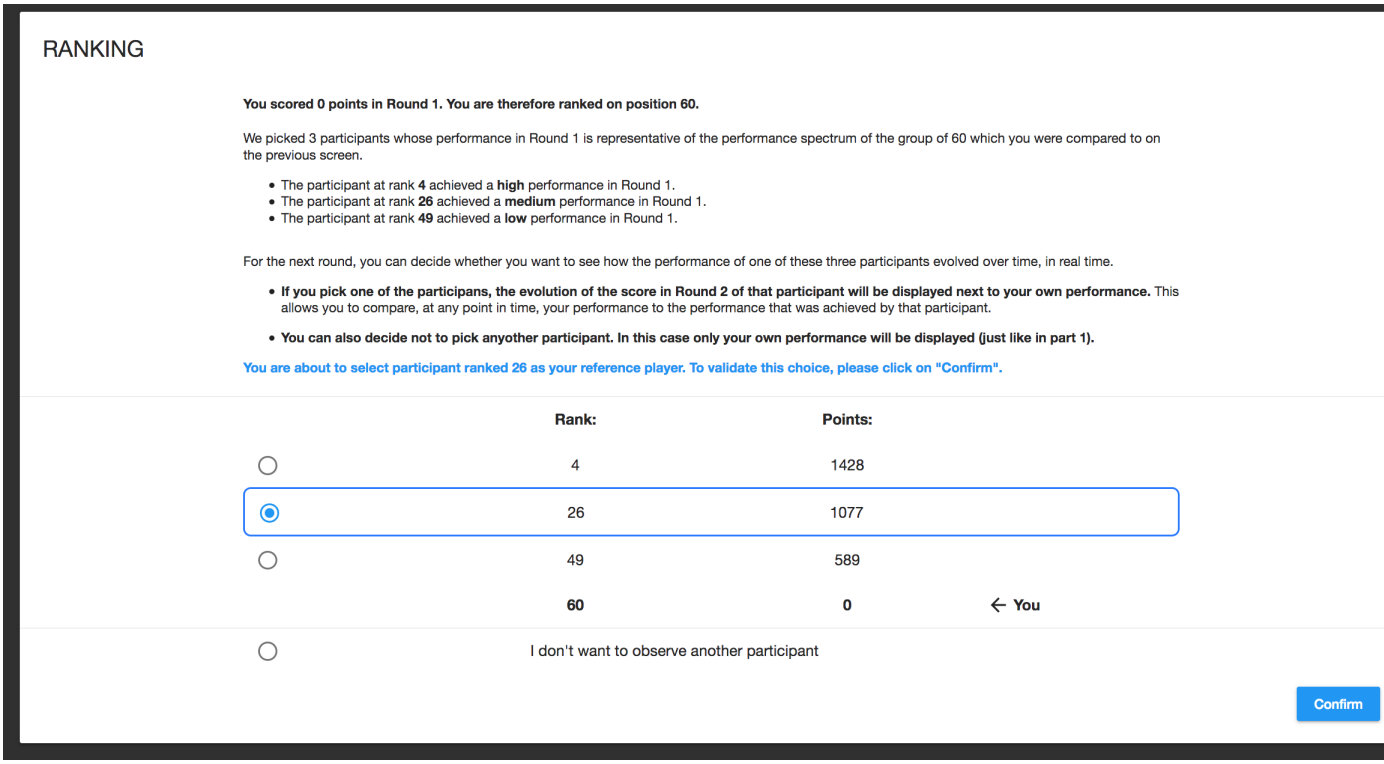


Figure B6: ENDO (choice screen). After getting information about the distribution of performance of the baseline players (Figure B2), partcipants in the ENDO treatments learn about the 3 potential reference players, and can choose which reference player (if any) to observe.

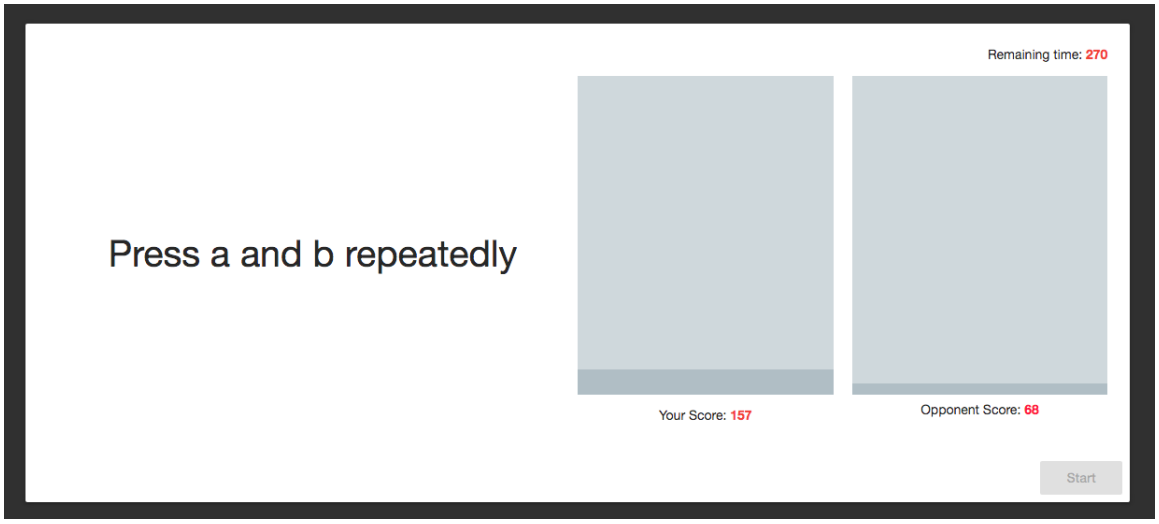


Figure B7: Screenshot for round 2 when participant sees a reference player. In this example, the participant's score in part 2 at this precise time (i.e. when 270 seconds are remaining) is 157 whereas the score of the reference player at this exact moment was 68.

Appendix C : Questionnaire measures

C.1 : Socio-demographics

- What is your gender? [male/female]
- In which year were you born? [1900-2010]
- What is your monthly gross income? [brackets]
- Which of the following best describes your race or ethnicity? [Caucasian / White, African American / Black, Hispanic/Latino, Asian American / Asian, Native American, Other]
- What category best describes your highest level of education? [8th grade or less, some high school, high school degree / GED, Some college, 2-year College Degree 4-year College Degree, Master's Degree, Doctoral Degree, Other]
- In which state do you currently reside? [list of states]
- Many people in the USA lean towards a political party. Which party do you lean towards? [Democrats, Republicans, Other, None]

C.2 : Post-effort questions

After both Parts 1) and 2), we ask

- On a scale from 1-5, how difficult did you find the task? [1 Not at all difficult - 5. Very difficult]
- On a scale from 1-5, how stressed have you been while completing the task? [1 Not at all stressed - 5. Very stressed]
- How satisfied are you with your performance? [1. Not at all satisfied - 5.very satisfied]

C.3 : Exit survey

To all players who get to see a reference player, we ask:

- Please describe in a few sentences how the performance of the other player affected your performance (open-ended).
- On a scale from -5 to +5, how did observing the performance of the other player affect your performance? [-5. Negatively affected my perf. , 0. Did not affect my perf. , +5. Positively affected my perf.]
- On a scale from -5 to +5, did observing the performance of the other player motivate you or discourage you? [-5. Discouraged me a lot, 0. Did not affect me, 5. Motivated me a lot]

- On a scale from 1 to 5, did observing the performance of the other player make you nervous? [1. Not at all nervous - 5. very nervous]
- On a scale from 1 to 5, to what degree did you feel in competition with the other player did you feel? [1. No competition at all - 5. very high competition]
- On a scale from 1 to 5, did observing the performance of the other player make the task more enjoyable for you? [1. Not at all more enjoyable - 5. Much more enjoyable]

In addition, we ask a set of "counterfactual questions" to assess how people think they would have performed, had they been assigned a different reference player. In the EXO (and EXO-BEST) treatments, for example, we ask :

- In the previous round, you observed the performance of the reference player who ranked 4th. Imagine that, instead of observing the reference player who ranked 4th, you had been assigned the reference player who was ranked 26. How would this have affected you? [A. It would have increased my performance, compared to the performance I achieved while observing the reference player ranked 4th. B. It would have decreased my performance, compared to the performance I achieved while observing the reference player ranked 4th. C. It would have made no difference.]
- Imagine that, instead of observing the reference player who ranked 4th, you had been assigned the reference player who was ranked 49. How would this have affected you? [A. It would have increased my performance, compared to the performance I achieved while observing the reference player ranked 4th. B. It would have decreased my performance, compared to the performance I achieved while observing the reference player ranked 4th. C. It would have made no difference.]
- Finally, imagine that instead of observing the reference player who ranked 3rd, you had been assigned NO reference player. How would this have affected you? [A. It would have increased my performance, compared to the performance I achieved while observing the reference player ranked 4th; B. It would have decreased my performance, compared to the performance I achieved while observing the reference player ranked 4th; C. It would have made no difference.]
- Could you have chosen a reference player, which reference player would you have chosen? [Participant ranked 4, participant ranked 26, participant ranked 49, None]

In the ENDO treatment, we ask

- In the previous round, you observed the performance of the reference player who ranked 4th. Please indicate in a few sentences why you have chosen to observe the performance of this reference player. (Open answer)
- Please describe in a few sentences how the performance of the other player affected your performance. (Open answer)
- On a scale from 1-5, do you regret to have chosen this reference player? [1. Not regrets at all, 5. A lot of regrets]

Finally, in the EXO-NO RP we ask the followin counterfactual questions:

- In the previous round, you could not observe the performance of a reference player. Imagine that you had been assigned the reference player who was ranked 4th. How would this have affected you? [A. It would have increased my performance, compared to not observing a reference player. B. It would have decreased my performance, compared to not observing a reference player. C. It would have made no difference.]
- Imagine that you had been assigned the reference player who was ranked 26th. How would this have affected you? [A. It would have increased my performance, compared to not observing a reference player. [B. It would have decreased my performance, compared to not observing a reference player. C. It would have made no difference.]
- Finally, imagine that you had been assigned the reference player who was ranked 59. How would this have affected you? [A. It would have increased my performance, compared to not observing a reference player. B. It would have decreased my performance, compared to not observing a reference player. C. It would have made no difference.]

while in the ENDO treatment, if a subject decided to see no reference player we ask:

- In the previous round, you decided not to observe a reference player. Please indicate in a few sentences why you made this choice.(open answer)
- On a scale from 1-5, do you regret to have chosen not to observe a reference player? [1. Not regrets at all, 5. A lot of regrets]