DISAPPOINTMENT AVERSION AND SOCIAL COMPARISONS IN A REAL-EFFORT COMPETITION

SIMON GÄCHTER, LINGBO HUANG and MARTIN SEFTON*

We present an experiment to investigate the source of disappointment aversion in a sequential real-effort competition. Specifically, we study the contribution of social comparison effects to the disappointment aversion previously identified in a two-person real-effort competition (Gill, D., and V. Prowse. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." American Economic Review, *102, 2012, 469–503). To do this we compare "social" and "asocial" versions of the Gill and Prowse experiment, where the latter treatment removes the scope for social comparisons. If disappointment aversion simply reflects an asymmetric evaluation of losses and gains we would expect it to survive in our asocial treatment. Alternatively, if losing to or winning against another person affects the evaluation of losses/gains, as we show would be theoretically the case under asymmetric inequality aversion, we would expect treatment differences. We find behavior in social and asocial treatments to be similar, suggesting that social comparisons have little impact in this setting. Unlike in Gill and Prowse we do not find evidence of disappointment aversion.* (*JEL* C91, D12, D81, D84)

I. INTRODUCTION

An important research program in behavioral economics has been the development of theories of reference-dependent preferences according to which people are loss averse—weighing losses more heavily than gains—around an expectations-based reference point. Such theories (e.g., Bell 1985; Delquié and Cillo 2006; Köszegi and Rabin 2006; Loomes and Sugden 1986) were originally developed and tested in nonstrategic settings (e.g., individual lottery choice experiments), but they have obvious relevance to contests, which quite naturally result in winners and losers, gains and losses.[1] In this study, we examine the behavioral consequences of competing against Nature or against another person. The reason this may matter is that competing against a person invites social comparisons that are not relevant in games against Nature.

Our framework for studying social comparison effects in a contest is the two-person sequential real effort competition studied by Gill and Prowse (2012) (hereafter GP). In their model, they show that a disappointment averse second mover (i.e., a second mover who is loss averse around an expectations-based reference point) responds negatively to a first mover's effort, and in their experiment they find significant evidence for this discouragement effect.

*We thank David Gill and Victoria Prowse for sharing the software, instructions, and data from their experiment, and for helpful discussions. We are also grateful for comments from the co-editors, Dave Malueg and Subhasish Chowdhury, two anonymous referees, Robin Cubitt, and participants at several conferences and seminars.

The copyright line for this article was changed on 29 November 2017 after original online publication.

Gächter: Professor, CESifo, IZA, and School of Economics, University of Nottingham, Nottingham, NG7 2RD, UK. Phone +44 (0) 115 84 66132, Fax + 44 (0) 115 95 14159, E-mail simon.gaechter@nottingham.ac.uk

Huang: Postdoctoral Researcher, Department of Economics, Monash University, Melbourne, Victoria 3800, Australia. Phone +61 399 055 482, E-mail lingbo.huang@outlook.com

Sefton: Professor, School of Economics, University of Nottingham, Nottingham, NG7 2RD, UK. Phone +44 (0) 115 84 66130, Fax +44 (0) 115 95 14159, E-mail martin.sefton@nottingham.ac.uk

1. See Gill and Stone (2010) for an analysis of endogenous expectations-based reference points to a competitive setting.

ABBREVIATIONS

GP: Gill and Prowse (2012)
MSM: Method of Simulated Moments

Economic Inquiry
(ISSN 0095-2583)
Vol. 56, No. 3, July 2018, 1512–1525
doi:10.1111/ecin.12498
Online Early publication September 21, 2017

On the basis of existing evidence on the importance of social comparisons and by formally modeling inequality aversion, we hypothesize that disappointment aversion is likely to be different when competing against another person compared to competing against Nature. To test this hypothesis, we conduct an experiment with two treatments. The SOCIAL treatment features a game between two human subjects playing in the roles of first and second mover, exactly as in the study by GP. The ASOCIAL treatment removes the scope for social comparisons by removing the first mover. Instead the second mover plays against Nature in a decision problem that, absent social comparison effects, corresponds to the second mover's decision problem in the SOCIAL treatment. Thus, if social comparisons do not affect behavior we expect behavior in the two treatments to be similar, whereas social comparison effects may lead to differences.

Our main result is that, contrary to our hypothesis, behavior is very similar across our treatments. This is the case whether we compare average efforts or examine the data at a more disaggregated level using regression analysis. Two other unexpected findings of our study are that we find much weaker prize effects than GP and, unlike them, we find no evidence for disappointment aversion.

The remainder of our paper is organized as follows. Section II introduces the GP framework. Section III discusses the literature that motivated our main hypothesis. Section IV presents the design of our study, and Section V the results. Section VI presents a discussion and conclusions.

II. DISAPPOINTMENT AVERSION IN THE GP FRAMEWORK

GP applied an expectations-based version of a disappointment aversion model to a real-effort competition. In their setting, a *first mover* and a *second mover* compete for a single prize by sequentially exerting efforts, with each player's chance of winning the prize being a probabilistic function of both efforts. Specifically, the second mover wins a money prize, v, with probability

$$P = (e_2 - e_1 + \gamma)/(2\gamma)$$

where e_1 and e_2 represent first mover and second mover effort, respectively, and $\gamma = 50$ in their (and our) experiment.

In GP's model, the second mover's utility is separable in the utility derived from monetary earnings and the disutility associated with effort. In their linearized model, the second mover's utility from monetary earnings in the event that she wins the prize is assumed to be $v + g_2(v - vP)$, where v is the "material utility" of the prize, g_2 is a preference parameter, and $g_2(v - vP)$ is the "gain-loss utility" associated with earning more than expected. Utility from monetary earnings in the event that the second mover fails to win the prize is given by $0 + l_2(0 - vP)$, where 0 is the material utility, l_2 is a preference parameter, and $l_2(0 - vP)$ is the gain-loss utility associated with earning less than expected. A second mover is defined as being disappointment averse if $\lambda_2 \equiv l_2 - g_2$ is strictly positive, that is, if she is loss averse around her expected monetary payoff. Letting $C_2(e_2)$ denote the second mover's effort cost, the second mover's expected utility is

$$(1)\quad EU_2\left(e_1, e_2\right) = P\left(v + g_2\left(v - vP\right)\right) + (1 - P)\left(0 + l_2\left(0 - vP\right)\right) - C_2\left(e_2\right) = vP + \lambda_2 vP(1 - P) - C_2\left(e_2\right).$$

GP assume that the second mover maximizes Equation (1), taking e_1 as given. They show that if $\lambda_2 = 0$ the optimal effort, e_2^*, is independent of e_1, but if the second mover is disappointment averse, $\lambda_2 > 0$, then e_2^* is always (weakly) decreasing in e_1. Assuming a quadratic effort cost function,

$$C_2\left(e_2\right) = be_2 + ce_2^2/2,$$

and a strictly concave objective function, GP show that (Proposition 3, p. 480):

$$\mathrm{d}e_2^*/\mathrm{d}e_1 = -\lambda_2 v/(2\gamma^2 c - \lambda_2 v) < 0.$$

Thus, there is a discouragement effect that becomes stronger if the second mover is more disappointment averse or if the prize is higher.

The GP experiment, which we replicate and describe in more detail below, uses a real effort task in which subjects position sliders on a screen. Effort is measured by the number of correctly positioned sliders. GP find significant evidence for a discouragement effect (i.e., e_2^* is decreasing in e_1), which is more pronounced the higher the prize. Using structural estimation to estimate the distribution of λ_2, they find significant heterogeneity across individuals and a significantly positive mean of λ_2.

The starting point for our study is based on the observation that the GP model is silent about the role of the first mover as a source of disappointment aversion. Suppose that the material utility of winning simply depends on the amount won, and the asymmetric weighting of gains and losses is akin to the loss aversion exhibited in many studies of individual decision making under uncertainty, where the gain-loss utility reflects the elation/disappointment of getting more/less money than expected. Under this interpretation it would not matter whether the second mover chooses e_2 to maximize Equation (1) after observing the effort choice of a first mover, or whether there is no first mover and the second mover chooses e_2 to maximize Equation (1) where e_1 is an exogenously given parameter of the probability of success function set by the experimenter.

However, we hypothesize that these two problems are behaviorally quite different. In the first case, the second mover is involved in a game against another person, while in the second case the second mover plays a game against Nature. As we argue in Section III, there is substantial evidence on social comparison effects that motivate our hypothesis.

III. THE IMPORTANCE OF SOCIAL COMPARISONS FOR EFFORT CHOICE

Our hypothesis is that the presence of rivals and the social comparisons they afford affect competitive behavior and hence also measure disappointment aversion. This hypothesis is based on a substantial literature in economics and psychology. Starting with the psychology literature, studies of social facilitation (Zajonc 1965) have long investigated the importance of the awareness of being evaluated by others and its influence on performance (e.g., Blascovich et al. 1999; Markus 1978). In economics, evidence is accumulating that social comparisons matter for effort choice (see, e.g., Falk and Ichino 2006; Mas and Moretti 2009; Gächter, Nosenzo, and Sefton 2013; Thöni and Gächter 2015; Herbst and Mas 2015; Gill et al. 2016). Observing others also influences risk-taking behaviors (e.g., Bougheas, Nieboer, and Sefton 2013; Cooper and Rege 2011; Dijk, Holmen, and Kirchler 2014; Fafchamps, Kebede, and Zizzo 2015; Linde and Sonnemans 2012; Schwerter 2016).

Further support for the potential influence of social comparisons in competitive settings comes from the study by Herrmann and Orzen (2008). They study the role of social preferences in a sequential rent-seeking contest. They compare a second-mover's response to the first-mover's investment when the first mover is a human subject and when the first mover's investment is a random number chosen by the computer. Herrmann and Orzen find second movers make higher investments when they play against another person compared with when they play against the computer. Similarly, Eisenkopf and Teyssier (2013) examine a simultaneous move game and find that average investment into a contest is higher in the presence of rivals.

There are other channels that can also lead to social comparison effects. More generally, the presence of the opponent might be behaviorally important because it might arouse emotions, such as *social* disappointment/elation from payoff comparisons and context-dependent joy of winning (e.g., Dohmen et al. 2011), in addition to the pecuniary reward.

To give an example of how social comparisons may affect behavior in the context of GP's experiment, consider the effect of asymmetric inequality aversion. Suppose that in a game against Nature the "material utility" from monetary earnings is $u_2(y_2) = y_2$, but in a game against another person the "material utility" from monetary payoffs depends not only on own payoff but also on the other person's payoff. Specifically, suppose the second mover is inequality averse as in the study by Fehr and Schmidt (1999), so that "material utility" from monetary payoffs is $u_2(y_2) = y_2 - \alpha_2 \max\{y_1 - y_2, 0\} - \beta_2 \max\{y_2 - y_1, 0\}$, where α_2 and β_2 $(\alpha_2 \geq \beta_2, \beta_2 < 1)$ are preference parameters measuring the second mover's marginal disutility from disadvantageous and advantageous inequality, respectively.[2]

Now, in the two-person real effort competition, the monetary payoff to the second mover is $y_2 \in \{0, v\}$, and so the effective prize spread, that is, the difference between the material utility of winning the prize and not winning the prize, is simply v in the game against Nature and $v(1 + \alpha_2 - \beta_2)$ in the game against another person. The upshot is that if the second mover is asymmetrically inequality averse (i.e., $\alpha_2 > \beta_2$) then the effective prize spread in a contest against another person is greater than that in a contest against Nature. Thus, when winning in a

2. Note, the terminology is different from that usually used in discussion of the Fehr-Schmidt model. Usually, material utility would refer to the utility from own pecuniary earnings, and in addition the agent gets disutility stemming from inequality. Here "material utility" includes inequality aversion.

competition means earning more than another person and losing means earning less, asymmetric inequality aversion leads to higher stakes, and in turn the higher stakes lead to a stronger discouragement effect.

Formally, with this specification of material utility the expression for expected utility (Equation (1)) changes to become

$$(2)\quad EU_2\left(e_1,e_2\right)=-\alpha_2 v+v\left(1+\alpha_2-\beta_2\right)P + \lambda_2 v\left(1+\alpha_2-\beta_2\right)P(1-P)-C_2\left(e_2\right),$$

and with the quadratic cost specification the discouragement effect becomes

$$de_2^*/de_1=-\lambda_2 v\left(1+\alpha_2-\beta_2\right) /\left(2\gamma^2 c-\lambda_2 v\left(1+\alpha_2-\beta_2\right)\right)<0.$$

Thus, with asymmetric inequality aversion, the discouragement effect is predicted to be stronger when facing a human opponent, and the effect is stronger for second movers with higher values of $(1+\alpha_2-\beta_2)$.

In summary, there are theoretical and empirical arguments why decisions taken in a social environment (a contest against another person) might differ from decisions taken in an asocial but otherwise identical environment (a contest against Nature). In Section IV, we describe how our experiment is designed to test this hypothesis.

IV. EXPERIMENTAL DESIGN AND PROCEDURES

A. Design

Our design compares two treatments, SOCIAL and ASOCIAL, in three waves of sessions. In each wave, there were six SOCIAL and three ASOCIAL sessions, with 20 subjects participating in each session. Thus, in each wave 120 subjects participated in our SOCIAL treatment (60 first movers and 60 second movers, as in GP) and 60 subjects participated in our ASOCIAL treatment. Across all three waves, 540 subjects participated in our experiment.

In sessions using our SOCIAL treatment, ten subjects were designated as first movers and another ten as second movers. Roles were randomly determined and remained the same for the whole duration of the session. Each participant then took part in two practice rounds and ten paying rounds. A round consisted of a sequential two-player game between a first mover and a second mover. First movers were repaired with second movers at the end of each round following a "no contagion" matching so that no subject's behavior in a given round can directly or indirectly affect the behavior of other participants that the subject is paired with at a later round. In practice rounds, each participant was paired with an automaton so that these experiences would not contaminate the matching protocol in the paying rounds.

At the beginning of the game a monetary prize drawn from {£0.10, £0.20, …, £3.90} is announced to the players. The first mover then has 120 seconds to position sliders on a computer screen (see GP for a more detailed discussion of the slider task). The number of correctly positioned sliders is the first mover's points score, and is denoted e_1. The second mover is then informed of the value of e_1 and has 120 seconds to position sliders. The number of sliders correctly positioned by the second mover is denoted e_2. At the end of the game, one of the players wins the prize and the other player gets nothing. The second mover wins with probability $(e_2-e_1+50)/100$. At the end of each round, each participant learned her own and her pair member's points score, her probability of winning the prize, and whether she was the winner or loser in that round.

The ASOCIAL treatment removes the scope for interpersonal comparisons by converting this two-player game into an individual decision-making task with as few changes as possible. At the beginning of the game, a prize and a "given number," n, is announced. The player then has 120 seconds to position sliders. For convenience, we refer to this player as the second mover and her number of correctly positioned sliders as e_2 (though of course, there is no first mover). At the end of the game, the second mover wins the prize with probability $(e_2-n+50)/100$. We used the values of e_1 and the realized prize values from the SOCIAL treatment to provide the given numbers and prizes in the ASOCIAL treatment.

To see the difference between treatments as experimental participants saw it we reproduce the slider screen heading seen by second movers in Figure 1. Panel A presents the information displayed in the SOCIAL treatment and panel B shows the information displayed to the subjects (who acted as if they were second movers) in the ASOCIAL treatment. The key differential information has been *italicized* (but not in the experiment).

FIGURE 1

Key Differential Information for Second Movers in the Two Treatments: (A) SOCIAL and (B) ASOCIAL

Key Differential Information for Second Movers in the Two Treatments

A SOCIAL	B ASOCIAL
The prize in pounds for this round is: …	The prize in pounds for this round is:…
The *first mover's points score* was: …	The *given number* for this round is:…
Currently, your points score is: …	Currently, your points score is: …

Unlike in the SOCIAL treatment where the second movers knew that the first movers were real human participants who were participating in the same session, the participants in the ASOCIAL treatment were simply told that their probability of winning depended on their points score relative to a "given number." We did not tell subjects that this number was generated by the choice of a subject in an earlier session, as that might have introduced a social element into the ASOCIAL treatment. All references to other players were removed from the instructions. These procedures were adopted so that the second movers in the SOCIAL treatment and the subjects in the ASOCIAL treatment dealt with as similar a decision problem as possible *except for the presence of a rival*. Furthermore, in order to keep subjects' practical experiences with the slider task as similar as possible in both treatments, the participants in the ASOCIAL treatment were asked to wait 2 minutes before they started their tasks just as second movers had to wait 2 minutes for their paired first movers to finish the task before they started their own tasks in the SOCIAL treatment.

B. *Differences between Waves*

In the first wave, we used the GP software to run our SOCIAL sessions.[3] However, after observing that average effort (i.e., correctly positioned sliders) was systematically lower than in GP we realized that the visual length of each slider was slightly shorter than in GP's experiment because of the smaller screen size of the computer monitors in the CeDEx lab. This made the task somewhat more difficult for our subjects. Therefore, in the second wave of sessions we modified the slider screen so that the visual length of each slider would be exactly the same as in GP's original experiment. The third wave was the same as the second with three exceptions. First, instead of allowing the in-built random number generator to draw the prizes in the SOCIAL sessions, we used the realized prize values from GP.[4] Second, we only recruited inexperienced participants who had taken part in at most one other experiment. Third, we conducted all experimental sessions on weekdays at the same time of the day (14:00–15:30 hours). These exceptions were made to enhance comparability with GP.

C. *Procedures*

At the beginning of each session, experimental instructions were handed out to participants in paper form and were read aloud by the experimenter. For the SOCIAL treatment, we used exactly the same instructions as in GP. The instructions for our ASOCIAL treatment were adapted accordingly. All instructions are reproduced in Appendix A in Appendix S1, Supporting Information. Average earnings for participants were £14.14, including a £4 show-up fee, for a session lasting about 90 minutes. All sessions were conducted in the CeDEx lab at the University of Nottingham using z-Tree software (Fischbacher 2007) and volunteer subjects recruited via ORSEE (Greiner 2015) from the undergraduate student subject

3. Available for download at https://www.aeaweb.org/aer/data/feb2012/20100346_data.zip.

4. Although we used the prize realizations from the GP experiment we decided to retain the GP instructions. These stated "In each paying round, there will be a prize which you may win. Each prize will be chosen randomly at the beginning of the round and will be between £0.10 and £3.90." We might instead have explained that the prize draws were made in an earlier experiment, but we decided that explaining this to subjects would be potentially confusing. In either case, the subjects would learn the outcome of the random draw at the beginning of the round.

FIGURE 2
Development of First Mover Effort

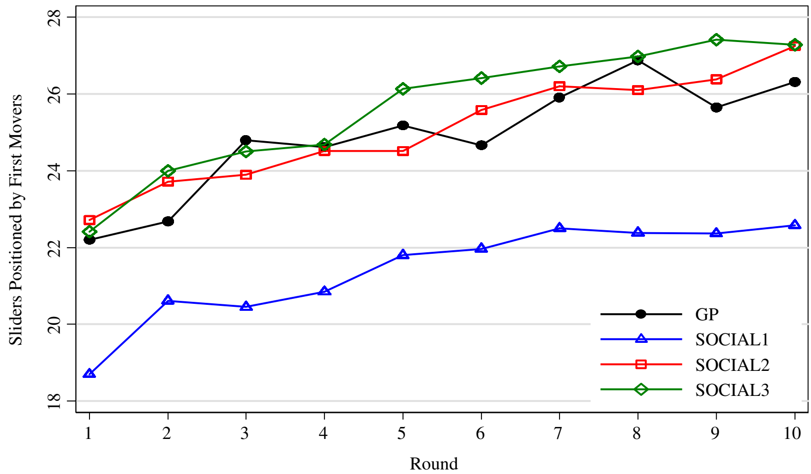


TABLE 1
Random Effects Regressions for First Mover Effort

	GP (1)	**SOCIAL1 (2)**	**SOCIAL2 (3)**	**SOCIAL3 (4)**
Prize	0.670*** (0.153)	0.230** (0.110)	0.243** (0.102)	0.282** (0.126)
Intercept	20.896*** (0.908)	18.284*** (0.588)	22.240*** (0.630)	21.868*** (0.639)
σ_ω	5.401	3.226	3.803	3.262
σ_ϵ	3.873	2.828	2.632	3.202
$N \times R$	600	600	600	600
Hausman test for random versus fixed effects	$\chi^2(10) = 0.00$ $p = 1.000$	$\chi^2(10) = 0.42$ $p = 1.000$	$\chi^2(10) = 1.88$ $p = .997$	$\chi^2(10) = 2.01$ $p = .996$

Notes: σ_ω denotes the standard deviation of the time invariant individual specific random effects and σ_ϵ denotes the standard deviation of the time varying idiosyncratic errors, which are i.i.d. over rounds and first movers. Standard errors are in parentheses. Round dummies (with the first round the omitted category) are included and are jointly significant at the 1% level in all cases.
***$p < .01$; **$p < .05$; *$p < .10$.

pool (excluding those who were then studying economics or psychology).

V. RESULTS

A. *First Mover Effort*

There are strong round effects in the data. This is seen in Figure 2, which shows average first mover effort by round for each wave (and, for comparison, in GP).[5] In our first wave, first mover effort was systematically lower than in GP. After modifying the slider screen for the second and third waves, average effort, and the development of effort across rounds, was much more in line with GP.

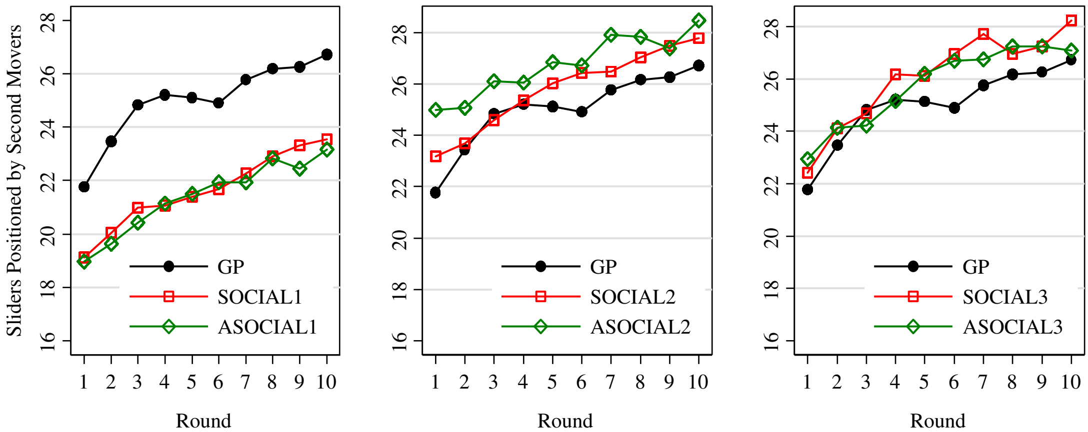
Note that Figure 2 does not condition efforts on prize values. To do this, we report random effects regressions, including round dummies to capture round effects. The results are reported in Table 1. Our estimates are similar across all three waves and, as expected, first movers supply more effort when the prize is higher. Note, however, that our estimate of the prize effect is somewhat lower than in GP.

B. *Second Mover Effort*

Figure 3 shows the development of average second mover effort over time in each of the

5. The GP data are available for download at https://www.aeaweb.org/articles?id=10.1257/aer.102.1.469.

FIGURE 3
Development of Second Mover Effort in the Three Waves



three waves.[6] For the sake of easy comparison, we include average GP second mover efforts in each panel. We find that second mover efforts are very similar in SOCIAL and ASOCIAL in each of the three waves. With the exception of the first wave, average second mover efforts tend to be somewhat higher in our data than in GP.

Recall that disappointment aversion predicts that a second mover would respond to higher first mover effort by decreasing her effort and even more so when competing for a higher prize. To test this discouragement effect, we use the same random effects panel data regression as in GP. Table 2 reports the estimates for each wave in SOCIAL and ASOCIAL.

Contrary to GP's estimates, which are reproduced in column (1), the coefficients on the regressors are generally insignificant: of 6×3 reported coefficients only two are significantly different from zero at the 10% level. To test whether second mover effort is neutral with respect to first mover effort, we test the joint significance of the coefficients on e_1 and Prize * e_1. In only one of the six cases, ASOCIAL1, is there a significant effect at the 10% level, and here the effect differs from GP in that there is a stronger discouragement effect at *lower* prize levels. Thus, we find only very limited evidence that second mover behavior is influenced by first mover effort, and we do not find the discouragement effect predicted by disappointment aversion. Note, however, that we find weak incentive effects in general: we also tested whether second mover efforts were sensitive to prizes (i.e., we tested the joint significance of Prize and Prize * e_1) and found a significant effect at the 10% level in only two cases (ASOCIAL1 and SOCIAL3; see Table 2).

C. *Structural Estimation*

We also repeated the structural estimation of disappointment aversion parameters following GP, using their preferred specification (GP, Table 3, p. 487). They assume a second mover's expected utility is as given in Equation (1), where the preference parameter λ_2 is assumed to be normally distributed in the population, with mean $\tilde{\lambda}_2$ and variance σ_λ^2. The cost function is specified as

$$C_2\left(e_2\right) = be_2 + c_{2,n,r}\, e_2^2/2,$$

where b is a parameter constant across subjects and rounds and the convexity parameter, $c_{2,n,r}$, for subject n in round r is given by

$$c_{2,n,r} = \kappa + \delta_r + \mu_n + \pi_{n,r},$$

where κ is constant across subjects and rounds, δ_r are round fixed effects, and μ_n are subject random

6. One participant from the first wave of the ASOCIAL treatment, one participant from the second wave of the ASOCIAL treatment, and one second mover from the third wave of the SOCIAL treatment are dropped from our data analysis because they appear to have been unable to position any slider correctly. GP also found one second mover did not position any slider correctly in their experimental sample and drop this participant from their main data analysis. Neither our, nor their, main findings are affected by the inclusion or exclusion of these participants.

TABLE 2
Random Effects Regressions for Second Mover Effort

	GP (1)	**SOCIAL1 (2)**	**ASOCIAL1 (3)**	**SOCIAL2 (4)**	**ASOCIAL2 (5)**	**SOCIAL3 (6)**	**ASOCIAL3 (7)**
e_1	0.044 (0.049)	0.019 (0.045)	−0.096* (0.051)	0.007 (0.051)	−0.006 (0.042)	−0.081 (0.057)	−0.008 (0.049)
Prize	1.639*** (0.602)	0.408 (0.452)	−0.207 (0.510)	0.268 (0.577)	−0.176 (0.470)	−0.901 (0.626)	−0.286 (0.537)
Prize * e_1	−0.049** (0.023)	−0.016 (0.021)	0.020 (0.023)	−0.006 (0.023)	0.014 (0.019)	0.044* (0.024)	0.018 (0.021)
Intercept	19.777*** (1.400)	18.585*** (1.100)	20.454*** (1.169)	22.749*** (1.423)	24.849*** (1.252)	24.100*** (1.551)	22.915*** (1.428)
σ_ω	4.288	3.760	2.768	3.759	4.688	3.285	4.847
σ_ϵ	3.852	2.473	2.777	3.003	2.429	3.173	2.753
H0: no e_1 effect	$\chi^2(2) = 7.09$	$\chi^2(2) = 0.75$	$\chi^2(2) = 5.28$	$\chi^2(2) = 0.09$	$\chi^2(2) = 1.48$	$\chi^2(2) = 3.35$	$\chi^2(2) = 1.97$
	$p = .029$	$p = .687$	$p = .071$	$p = .956$	$p = .477$	$p = .188$	$p = .374$
H0: no prize effect	$\chi^2(2) = 12.08$	$\chi^2(2) = 1.13$	$\chi^2(2) = 4.84$	$\chi^2(2) = 1.18$	$\chi^2(2) = 4.24$	$\chi^2(2) = 6.15$	$\chi^2(2) = 3.49$
	$p = .002$	$p = .568$	$p = .089$	$p = .555$	$p = .120$	$p = .046$	$p = .175$
$N \times R$	590	600	590	600	590	590	600
Hausman test for random versus fixed effects	$\chi^2(12)=2.60$	$\chi^2(12) = 1.61$	$\chi^2(12) = 0.08$	$\chi^2(12) = 0.49$	$\chi^2(12) = 0.05$	—	$\chi^2(12) = 1.06$
	$p = 0.998$	$p = 1.000$	$p = 1.000$	$p = 1.000$	$p = 1.000$	—	$p = 1.000$

Notes: σ_ω denotes the standard deviation of the time invariant individual specific random effects and σ_ϵ denotes the standard deviation of the time varying idiosyncratic errors, which are i.i.d. over rounds and second movers. Standard errors are in parentheses. Round dummies (with the first round the omitted category) are included and are jointly significant at the 1% level in all cases. The Hausman test statistic is missing in SOCIAL3 because the estimated covariance matrix of the difference between random and fixed effects estimators is not invertible. Subjects who failed to position any sliders correctly in any round are excluded from the analysis.

***$p < .01$; **$p < .05$; *$p < .10$.

TABLE 3
Structural Estimates of Disappointment Aversion

	GP (1)	SOCIAL1 (2)	ASOCIAL1 (3)	SOCIAL2 (4)	ASOCIAL2 (5)	SOCIAL3 (6)	ASOCIAL3 (7)
$\tilde{\lambda}_2$	1.767*** (0.763)	0.458 (0.446)	0.856 (1.198)	−0.591 (0.647)	−1.929*** (0.490)	−0.124 (0.687)	−0.629 (0.493)
σ_λ	1.871** (0.875)	0.777* (0.467)	0.780 (1.310)	0.794 (0.507)	1.408*** (0.243)	2.253 (1.500)	0.680 (0.690)
b	−0.442*** (0.021)	−0.551*** (0.032)	−0.566*** (0.032)	−0.496*** (0.039)	−0.644*** (0.030)	−0.576*** (0.028)	−0.617*** (0.020)
κ	2.008*** (0.101)	1.675*** (0.147)	1.998*** (0.098)	1.882*** (0.129)	1.986*** (0.142)	1.927*** (0.109)	1.924*** (0.125)
σ_μ	0.692*** (0.127)	0.434*** (0.069)	0.441*** (0.065)	0.581*** (0.082)	0.652*** (0.107)	0.525*** (0.068)	0719*** (0.131)
σ_π	0.369*** (0.075)	0.279*** (0.039)	0.338*** (0.071)	0.445*** (0.111)	0.218*** (0.032)	0.344*** (0.097)	0.340*** (0.091)
de_2/de_1 (v = £0.10 , low λ_2)	−0.000 (0.001)	0.000 (0.001)	−0.000 (0.001)	0.000 (0.001)	0.001 (0.001)	0.001 (0.002)	0.001 (0.001)
de_2/de_1 (v = £2, average λ_2)	−0.033* (0.017)	−0.006 (0.006)	−0.015 (0.016)	−0.001 (0.003)	0.000 (0.000)	−0.003 (0.007)	−0.000 (0.002)
de_2/de_1 (v = £3.90, high λ_2)	−0.135*** (0.049)	−0.031* (0.018)	−0.059 (0.064)	−0.024 (0.018)	−0.036*** (0.011)	−0.068 (0.051)	−0.019 (0.017)
$N \times R$	590	600	590	600	590	590	600
OI test	14.030 [0.829]	33.752 [0.028]	33.450 [0.030]	27.076 [0.133]	22.691 [0.304]	29.708 [0.075]	30.043 [0.069]

Notes: The estimation method is MSM and follows the preferred model in GP, Table 3. Standard deviations of the transitory and persistent unobservables in the cost of effort function, σ_π and σ_μ, are computed from the estimates of the parameters of the Weibull distribution. Estimates of κ, σ_π, and σ_μ have been multiplied by 100. All specifications further include dummy variables for each of rounds 2–10 inclusive. Reaction functions and gradients were produced by simulation methods. Using the estimated parameters of the cost of effort function for round 5, we simulated a large number of second mover optimal efforts conditional on specific values of first mover effort and the prize, and computed the mean best response. The reaction functions are linear. Low, average, and high λ_2 refer to the 20th, 50th, and 80th percentiles of the distribution of λ_2, respectively. Standard errors are in parentheses. Newey overidentification (OI) test statistics are reported with p values shown in brackets. Subjects who failed to position any sliders correctly in any round are excluded from the analysis. Four sliders are added to each effort score in SOCIAL1 and ASOCIAL1.

***$p < .01$; **$p < .05$; *$p < .10$.

effects assumed to be independent draws from a Weibull distribution with scale parameter φ_μ and shape parameter ϕ_μ. The final term, $\pi_{n,r}$, is a random shock that varies across subjects and rounds and follows a Weibull distribution with scale parameter ϕ_π and shape parameter φ_π.

There are 17 parameters to be estimated: $\tilde{\lambda}_2, \sigma_\lambda^2, b, \kappa, \phi_\mu, \varphi_\mu, \phi_\pi$, and φ_π, in addition to the nine round fixed effects (the first round providing the omitted category). Estimation is done using the method of simulated moments (MSM). First, selected moments of the experimental data are calculated (see Appendix B in Appendix S1 for the observed moments). Next, using trial values of the unknown parameters and the first mover efforts and prize values from the experimental data, 30 simulated samples of second mover efforts are created, each involving ten rounds of effort choices by 60-second movers. A goodness of fit metric is then computed based on the differences between the average moments across the simulated samples and the corresponding moments in the experimental data. This process is repeated using a simulated annealing method to search over the parameter space for the MSM estimates that maximize the goodness of fit.

Our estimates (including a reestimation using GP data) are reported in Table 3 below.[7] Estimates of the cost function are broadly in line with GP. The estimate of b is significantly negative and there is significant within-subject and between-subject variation in the cost of effort function, with more variation between- than within-subject. However, our results are quite different when it comes to the λ_2 parameter, which measures the difference between the second mover's marginal utility of earnings in the loss and the gain domain (where expected earnings define the reference point). Whereas the estimate of the mean of the distribution, $\tilde{\lambda}_2$, is positive and significant in GP, our estimates are much lower and generally insignificant. Indeed, in four of the six cases, the estimate is negative. In the second wave of ASOCIAL the negative estimate of $\tilde{\lambda}_2$ is significant. This is an unexpected finding and taken at face value it would mean that subjects weight gains more heavily than losses. We are reluctant to put much weight on this finding, as it is clear from Table 3 that the estimates vary considerably across waves. Moreover, in four of the six cases the model over-identification test is rejected at the 10% level.

Building on the observation that asymmetric inequality aversion affects the strength of the discouragement effect, we also estimated a richer model in which expected utility is given by Equation (2), while retaining the same cost function specification. To estimate the richer model, we used a subject random effect specification for the social preference parameter, $(1 + \alpha_2 - \beta_2)$, assuming that this follows a Weibull distribution.[8] We fit simulated moments to the same set of moments as before. The richer model has two additional parameters corresponding to the scale and shape parameters of the social preference parameter distribution. MSM estimates are presented in Table 4. Rather than report estimates of the scale and shape parameters of the social preference parameter distribution, we report the implied estimates of the mean and standard deviation of this distribution.

For the GP data, we observe significant variation in social preference parameters, but this has little effect on the qualitative features of the other estimates. As with the simpler model, the estimated mean of the distribution of disappointment aversion parameters is significantly positive, and there is a significant estimated discouragement effect at medium and high prize values. Also, in our data, estimating the richer model does not have much of an effect on cost function estimates. However, estimation of the richer model for our data does not deliver a consistent picture of how preference parameters affect behavior. The estimates of $\tilde{\lambda}_2$ are insignificant in all cases, and the estimates of the mean of the distribution of social preferences are also insignificantly different from one. Given our earlier evidence that prize effects are weak in our data, perhaps it is unsurprising that we fail to observe significant results from structural estimation of more nuanced prize

7. Note, although we used the exact code provided by GP to create the GP column, the estimates differ slightly from those reported in their paper due to simulation noise. For our datasets, we also used the GP code, except that we had to increase the bounds of the parameter space. For our first wave the code did not deliver MSM estimates. The reason is that some of the observed moments are calibrated from the GP sample (e.g., the proportion of second mover efforts exceeding 35). In our first wave average efforts, aggregating over all subjects and rounds, are lower than in GP by about four sliders, and there are no observations where second mover efforts exceed 35. In order to calculate estimates using the GP code and the moments selected by GP, we added four sliders to each score in the first wave.

8. Note that the richer model essentially allows different individuals to have different effective prize spreads for a given prize value. We considered it important to impose a positive effective prize spread (i.e., *ceteris paribus*, winning is preferred to losing) and so used a distribution with non-negative support for the social preference parameter.

TABLE 4
Structural Estimates of the Distributions of Disappointment Aversion and Social Preference Parameters

	GP (1)	SOCIAL1 (2)	ASOCIAL1 (3)	SOCIAL2 (4)	ASOCIAL2 (5)	SOCIAL3 (6)	ASOCIAL3 (7)
$\tilde{\lambda}_2$	1.537*** (0.619)	1.797 (1.292)	1.991 (1.675)	−1.108 (3.373)	−1.861 (1.549)	0.089 (1.956)	−1.814 (1.685)
σ_λ	2.304*** (0.384)	2.440 (1.700)	1.748 (1.612)	1.111 (2.892)	1.401 (2.151)	2.108* (1.146)	0.820 (1.276)
b	−0.537*** (0.015)	−0.623*** (0.018)	−0.569*** (0.012)	−0.572*** (0.021)	−0.586*** (0.029)	−0.524*** (0.021)	−0.626*** (0.018)
κ	1.780*** (0.047)	1.846*** (0.078)	1.818*** (0.049)	1.815*** (0.072)	1.753*** (0.122)	1.726*** (0.084)	1.932*** (0.102)
σ_μ	0.460*** (0.054)	0.425*** (0.059)	0.306*** (0.046)	0.448*** (0.064)	0.492*** (0.040)	0.424*** (0.088)	0.794*** (0.151)
σ_π	0.586*** (0.203)	0.241*** (0.023)	0.272*** (0.045)	0.262*** (0.059)	0.260*** (0.100)	0.320*** (0.084)	0.342*** (0.131)
$\text{Mean}_{1+\alpha-\beta}$	1.375 (0.397)	0.646 (0.597)	0.469 (0.485)	0.629 (1.190)	0.833 (0.960)	0.874 (0.402)	0.649 (0.416)
$\sigma_{1+\alpha-\beta}$	2.169*** (0.703)	0.820 (0.962)	0.520 (0.828)	0.846 (1.632)	1.389 (1.444)	1.055 (0.716)	1.171 (0.762)
de_2/de_1 ($v=$ £0.10, low λ_2)	0.001 (0.003)	0.001 (0.004)	0.002 (0.020)	0.003 (0.025)	0.007 (0.032)	0.001 (0.003)	0.001 (0.002)
de_2/de_1 ($v=$ £2, average λ_2)	−0.038* (0.019)	−0.017 (0.016)	−0.016 (0.016)	−0.001 (0.004)	−0.001 (0.005)	−0.016 (0.024)	−0.002 (0.008)
de_2/de_1 ($v=$ £3.90, high λ_2)	−0.111*** (0.036)	−0.068 (0.052)	−0.051 (0.044)	−0.009 (0.019)	−0.016 (0.017)	−0.086 (0.066)	−0.023 (0.032)
$N \times R$	590	600	590	600	590	590	600
OI test	14.296 [0.766]	41.381 [0.002]	26.818 [0.109]	7.144 [0.993]	15.333 [0.701]	27.618 [0.091]	23.181 [0.230]

Notes: The estimation method is MSM and follows a richer model with social preference parameters. Standard deviations of the transitory and persistent unobservables in the cost of effort function, σ_π and σ_μ, and mean and standard deviation of the individual specific social preference parameters, $\text{Mean}_{1+\alpha-\beta}$ and $\sigma_{1+\alpha-\beta}$ are computed from the estimates of the parameters of the Weibull distribution. Estimates of κ, σ_π, and σ_μ have been multiplied by 100. All specifications further include dummy variables for each of rounds 2–10 inclusive. Reaction functions and gradients were produced by simulation methods. Using the estimated parameters of the cost of effort function for round 5 and the estimated social preference parameters, we simulated a large number of second mover optimal efforts conditional on specific values of first mover effort and the prize, and computed the mean best response. The reaction functions are linear. Low, average, and high λ_2 refer to the 20th, 50th, and 80th percentiles of the distribution of λ_2, respectively. Standard errors are in parentheses. The estimates of $\text{Mean}_{1+\alpha-\beta}$ are tested against 1; all other estimates are tested against 0. Newey overidentification (OI) test statistics are reported with p values shown in brackets. Subjects who failed to position any sliders correctly in any round are excluded from the analysis. Four sliders are added to each effort score in SOCIAL1 and ASOCIAL1.

***$p < .01$; **$p < .05$; *$p < .10$.

TABLE 5
Random Effects Regressions for Second Mover Effort (Pooling SOCIAL and ASOCIAL)

	Wave 1 **(1)**	**Wave 2** **(2)**	**Wave 3** **(3)**	**All waves** **(4)**
e_1	0.023 (0.047)	0.015 (0.046)	−0.080 (0.053)	−0.022 (0.027)
Prize	0.454 (0.476)	0.231 (0.524)	−0.995* (0.582)	−0.200 (0.287)
Prize * e_1	−0.018 (0.022)	−0.004 (0.021)	0.048** (0.022)	0.014 (0.012)
ASOCIAL	2.068 (1.567)	1.642 (1.839)	−2.027 (2.060)	0.379 (1.012)
ASOCIAL * e_1	−0.124* (0.067)	−0.030 (0.066)	0.070 (0.074)	−0.015 (0.038)
ASOCIAL * Prize	−0.729 (0.673)	−0.380 (0.741)	0.789 (0.816)	−0.010 (0.406)
ASOCIAL * Prize * e_1	0.042 (0.031)	0.016 (0.029)	−0.032 (0.032)	0.003 (0.017)
Wave 1				−4.578*** (0.017)
Intercept	18.499*** (1.115)	22.983*** (1.316)	24.527*** (1.476)	23.677*** (0.742)
σ_ω	3.307	4.245	4.147	3.897
σ_ϵ	2.622	2.739	2.971	2.784
H_0: no *treatment* effect	$\chi^2(4) = 4.99$ $p = .288$	$\chi^2(4) = 1.70$ $p = .790$	$\chi^2(4) = 1.23$ $p = .873$	$\chi^2(4) = 0.69$ $p = .953$
$N \times R$	1,190	1,190	1,190	3,570
Hausman test for random versus fixed effects	— —	$\chi^2(15) = 0.38$ $p = 1.000$	— —	$\chi^2(15) = 5.29$ $p = .989$

Notes: σ_ω denotes the standard deviation of the time invariant individual specific random effects and σ_ϵ denotes the standard deviation of the time varying idiosyncratic errors, which are i.i.d. over rounds and second movers. Standard errors are in parentheses. Round dummies (with the first round the omitted category) are included and are jointly significant at the 1% level in all cases. The Hausman test statistic is missing in Wave 1 and Wave 3 because the estimated covariance matrix of the difference between random and fixed effects estimators is not invertible. Subjects who failed to position any sliders correctly in any round are excluded from the analysis.

***$p < .01$; **$p < .05$; *$p < .10$.

effects, such as asymmetric weighting of gains and losses relative to expectations and variation in effective prize spreads across subjects.

D. *Treatment Differences*

Since our main focus is whether interpersonal comparisons affect behavior, we also compared second mover behavior across our SOCIAL and ASOCIAL treatments. To do this, our first approach exploits the fact that in our design each SOCIAL second mover has a counterpart in the ASOCIAL treatment who had to complete the same slider task for the same prize and same monetary incentives. For each SOCIAL second mover, we take the average number of sliders positioned across all ten paying rounds, and compare this average to that of their counterpart in the ASOCIAL treatment, using a Wilcoxon signed-ranks test. In none of the waves do we observe a significant treatment effect (first wave $p = .497$, second wave $p = .280$, third wave $p = .979$).

Second, we reran random effects regressions of second mover efforts on prizes, e_1 and an interaction term for the pooled SOCIAL + ASOCIAL samples, adding treatment dummies and treatment interactions. The results are in Table 5. In addition, we report the test of the joint significance of the treatment dummies and treatment interactions. In none of the three waves is there a significant treatment effect. We also find neither a treatment effect nor a significant influence of e_1 and prizes when we pool all waves (including a dummy variable for the first wave to capture the level shift caused by modifying the slider screen; see Section III.A).

VI. DISCUSSION AND CONCLUSION

In this study, we have tested the hypothesis that disappointment aversion, and the discouragement effect produced by disappointment aversion, will be different in a competition against another person and a competition against Nature. Our experiments reject this hypothesis for the environment in which we studied this question—GP's two-person sequential effort competition using the slider task as a real effort performance measurement. We conclude that social comparison effects do not influence estimates of disappointment aversion. Of course, this does not necessarily imply that our finding is generalizable to other environments or real effort tasks.

An unexpected finding of our experiment is that, in contrast to GP, we failed to find support for disappointment aversion in any wave of any treatment. Despite this being a "null result" it is nevertheless important to report. Our results, taken together with GP, provide mixed evidence on models of expectation-based reference dependence (e.g., Kőszegi and Rabin 2007). These models assume individuals evaluate gains and losses asymmetrically around a reference point, where the reference point is expectations-based and depends on the choice made. The assumption that individuals, in making a choice, take into account that the choice will affect the reference point about which gains and losses are evaluated seems quite sophisticated, but whether actual behavior reflects such behavioral assumptions is an empirical question. Whereas GP find support, this is not echoed in our study. Similarly, Abeler et al. (2011) and Ericson and Fuster (2011) find choices consistent with endogenous expectations-based reference points, whereas Holzmeister et al. (2015) and Altmejd et al. (2015) who replicated Abeler et al. (2011) and Ericson and Fuster (2011), respectively, find no or only weakly significant support. Gneezy et al. (2017) also replicated the Abeler et al. (2011) design, although in additional treatments they found important deviations from expectations-based reference dependence.

Another unexpected finding of our study was that in all three of our waves we observed much weaker first mover prize effects than GP, and, contrary to GP, almost no second mover prize effects. These weaker prize effects may explain the absence of discouragement effects. If second movers are not responsive to monetary prizes, either because they place little value on the prize or because their marginal cost of effort is too high then it is perhaps not surprising that they are unresponsive to first mover efforts. It is worth noting, however, that our observed means and dynamics in effort over time are very similar to GP; if subjects have a positive marginal cost of effort and place no value on the prize they should not be positioning sliders at all. Our result of a weak or no prize effect is consistent with Araujo et al. (2016) who perform a between-subject comparison of slider task performance varying incentives and find only weak incentive effects. There are, however, other studies that do find that slider positioning efforts respond positively and significantly to increases in financial incentives (e.g., Abeler and Jäger 2015; Lee 2015; Neckermann, Warnke, and Bradler 2014). Note that nonmonotonic responses to changes in piece rates are observed in between-subject studies of other real-effort tasks (e.g., Ariely et al. 2009; Gneezy and Rustichini 2000).

In conclusion, we believe our paper makes two contributions to the literature. First, we have shown that social comparison effects do not matter in the contest we studied. Second, our study contributes to recent replication efforts in the experimental economics literature (Camerer et al. 2016)[9] to establish a body of knowledge about results that are robust to replication. In particular, we show that in the contest we study, reference-point effects are surprisingly weak.

ACKNOWLEDGMENT

This work was supported by the Economic and Social Research Council [grant numbers ES/K002201/1, ES/J500100/1]; and the European Research Council Advanced Investigator [grant number 295707 COOPERATION].

REFERENCES

Abeler, J., A. Falk, L. Goette, and D. Huffman. "Reference Points and Effort Provision." *American Economic Review*, 101, 2011, 470–92.

Abeler, J., and S. Jäger. "Complex Tax Incentives." *American Economic Journal: Economic Policy*, 7, 2015, 1–28.

Altmejd, A., A. Dreber, E. Heikensten, and M. Johannesson. "Replication of 'Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments' by Ericson, K.M.M./Fuster, A. (2011) in: The Quarterly Journal of Economics, 126(4), pp. 1879–1907." Experimental Economics Replication Project, 2015. https://experimentaleconreplications.com/finalreports/Ericson&Fuster2011.pdf

Araujo, F. A., E. Carbone, L. Conell-Price, M. W. Dunietz, A. Jaroszewicz, R. Landsman, D. Lamé, L. Vesterlund, S. W. Wang, and A. J. Wilson. "The Slider Task: An Example of Restricted Inference on Incentive Effects." *Journal of the Economic Science Association*, 2, 2016, 1–12.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar. "Large Stakes and Big Mistakes." *Review of Economic Studies*, 76, 2009, 451–69.

Bell, D. E. "Disappointment in Decision-Making under Uncertainty." *Operations Research*, 33, 1985, 1–27.

Blascovich, J., W. B. Mendes, S. B. Hunter, and K. Salomon. "Social 'Facilitation' as Challenge and Threat." *Journal of Personality and Social Psychology*, 77, 1999, 68–77.

Bougheas, S., J. Nieboer, and M. Sefton. "Risk-Taking in Social Settings: Group and Peer Effects." *Journal of Economic Behavior & Organization*, 92, 2013, 273–83.

Camerer, C. F., A. Dreber, E. Forsell, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, E. Heikensten, F. Holzmeister, T. Imai, S. Isaksson, G. Nave, T. Pfeiffer, M. Razen, and H. Wu. "Evaluating Replicability of Laboratory Experiments in Economics." *Science*, 351, 2016, 1433–6.

Cooper, D. J., and M. Rege. "Misery Loves Company: Social Regret and Social Interaction Effects in Choices under Risk and Uncertainty." *Games and Economic Behavior*, 73, 2011, 91–110.

9. See also https://experimentaleconreplications.com/finalreports/Abeler%20et%20al.%202011.pdf

Delquié, P., and A. Cillo. "Disappointment without Prior Expectation: A Unifying Perspective on Decision under Risk." *Journal of Risk and Uncertainty*, 33, 2006, 197–215.

Dijk, O., M. Holmen, and M. Kirchler. "Rank Matters—The Impact of Social Competition on Portfolio Choice." *European Economic Review*, 66, 2014, 97–110.

Dohmen, T., A. Falk, K. Fliessbach, U. Sunde, and B. Weber. "Relative versus Absolute Income, Joy of Winning, and Gender: Brain Imaging Evidence." *Journal of Public Economics*, 95, 2011, 279–85.

Eisenkopf, G., and S. Teyssier. "Envy and Loss Aversion in Tournaments." *Journal of Economic Psychology*, 34, 2013, 240–55.

Ericson, K. M. M., and A. Fuster. "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments." *Quarterly Journal of Economics*, 126, 2011, 1879–907.

Fafchamps, M., B. Kebede, and D. J. Zizzo. "Keep Up with the Winners: Experimental Evidence on Risk Taking, Asset Integration, and Peer Effects." *European Economic Review*, 79, 2015, 59–79.

Falk, A., and A. Ichino. "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 24, 2006, 38–57.

Fehr, E., and K. M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 1999, 817–68.

Fischbacher, U. "Z-Tree: Zurich Toolbox for Readymade Economic Experiments." *Experimental Economics*, 10, 2007, 171–8.

Gächter, S., D. Nosenzo, and M. Sefton. "Peer Effects in Pro-Social Behavior: Social Norms or Social Preferences?" *Journal of the European Economic Association*, 11, 2013, 548–73.

Gill, D., Z. Kissova, J. Lee, and V. Prowse. "First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision." Department of Economics Discussion Paper No. 783, University of Oxford, 2016,

Gill, D., and V. Prowse. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review*, 102, 2012, 469–503.

Gill, D., and R. Stone. "Fairness and Desert in Tournaments." *Games and Economic Behavior*, 69, 2010, 346–64.

Gneezy, U., L. Goette, C. Sprenger, and F. Zimmermann. "The Limits of Expectations-Based Reference Dependence." *Journal of the European Economic Association*, 15, 2017, 861–76.

Gneezy, U., and A. Rustichini. "Pay Enough or Don't Pay at All." *Quarterly Journal of Economics*, 115, 2000, 791–810.

Greiner, B. "Subject Pool Recruitment Procedures: Organizing Experiments with Orsee." *Journal of the Economic Science Association*, 1, 2015, 114–25.

Herbst, D., and A. Mas. "Peer Effects on Worker Output in the Laboratory Generalize to the Field." *Science*, 350, 2015, 545–9.

Herrmann, B., and H. Orzen. "The Appearance of Homo Rivalis: Social Preferences and the Nature of Rent-Seeking." CeDEx Discussion Paper No. 2008-10, University of Nottingham, 2008.

Holzmeister, F., J. R. Huber, M. Kirchler, and M. Razen. "Replication of 'Reference Points and Effort Provision' by Abeler, J./Falk, A./Goette, L./Huffman, D. (2011) in: The American Economic Review, 101(2), pp. 470–492." Experimental Economics Replication Project, 2015. https://experimentaleconreplications.com/finalreports/Abeleret al.2011.pdf

Köszegi, B., and M. Rabin. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121, 2006, 1133–65.

———. "Reference-Dependent Risk Attitudes." *American Economic Review*, 97, 2007, 1047–73.

Lee, J. "How Does Competition Affect People Economically and Emotionally?" Manuscript, Department of Economics, Cornell University, 2015. https://drive.google.com/file/d/0B-ml6reB6PwFYXctaE5YMkpvS0E/view?usp=sharing

Linde, J., and J. Sonnemans. "Social Comparison and Risky Choices." *Journal of Risk and Uncertainty*, 44, 2012, 45–72.

Loomes, G., and R. Sugden. "Disappointment and Dynamic Consistency in Choice under Uncertainty." *Review of Economic Studies*, 53, 1986, 271–82.

Markus, H. "Effect of Mere Presence on Social Facilitation: An Unobtrusive Test." *Journal of Experimental Social Psychology*, 14, 1978, 389–97.

Mas, A., and E. Moretti. "Peers at Work." *American Economic Review*, 99, 2009, 112–45.

Neckermann, S., A. J. Warnke, and C. Bradler. "Rewards and Performance: A Comparison across Creative and Routine Tasks." *Academy of Management Proceedings*, 2014, 2014, 14335.

Schwerter, F. "Social Reference Points and Risk Taking." Discussion Paper, University of Bonn, 2016.

Thöni, C., and S. Gächter. "Peer Effects and Social Preferences in Voluntary Cooperation: A Theoretical and Experimental Analysis." *Journal of Economic Psychology*, 48, 2015, 72–88.

Zajonc, R. B. "Social Facilitation." *Science*, 149, 1965, 269–74.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Instructions and Observed Moments