



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Performance curiosity

Alós-Ferrer, Carlos ; García-Segarra, Jaume ; Ritschel, Alexander

DOI: <https://doi.org/10.1016/j.joep.2017.08.002>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-151877>

Journal Article

Accepted Version

Originally published at:

Alós-Ferrer, Carlos; García-Segarra, Jaume; Ritschel, Alexander (2018). Performance curiosity. *Journal of Economic Psychology*, 64:1-17.

DOI: <https://doi.org/10.1016/j.joep.2017.08.002>

Performance Curiosity*

Carlos Alós-Ferrer[†] Jaume García-Segarra[‡] Alexander Ritschel[§]
University of Cologne University of Cologne University of Cologne

This Version: February 2018

Abstract

We show that performance curiosity – the desire to know one’s own (relative) performance – can trump inequality aversion. In two experiments (combined $N = 450$), participants chose between an equal allocation and a performance-based one after generating surplus in a real-effort task. In the experimental treatment, choosing an equal allocation came at the cost of not knowing the own performance, which led to a substantial increase of performance-based choices in comparison with the control treatment. The effect seems especially pronounced for women, but the gender effect is due to a difference in expectations regarding performance. Interestingly, the manipulation equalized the proportion of equal allocation choices between males and females compensating for their difference in expectations.

JEL Classification: C91 · D03 · J16

PsycInfo Codes: 2360 · 3020

Keywords: Social Preferences · Egalitarian Behavior · Expectations · Performance Curiosity

Published. This is an author-generated version of a research manuscript which has been published in the *Journal of Economic Psychology*. It is self-archived in scientific repositories and authors’ webpages for the purpose of facilitating scientific discussion. Please cite the published version:

Alós-Ferrer, C., García-Segarra, J., and Ritschel, A. (2018). Performance Curiosity. *Journal of Economic Psychology*, 64: 1–17.

*We thank Anja Achtziger, Gary Charness, Urs Fischbacher, Uri Gneezy, David Laibson, Daniel Navarro-Martínez, the co-editor Stefan Schulz-Hardt, and two anonymous referees for helpful comments. Alexander Ritschel was financed by the Research Unit “Psychoeconomics” (FOR 1882) of the German Research Foundation (DFG). The Department of Economics at the University of Cologne gratefully acknowledges financial support from the DFG to build the Cologne Laboratory for Economic Research.

[†]Corresponding author: carlos.alos-ferrer@econ.uzh.ch. Department of Economics, University of Zurich. Blümlisalpstrasse 10, 8006 Zurich, Switzerland.

[‡]Department of Economics, University of Cologne. Albertus-Magnus-Platz, 50923 Cologne, Germany.

[§]Department of Economics, University of Zurich. Blümlisalpstrasse 10, 8006 Zurich, Switzerland.

1 Introduction

When different people exert different levels of effort, an egalitarian distribution of jointly-generated proceeds is not necessarily “fair.” In this sense, egalitarianism might not always be socially desirable. If agents anticipate the results of their effort, social conventions dictating egalitarian allocations can greatly diminish incentives to exert effort to achieve a high performance.

This simple observation points to a conflict arising from different human motivations. On the one hand, most human beings strive to perform well. Indeed, psychological research has identified *achievement motivation* as a basic (intrinsic) motive (e.g., McClelland et al., 1989; Brunstein and Heckhausen, 2008). This goes hand-in-glove with the view that individual effort should be rewarded. On the other hand, research in behavioral economics has identified *inequality aversion*, which implies a preference in favor of egalitarian outcomes even if they result in a reduction of the own payoffs, as an important factor in decisions concerning (re)distribution (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). When confronted with the decision to allocate the proceeds from individual effort, these two motivations result in opposite tendencies, which are at the heart of discussions on many socioeconomic issues, ranging from performance pay within firms to redistribution of income through tax systems.

In this work, we contribute to the investigation into the motivations underlying preferences among distributional allocations. We aim to show that preferences for egalitarian redistribution, as opposed to rewarding individual performance, hang in a fragile balance and can be significantly reduced with subtle interventions. In particular, we focus on a manipulation derived from *performance curiosity*, defined as the desire to know the own performance (especially in relative terms). The rationale is that the motivation to achieve and perform well is linked to self-reputation and self-image concerns, which are themselves closely linked to the notion of *identity* (Mazar et al., 2008; Bénabou and Tirole, 2011). Obviously, information on the own performance has a crucial impact on the self-image. We hence postulate that human beings might be willing to give up the benefits of their own effort to appease their inequality aversion, but this willingness only lasts as long as they at least receive the information on how well they have performed. Specifically, if social conventions, on top of imposing an egalitarian redistribution, go to the extreme of eliminating (relative) performance feedback, we postulate that preferences for egalitarian distributions will be greatly reduced.

In this paper, we report on two experiments with an innovative design which pits inequality aversion against curiosity regarding the own performance in a real-effort task. The task is used to generate surplus, and is followed by a distributional allocation decision within a group. The design creates a tradeoff between an egalitarian allocation and receiving information on performance in the real-effort task. We show that this subtle manipulation can shift preferences away from egalitarian allocations.

In our first experiment ($N = 180$) participants generated income by working on a real-effort task and subsequently decided on the allocation of the joint proceedings within a small group. In the control treatment, participants chose between an egalitarian allocation and a performance-based one, leading to the standard observation of inequality aversion. In the experimental (“No Info”) treatment, the only change was that opting for the egalitarian distribution came at the cost of not knowing the actual performance and ranking in the real-effort task. The design highlights *relative performance*, because the information, if revealed, includes the performance and rank of all group members, enabling social comparisons. This change had a large effect, with participants choosing the performance-based distribution rather than the egalitarian one. In the second experiment ($N = 270$) we replicated the results of the first experiment and also added a third treatment and additional questionnaires to test for further explanations of the basic effect.

Our tasks were explicitly incentivized. Hence, given the own expected performance, participants could easily compute which of the two allocations would maximize their (expected) monetary rewards. The allocations were designed in such a way that only those expecting to be strictly above average would be better off under the performance-based allocation, with all others being better off under the egalitarian allocation. Hence, we elicited expected performance in the real-effort task and controlled for it by testing the basic hypotheses also within different groups (in terms of expectations) and by including the corresponding variable as a control in our regressions. As was to be expected, a higher expectation led to a higher percentage of performance-based choices, but the basic effect remains clearly significant.

In view of evidence on gender differences and competition (e.g., Gneezy et al., 2003; Niederle and Vesterlund, 2007), we were also interested in gender effects. For this reason, we were led to a design with perfectly balanced participation across gender and a sample size large enough to examine gender differences. Specific hypotheses regarding those can also be derived from research on inequality aversion. Some studies in this field have found gender differences, as reflected, e.g., in the proportion of egalitarian allocations in Dictator and Ultimatum games (see, e.g., Fehr et al., 2006). However, the evidence is mixed. Croson and Gneezy (2009) argue that there is no difference in social preferences, and that the observed effects arise due to gender differences in the sensitivity to cues in experimental contexts. The latter hypothesis is motivated by research in psychology establishing that women are more sensitive to social cues and feedback than men (Gilligan, 1982; Roberts and Nolen-Hoeksema, 1989). Since our manipulation involves the provision of feedback, it is natural to expect that the basic effect should be especially pronounced for women, compared to men. Indeed, the treatment effect in our experiments was clearly stronger among women, with behavior in the control treatment exhibiting clear gender differences which vanished in the No Info treatment. Controlling for expected performance allows to uncover the roots of this gender effect: women had a lower expected performance than men, leading to a lower percentage of performance-

based choices in the control treatment (due to extrinsic, monetary rewards) which was overcome in the experimental treatment.

We conclude that performance curiosity, i.e. wanting to know one's own (relative) performance, can counteract inequality aversion as a motivation and tilt decisions away from egalitarian distributions and toward performance-based ones. Additionally, creating a tradeoff between information and egalitarian allocations substantially reduces gender differences in behavior, with women becoming more willing to accept performance-based allocations. This is a potentially important insight for incentive design and the reduction of gender differences.

The paper is organized as follows. Section 2 discusses some related literature strands and helps place our experiments in a broader context. Section 3 describes the design of Experiment 1 and presents the analysis of the data. Section 4 does the same for Experiment 2. Section 5 discusses the results. Appendix A analyzes some closely related constructs which were measured through questionnaires in our experiments and discards them as possible alternative explanations. Appendix B contains the experimental instructions.

2 Related Literature

Our results are in agreement with the literature showing that preferences for egalitarian allocations, or, more generally, reducing inequality, seem not to be stable. For instance, dictator-game giving is reduced if the perception of anonymity and social distance is increased (Hoffman et al., 1994; Charness and Gneezy, 2008; Franzen and Pointner, 2012). Also, subtle content-free psychological manipulations cause radical shifts of behavior (Achtziger et al., 2015, 2016, 2018). Further, part of the motivation for reducing inequality might be related to the desire to show others that one is not selfish (Bardsley, 2008; Cappelen et al., 2013). Further, unequal payoffs are often deemed acceptable, provided some justification is given, e.g. merit, entitlement, or needs (Forsythe et al., 1994; Hoffman et al., 1994; Konow, 2003; Fershtman et al., 2012). This already points out that the balance between rewarding performance and distributing resources equitably can be shifted if the link to performance is emphasized (e.g., through merit or entitlement).

The studies reported here are also related to the extensive social-psychological literature on social comparisons, which goes back to Festinger (1954). This literature stresses the human desire to evaluate oneself. The very first hypothesis is that people have a basic need to evaluate themselves and their performance. The second hypothesis is that social comparisons are used to fulfill this desire since objective evaluations are often unavailable or difficult to obtain (see also Moore and Klein, 2008). For instance, Trope (1980) asked experimental participants to rate computer games requiring eye-hand coordination, but which differed in how accurately skill was linked to outcomes. Participants preferred games with accurate feedback, bearing the potential negative consequences for their self-esteem, over games which were not diagnostic for the participant's skill.

Recent studies in economics have also clearly shown that information on the own (relative) performance has a strong impact on motivation. Several studies have shown that the provision of feedback has a positive effect on performance in the laboratory and in the field (Hannan et al., 2008; Azmat and Iriberry, 2010; Blanes i Vidal and Nossol, 2011; Kuhnen and Tymula, 2012). Azmat and Iriberry (2016) have recently shown that information on relative performance increases effort and affects satisfaction when payoff depends on performance (piece-rate wage), but has no effect in either when pay is independent of performance (flat rate). Schoenberg and Haruvy (2012) gave participants in an experimental asset market information on either the highest- or lowest-earning trader in the market, period by period. Market prices were significantly higher when the information of the highest-performing trader was presented, and the satisfaction of the traders was higher if they had a good relative position relative to the presented information.

In agreement with developments in research on social comparisons as quoted above, research in economics has also argued that people may care about relative performance, as captured by rankings, even when rankings have no financial consequences, because of its impact on self-image (Bénabou and Tirole, 2006, 2011; Köszegi, 2006).¹ For instance, Berger et al. (2013) show that ratings on performance are more effective for increasing productivity if there is a high level of differentiation in the feedback.² However, this strand of the literature has typically concentrated on the link between information on performance and performance itself, while our aim is to establish the link between information on performance and *preferences* on distributional allocations.

Of course, the concept of curiosity (in a general sense) has received a great deal of attention in psychology. An influential curiosity-related construct is *lay epistemic theory* (Kruglanski, 1989), which describes the factors involved in the general knowledge formation process. It describes four epistemic motivations related to the need for specific or nonspecific closure, and the need to avoid them. Those determine the length of hypothesis generation and testing sequence to evaluate evidence and form a belief. Although similarities between social comparison processes (Festinger, 1954) and lay epistemic theory seem apparent, the latter does not suggest a general drive towards social comparison and weighs informational sources by their relevance, not by their similarity towards oneself.

¹Work in evolutionary game theory has shown that relative-payoff concerns, which go hand-in-glove with imitating behavior, can tilt long-run predictions away from Nash equilibria (Vega-Redondo, 1997; Alós-Ferrer and Ania, 2005) and towards more competitive ones in experimental markets. This prediction has been confirmed in several Cournot-oligopoly experiments, e.g. Huck et al. (1999) and Offerman et al. (2002). In alignment with those, Fatas et al. (2015) found that the availability of a relative-performance measure led to more competitive outcomes.

²Payment schemes based on *relative* performance can sometimes decrease productivity. In a field experiment, Bandiera et al. (2005) found that productivity under piece rates was substantially higher than under relative incentives, because workers took into account the negative externality their effort imposed on others. However, an altruistic motivation was ruled out because this effect was only present when monitoring was possible.

Another important construct is *epistemic curiosity* (Loewenstein, 1994; Litman et al., 2005), defined as “the desire for knowledge that motivates individuals to learn new ideas, eliminate information-gaps, and solve intellectual problems” (Litman, 2008, p. 1586). Epistemic curiosity is subdivided in two broad categories. The first is *interest-type curiosity*, which involves the anticipated pleasure of new discoveries. The second is *deprivation-type curiosity*, which is the need to reduce uncertainty and eliminate undesirable knowledge gaps. To check for the possible relation between these concepts and performance curiosity, we included the epistemic curiosity questionnaire by Litman and Mussel (2013) in Experiment 2 (see Appendix A).

Our explanation for the relevance of performance curiosity hinges on the value of feedback for the own self-image. Psychological research (Sedikides and Strube, 1995, 1997) has identified four cardinal *self-motives* as relevant to the development, maintenance, and modification of self-views. These are *self-enhancement* (the desire to see oneself positively), *self-verification* (the desire to confirm a preexisting view of oneself), *self-assessment* (the desire to know the truth about oneself), and *self-improvement* (the desire to improve oneself). To check the possible relation between these concepts and performance curiosity, we included the questionnaire by Gregg et al. (2011) in Experiment 2.

A study by Van de Ven et al. (2005) investigated how curiosity regarding an uncertain outcome contributes to the *endowment effect*, which causes a disparity between selling and buying prices. In an experiment involving lotteries, sellers were more curious about the outcome than buyers and the curiosity positively correlated with the minimum selling price. A second experiment involved tokens initially held by sellers whose actual value (exchange rate) was uncertain. Post-trade information about that value differed across conditions. Minimum selling prices were significantly higher for sellers who would not learn about the exchange rate of a token they sold, compared to sellers who would learn about the exchange rate independently of whether they sold it or not. This showed a willingness to pay for the satisfaction of curiosity, which added to the endowment effect.

Last, our experiments are related to the literature on gender effects in competition. In our case, the decision on how to redistribute the joint earnings of the group came after performing the task. Therefore, we do not deal with performance differences or with preferences for competition, but rather with preferential choices in distributions. Still, our work is obviously related to those strands of the literature at a conceptual level. There are well-established gender differences in the attitudes toward competition. Gneezy et al. (2003) and Niederle and Vesterlund (2007) showed that women tend to shy away from competition in real-effort tasks, and Gneezy et al. (2009) showed that matrilineal and patriarchal societies differ with regard to competitive behavior.³ A recent

³Our design bears procedural similarities to Task 4 in Niederle and Vesterlund (2007) in the sense that participants perform the task without knowing the incentive rule and choose the compensation scheme later. Niederle and Vesterlund (2007) offered a choice between piece-rates or competitive tournaments, while our design involves neither, but rather different distributions of the joint income generated by the group.

field experiment by Azmat et al. (2016) has shown that women perform worse than men when stakes are high, while the opposite is true when the stakes are low.

3 Experiment 1

3.1 Design and Procedures

The experiment was conducted at the Cologne Laboratory for Economic Research and programmed in *z-Tree* (Fischbacher, 2007).⁴ Participants were recruited from the University of Cologne’s pool, excluding psychology students, using ORSEE (Greiner, 2015). The experiment was a between-subject design with 180 participants, 45 females and 45 males in each of the two treatments. Average payoff was EUR 13.25 (USD 17.80) for a session that lasted around 60 minutes. Subjects were assigned to groups of five players, but they did not know the identities of the other group members.⁵ Further, both the task and the subsequent allocation decision were made individually and there was no interaction of any kind except in the determination of payoffs at the end of the experiment. Initially, subjects worked individually in a real-effort task requiring to add up sets of five two-digit numbers (e.g. $30+87+19+16+38=\square$) for a predetermined time. Each correct answer generated 10 points, with an exchange rate of 5 Euro cents per point, but no feedback on the correctness of the responses was provided until the end of the experiment. Subjects were not allowed to use calculators or other electronic devices, but they were allowed to use scratch paper. This real-effort task was proposed in Niederle and Vesterlund (2007) and has been often used in experiments (e.g., Azmat and Iriberry, 2016).⁶

The real-effort task lasted for eight minutes, which is relatively long. This duration was selected to put enough weight on the effort side rather than on the skills of the subjects. Also, a shorter task duration would have increased the probability of ties inside each group. After performing the task, the experiment moved to the decision-making part without any feedback being provided.

Participants were asked to make a decision on how they would prefer to distribute the joint amount of points generated by the group. They were informed that one of the group members would be selected at random and his or her decision would be implemented as stated.⁷ Two distributions were available. The first was an *equal split*, where all group

⁴The data and the codebooks for both experiments can be downloaded from <http://osf.io/5hkj6>.

⁵In particular, they had no information on the gender composition of their group. This is particularly important, because gender composition might affect decision making and performance within a group (Gneezy et al., 2003; Apesteguía et al., 2012).

⁶Women often think that their performance in math will be worse than that of men, which sometimes leads to actually worse performance due to “stereotype threat” (Spencer et al., 1999; Gneezy et al., 2003). However, a meta-analysis of 100 studies on gender differences in math performance showed that there is no gender difference in arithmetic or algebraic performance (Hyde et al., 1990).

⁷Subjects were informed that they were part of a group at the beginning of the experiment, and they were aware that after finishing the real-effort task, they would decide among two distribution rules for sharing the joint surplus, with one decision implemented at random. However, they were not informed about the specifics of the possible distribution rules until the decision had to be made.

Table 1: Treatment Overview: Distribution and Information Revelation

Control Treatment				No Info Treatment			
Equal Split		Performance-Based		Equal Split		Performance-Based	
Rank	Share (%)	Rank	Share (%)	Rank	Share (%)	Rank	Share (%)
1	20	1	40	-	20	1	40
2	20	2	30	-	20	2	30
3	20	3	15	-	20	3	15
4	20	4	10	-	20	4	10
5	20	5	5	-	20	5	5

members earned the same amount of points. The second was a *performance-based split*, leading to a different amount of points depending on the relative rank within the group (see Table 1). Specifically, the performance-based split made the three worst-placed participants earn less than under the equal split, in order to create a clear tradeoff between monetary rewards and feedback provision for the median participant.⁸

The key manipulation between treatments was the amount of information that was revealed with each possible distribution. In the *Control Treatment* participants chose one of the two distributions, but later the ranking and the number of correctly solved calculations was revealed independently of which distribution was finally implemented. However, in the *No Info Treatment* participants were told that the ranking and the number of correctly solved calculations would only be revealed if the performance-based split was actually the one chosen and implemented. The alternatives are detailed in Table 1.

In the absence of self-image concerns, the difference across treatments should not affect behavior. In the presence of performance curiosity, however, we should observe a higher proportion of performance-based choices in the No Info treatment than in the control treatment.

To control for various other possible explanations for differences in behavior across treatments we also elicited *self-efficacy* before and after the real-effort task. That is, participants were asked how many additions they thought they could solve correctly in eight minutes. Further, they were asked to report their expected relative performance in a 7 seven-point scale ranging from “Far below average” to “Far above average.” To control for individual differences in the regression analysis in attitudes and motivation, we also elicited risk aversion attitudes using the lottery questionnaire of Holt and Laury (2002) and the reduced Achievement Motive Scale of Lang and Fries (2006), using the German version of Dahme et al. (1993) (the analysis of those is relegated to Appendix

⁸The performance-based distribution allocates 15%, 10%, and 5% of the joint proceeds to the 3rd, 4th, and 5th group member, respectively. Hence choosing the performance-based distribution entails a clear monetary cost for participants who expect to be average or worse. The participants ranked first and second are rewarded with 40% and 30% of the joint proceeds, that is, by a factor of 2 and 1.5, respectively, compared to the equal-split distribution. This makes clear that the distribution indeed rewards performance.

A). At the end of the experiment participants received feedback (if any) on a screen where their rank and the number of correctly solved additions were highlighted.

3.2 Distribution choices

The percentage of performance-based choices in both distributions is illustrated in Figure 1(a). In the control treatment, there is no tradeoff since the same information is revealed independently of which distribution is implemented. Hence, this treatment is a pure test of inequality aversion. In agreement with standard results in the social preferences literature, only 35 of the 90 participants (38.89%) in this treatment chose the performance-based distribution. In contrast, this proportion rose to 70.00% in the No Info treatment (63 out of 90). The difference is highly significant according to a test of proportions ($z = -4.191$, $p < 0.0001$).⁹ Thus, subjects chose the performance-based distribution in the No Info treatment more often than in the control treatment. Since the only difference between treatments is the information that is revealed when the equal split is implemented, the difference in results is due to the tradeoff regarding information on performance. We hence conclude that performance curiosity overcomes inequality aversion in our data.

Since we expected a gender effect, we analyzed the behavior of the women and men separately. Figure 1(b) shows the proportions of participants choosing the performance-based split across treatments by gender. The overall effect observed in Figure 1(a) seems to be driven by female behavior. While only 10 women out of 45 (22.22%) chose the performance-based split in the control treatment, the proportion rose to 68.89% (31 out of 45) in the No Info treatment (test of proportions, $z = -4.445$, $p < 0.0001$). The difference goes in the same direction for men, with 55.56% of performance-based choices in the control treatment vs. 71.11% in the No Info treatment, but the difference misses significance ($z = -1.531$, $p = 0.1257$). As a result, while females chose the equal split more often than males in control treatment (test of proportions, $z = 3.2434$, $p = 0.0012$), the difference disappeared in the No Info treatment ($z = 0.2300$, $p = 0.8181$). We delay the analysis of the interaction between treatment and gender to the regression analysis below (Table 2, Model 2; Section 3.4).

That is, while men choose the performance-based distribution more often, as soon as a tradeoff involving performance curiosity is introduced, women behave exactly as men.

3.3 Expected Performance

There is one obvious, rational reason for some participants to choose the performance-based split. If the participant believes that he or she is likely to be strictly above the group's average, the individual payoff under the performance-based split is higher than under the egalitarian allocation. Reciprocally, if the participant believes to be at or

⁹A Fisher's exact test delivers the same conclusion.

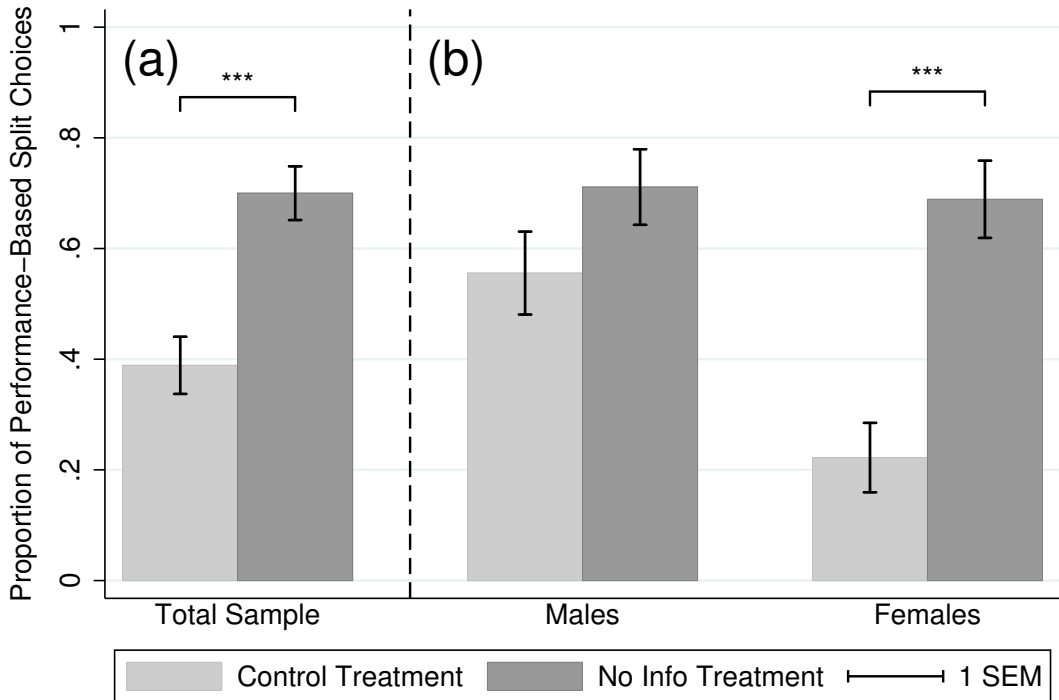


Figure 1: Experiment 1, Proportion of Performance-Based Splits. (a) Proportion of participants choosing the performance-based split across treatments. (b) Proportion of participants choosing the performance-based split by treatment and gender. *** $p < 0.01$, test of proportions

below average, the individual payoff is higher under the egalitarian allocation (recall Table 1). Hence, we need to control for expected performance.

The median number of correctly solved additions was 15 (ranging from 5 to 43), and the median error rate was 0.19 (ranging from 0.03 to 0.69).¹⁰ Crucially, participants were asked their expected relative performance (after completing the real-effort task) in a seven-point scale. The expected relative performance did not differ across treatments according to a Mann-Whitney-Wilcoxon (MWW) test (control treatment, mean 4.089; No Info treatment, 4.300; $z = -1.055$, $p = 0.2913$). We classified participants into three groups according to their beliefs, i.e. below, exactly, and above average.¹¹

Figure 2(a) disentangles the proportions of performance-based choices within the three expectation groups. The performance-based split was chosen less often in the

¹⁰After completing the real-effort task but before making the distribution choice, participants were asked about their expected performance, that is, how many additions they thought they had solved correctly. The average expectation was 15.2 additions in the control treatment and 17.3 in the No Info treatment. The difference was not significant according to a Mann-Whitney-Wilcoxon test ($z = -1.457$, $p = 0.1452$).

¹¹For simplicity, the question on expected relative performance did not refer to a group. Since allocation to groups was random, the participants' expectations translate into expected ranking within the group.

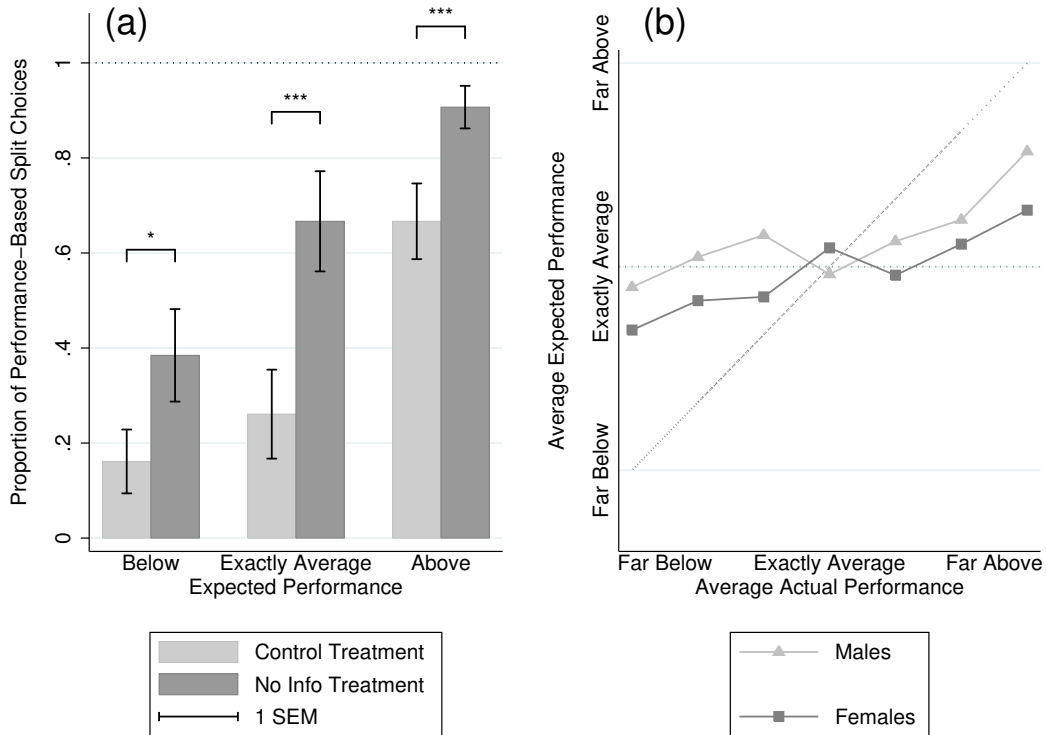


Figure 2: Experiment 1, Proportion of Performance-Based Splits and Expected Performance. (a) Proportion of participants choosing the performance-based split by treatment and expected ranking. (b) Expected versus actual performance by gender

control treatment than in the No Info treatment across all three groups. For participants who expected to be above average, the performance-based split was chosen 24 out of 36 times (66.67%) in the control treatment, compared to 39 out of 43 (90.70%) in the No Info treatment (test of proportions, $z = -2.647$, $p = 0.0081$). For participants who expected to perform exactly on average, the performance-based split was chosen 6 out of 23 times (26.09%) in the control treatment, compared to 14 out of 21 (66.67%) in the No Info treatment ($z = -2.700$, $p = 0.0069$). For participants who expected to be below average, the performance-based split was also chosen less often in the control treatment (16.13%; 5 of 31) than in the No Info treatment (38.46%; 10 of 26), but the difference narrowly missed significance ($z = -1.907$, $p = 0.0565$). Additionally, comparing the proportions of choices across groups, one sees that, as was to be expected, a higher expected performance resulted in a higher proportion of performance-based choices (see regression analysis below).¹²

We did not find any difference in beliefs across treatments. However, women believed that they had performed worse (average 3.800 in the 7-point scale), compared to men's expectations (average 4.589 in the 7-point scale). The difference is highly significant

¹²A separate probit regression showed no significant interaction between treatment and expected performance.

Table 2: Experiment 1, Probit Regressions on Performance-based Choices

	Model 1	Model 2	Model 3
No Info Treatment	0.807*** (0.1930)	0.417 (0.2725)	0.679** (0.3031)
Female		-0.904*** (0.2801)	-0.395 (0.3137)
No Info X Female		0.840** (0.3946)	0.440 (0.4350)
Expected Performance in Septiles			0.529*** (0.0930)
Constant	-0.282** (0.1340)	0.140 (0.1875)	-0.224 (0.2126)
LogLikelihood	-115.120	-109.701	-90.676
Wald Test	17.458***	26.590***	50.785***
Linear combination tests:			
No Info + No Info X Female		1.257*** (0.2854)	1.119*** (0.3102)
Female + No Info X Female		-0.064 (0.278)	0.045 (0.303)

Notes. Standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

(MWW test, $z = 3.924$, $p = 0.0001$). Figure 2(b) plots the expected performance (average report in the 7-point Likert scale) for males and females, conditional on the actual population septiles in the experiment, as computed from the number of correctly solved additions. This figure makes apparent that females tend to underestimate their performance compared to males, as previously pointed out, e.g., by Gneezy et al. (2003).¹³

3.4 Regression Analysis

Recapitulating, in the previous subsections we have seen that performance curiosity shifts choices toward the performance-based split and away from the egalitarian allocation, and that this effect is far stronger for women. However, we have also seen that choices are influenced by the expected ranking, and that there are differences in beliefs across genders. In view of these results, our next step is to conduct regression analyses controlling for expected ranking. This will also allow us to investigate the determinants of the gender effect.

We hence turn to probit regressions on the choice of the performance-based split accounting for treatment, gender, and expected ranking. We report on three different model specifications in Table 2.

¹³Although the task has been determined to be gender-neutral (see Niederle and Vesterlund, 2007), in our sample women actually performed worse than men, with an average 18.4 correct additions for males and 15.0 for females (MWW test, $z = 3.349$, $p = 0.0008$). Causality, of course, might go in the other direction because of stereotype threat (recall Footnote 6).

Model 1 investigates the basic treatment effect without further controls. The treatment dummy is highly significant and positive, reproducing in the regression the basic message of the test of proportions in Section 3.2, as illustrated in Figure 1(a). Model 2 adds gender and its interaction with the treatment dummy. The reference group is hence made of males in the control treatment. The treatment dummy becomes non-significant, reflecting that the treatment effect misses significance for males as found in Section 3.2 (recall Figure 1(b)). In contrast, the interaction of the treatment dummy and the female dummy is significant, showing that women react to the manipulation more strongly than men. A post-hoc linear combination test (No Info + No Info \times Female, bottom of Table 2) reveals that the treatment effect for women is highly significant ($p < 0.0001$). The female dummy is negative and highly significant, showing that women choose the egalitarian allocation more often than men in the control treatment. However, this gender difference does not exist in the No Info treatment, as shown by the corresponding linear combination test (Female + No Info \times Female, bottom of Table 2; $p = 0.8181$).

Model 3 is the key regression, which introduces the expected ranking of participants and sheds light into the determinants of the gender differences. First, the independent variable “Expected Performance in Septiles”, centered at 0, is highly significant and positive, confirming that the higher the expectations the more likely it is that the participant chooses the performance-based split, as higher expectations make the latter split more attractive. Once we control for this natural effect, the treatment dummy becomes highly significant and positive, indicating that performance curiosity is a significant driver of behavior for males. The linear combination test confirms that the treatment effect is also highly significant for females. That is, after controlling for the expectations we see that both male and female participants are more likely to choose the performance-based split when inequality aversion conflicts with performance curiosity.

In contrast, in Model 3 the female dummy and its interaction with the treatment dummy are not significant anymore. A linear combination test also shows that there is no gender difference in the No Info treatment ($p = 0.8822$). That is, the entire gender effect observed in previous regressions and tests originates exclusively in differences in expectations. In other words, since females have lower expectations, the performance-based split is less attractive, and hence chosen less often by females than by males. Once we control for expectations, we see that the effects of performance curiosity are present for both genders.

4 Experiment 2

The purpose of our second experiment was threefold. First, the experiment was conceived as a replication in order to establish the reliability of the effects. Accordingly, the first two treatments of Experiment 2 were identical to those of Experiment 1. Second, we added a third treatment to the design, the *Only Winner Treatment*, in order to test whether revealing only partial information, namely whether one is the winner or

not, is enough to produce the same effect we found in Experiment 1. Human beings have intrinsic preferences for winning, perceived as a reward in itself. That is, they choose to exert effort to win (Fershtman et al., 2012) regardless of the payoff distribution resulting from their eventual victory. Hence, one could speculate that the effects we uncovered in Experiment 1 are actually derived from this motivation only, that is, that the desire to know whether one is the winner is the actual driver behind the effect. In the new treatment, the information regarding whether one is the winner or not is revealed independently of the distribution choice. Hence, if the desire to know whether one is the winner were enough to satisfy the need for information about performance, in the new treatment the effect should disappear.

Third, we added additional controls (personality scales) to the design in order to test for the robustness of the effects with respect to individual differences. The new controls were of two kinds, related, first, to general curiosity, and, second, to self-motives. Specifically, we focused on epistemic curiosity (Loewenstein, 1994; Litman et al., 2005), using the epistemic curiosity questionnaire by Litman and Mussel (2013), and self-motives (Sedikides and Strube, 1995, 1997), using the questionnaire by Gregg et al. (2011). Both constructs are described in Section 2 and their analysis is relegated to Appendix A.

4.1 Design and Procedures

The experiment was a between-subject design with three treatments, again conducted at the Cologne Laboratory for Economic Research, and programmed in z-Tree. We recruited 270 participants through ORSEE from the University of Cologne’s pool, excluding psychology students. The sample was perfectly balanced by gender, leaving 45 females and 45 males for each of the three treatments. The average payoff was EUR 10.56 (USD 14.21) for a session that lasted around 55 minutes.

The experiment differed from Experiment 1 in two respects. First, since after Experiment 1 we concluded that neither risk attitudes nor achievement motivation were determinant for the treatment effect, we replaced those questionnaires by the ones measuring epistemic curiosity and self-motives as explained above. Since questionnaires were placed at the end of the experiment, the change could not affect behavior and the experiments remain fully comparable. Second, we added a third treatment, but the other two treatments were identical to the control and No Info treatments of Experiment 1 (recall Table 1), hence the new experiment contains a pure replication.

The third treatment is the *Only Winner Treatment*. In this treatment, the egalitarian and performance-based allocations are identical to those of other treatments, but the associated information is different. If the performance-based split is chosen and implemented, as in the other two treatments all information regarding how many additions have been correctly solved by each member of the group and the resulting ranking is revealed. The difference is that, if the equal split is the one chosen and implemented,

Table 3: Only Winner Treatment in Experiment 2

Only Winner Treatment			
Equal Split		Performance-Based	
Rank	Share (%)	Rank	Share (%)
1	20	1	40
-	20	2	30
-	20	3	15
-	20	4	10
-	20	5	5

then the only information revealed is how many additions were correctly solved *by the winner*, plus a signal indicating whether the participant is the winner or not. See Table 3 for a summary of the treatment.

The rest of Experiment 2 was identical to Experiment 1 regarding the real-effort task, the procedure for implementing the chosen allocation, the order of tasks and decisions, self-efficacy elicitation, and the final possibility to privately receive feedback.

4.2 Distribution Choices

Figure 3 shows the proportions of participants choosing the performance-based split across the three treatments, for the whole sample (a) and split by gender (b). We successfully replicated the main result of Experiment 1. There were significantly more performance-based choices in the No Info treatment (75.56% of 90) than in the control treatment (43.33% of 90) (test of proportions, $z = -4.402$, $p_{\text{adj}} < 0.0001$, adjusted for multiple comparisons following the Holm-Bonferroni method). Further, the proportion of performance-based choices in the Only Winner treatment (60.00% of 90) was significantly smaller than in the No Info treatment ($z = 2.233$, $p_{\text{adj}} = 0.0256$), and larger than in the control treatment ($z = -2.237$, $p_{\text{adj}} = 0.0505$), although the latter difference narrowly misses significance. That is, the new treatment lies clearly “in the middle” between the two previous ones.

Females chose the performance-based split less often than males in the control treatment (females, 33.33%; males, 53.33%; $z = 1.915$, $p = 0.0556$), although the difference narrowly misses significance, whereas there were no gender differences in the No Info treatment (females, 71.11%; males, 80.00%; $z = 0.981$, $p = 0.3265$) or in the Only Winner treatment (females, 53.33%; males, 66.67%; $z = 1.291$, $p = 0.1967$). That is, the apparently higher degree of inequality aversion of women in the control treatment disappears in both of the treatments with an informational tradeoff.

As in Experiment 1, when looking at females only, we also observe significantly more performance-based choices in the No Info treatment (71.11%) than in the control treatment (33.33%) (test of proportions, $z = -3.588$, $p_{\text{adj}} = 0.0010$). Further, the effect is also significant for males (No Info treatment, 80.00%; control treatment, 53.33%;

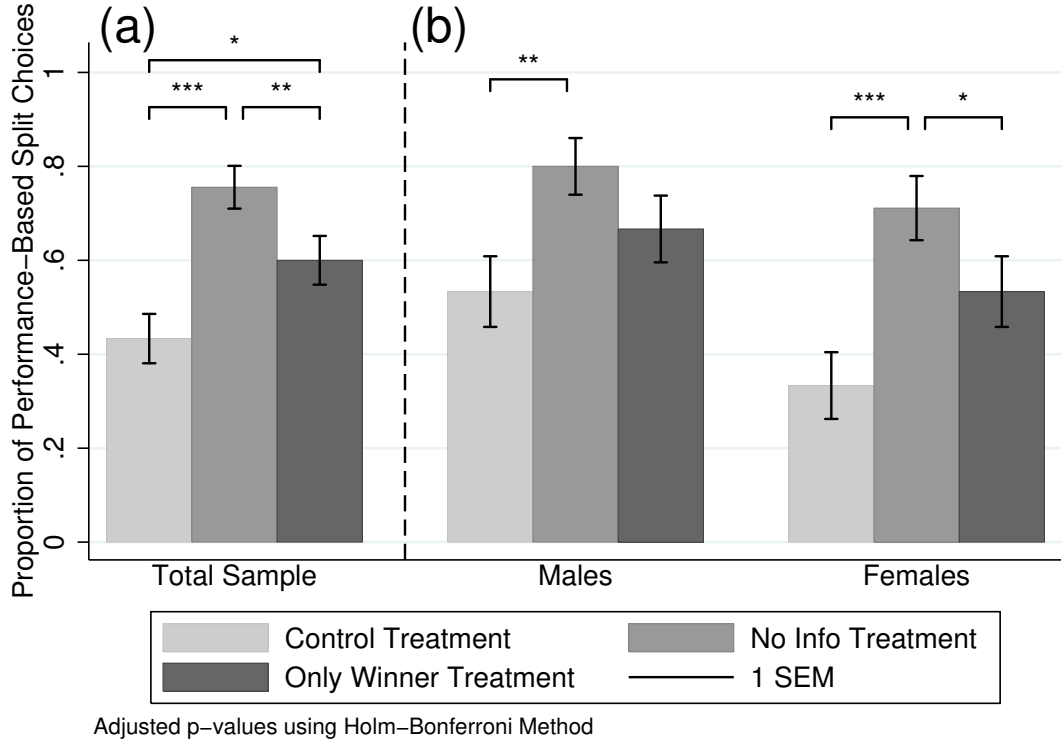


Figure 3: Experiment 2, Proportion of Performance-Based Splits. (a) Proportion of participants choosing the performance-based split across treatments. (b) Proportion of participants choosing the performance-based split by treatment and gender. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

$z = -2.683$, $p_{\text{adj}} = 0.0219$). That is, although the effect for males missed significance in Experiment 1, it is significant in the replication, and we conclude that both men and women chose the performance-based split significantly more often when information was not revealed in the equal split. Looking at females only, the proportion of performance-based choices in the Only Winner treatment (53.33%) again lies between the control and the No Info treatments, although the effects miss significance (No Info vs. Only Winner, $z = 1.739$, $p_{\text{adj}} = 0.0820$; control vs. Only Winner, $z = -1.915$, $p_{\text{adj}} = 0.1111$). The proportion of performance-based choices for males in the Only Winner treatment (66.7%) was also between those of the other treatments, but the effects are not significant (No Info vs. Only Winner, $z = 1.430$, $p_{\text{adj}} = 0.3053$; control vs. Only Winner, $z = -1.291$, $p_{\text{adj}} = 0.1967$).¹⁴

If partial information, namely whether one was or not the winner, was enough to satisfy performance curiosity, the effect seen in Experiment 1 should disappear in the Only Winner treatment, because the information regarding the winner was available in both allocations. Quite to the contrary, we observe that the effect is already present in the Only Winner treatment. Hence, we discard that revealing partial information

¹⁴The interaction between gender and treatments was not significant in an additional probit regression.

about the winner is enough to satisfy performance curiosity. The general picture that arises so far is that Experiment 2 successfully replicated the results of Experiment 1, but additionally the new treatment, where the information tradeoff also exists but is less severe than in the No Info treatment (because less information is hidden under the equal split) is simply midway between the control treatment and the No Info one. That is, the stronger the informational tradeoff, the stronger the influence of performance curiosity, leading the participants to choose the fully informative performance-based split.

4.3 Expected Performance

The median number of correctly solved additions in Experiment 2 was 15 (ranging from 3 to 36), and the median error rate was 0.15 (ranging from 0 to 0.73).¹⁵ As in Experiment 1, we control for the individual beliefs regarding the expected ranking, elicited after the completion of the real-effort task but before the distribution choice. Figure 4(a) illustrates the expected own ranking of the participants. In contrast to Experiment 1, there were some minor but significant differences in expectations across treatments. Specifically, the average expected septile for participants in the No Info treatment (4.378) was significantly higher than that of the participants in the Only Winner treatment (3.889; $z = 2.487$, $p_{\text{adj}} = 0.0387$), and also higher than that of the participants in the control treatment (3.978; MWW test, $z = 2.100$, $p_{\text{adj}} = 0.0714$) although the difference missed significance. There were no significant differences in expectations between the control and Only Winner treatments ($z = 0.430$, $p_{\text{adj}} = 0.6674$).

As in Experiment 1, the proportion of performance-based choices was significantly higher in the No Info treatment than in the control treatment, both for participants who expected to perform above average (control, 73.53% of $N = 34$; No Info, 97.67% of $N = 43$; test of proportions, $z = -3.130$, $p = 0.0017$) and for those who expected to perform on average (control, 23.81% of $N = 21$; No Info, 72.00% of $N = 25$; $z = -3.256$, $p = 0.0011$). The difference for participants who expected to perform below the average was not significant (control, 25.71% of $N = 35$; No Info, 36.36% of $N = 22$; $z = -0.856$, $p = 0.3922$). However, we do not observe “last-place aversion” (Kuziemko et al., 2014) in the sense that for those subjects the equal split distribution is not chosen significantly more often in the No Info treatment, where they could hide their expected poor placement.

As commented above, the proportion of performance-based choices in the Only Winner treatment lies between the proportion of performance-based split in the control and No Info treatments. Of course, this points to weaker effect sizes when comparing with

¹⁵As in Experiment 1, participants were asked how many additions they thought they had solved correctly (after completing the real-effort task but before making the distribution choice). The average expectation was 16.2 additions in the control treatment, 17.3 additions in the No Info treatment, and 15.0 additions in the Only Winner treatment. The differences were generally not significant according Mann-Whitney-Wilcoxon tests, adjusted for multiple testing (control vs. No Info, $z = -1.177$, $p_{\text{adj}} = 0.2393$; control vs. Only Winner, $z = 1.344$, $p_{\text{adj}} = 0.3577$; No Info vs. Only Winner, $z = 2.217$, $p_{\text{adj}} = 0.0799$).

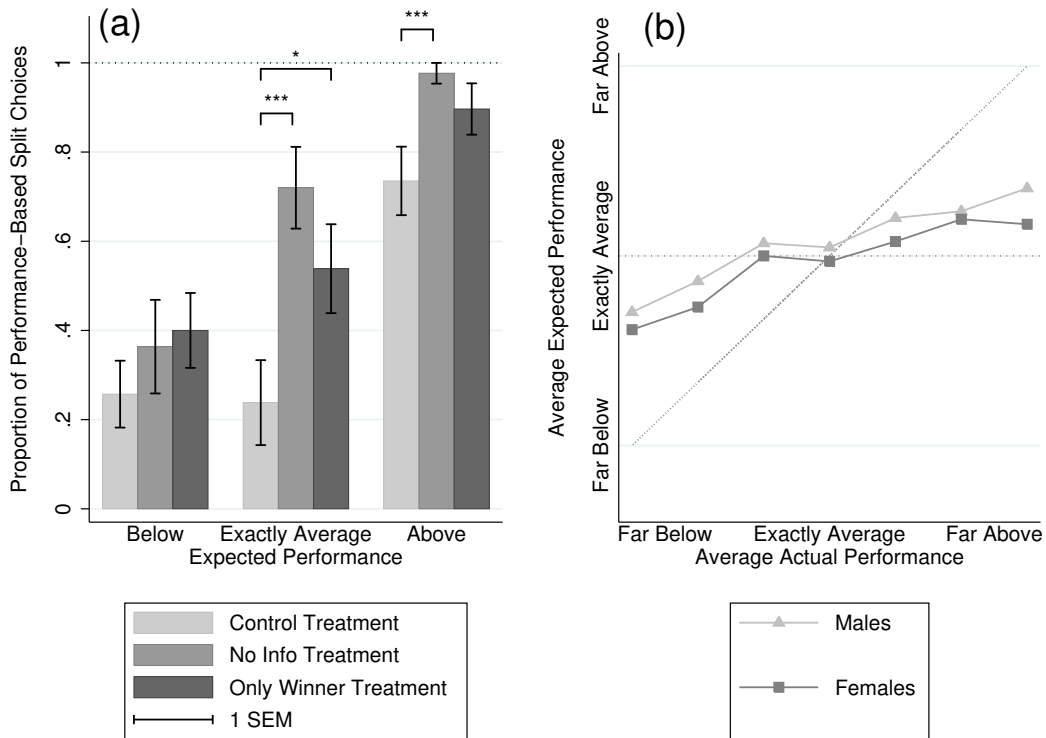


Figure 4: Experiment 2, Proportion of Performance-Based Splits and Expected Performance. (a) Proportion of participants choosing the performance-based split by treatment and belief. (b) Expected versus actual performance by gender

either of the other two treatments. As a consequence, when splitting the data into the three expected performance ranges, the differences are in general not significant.

Figure 4(b) plots the expected performance (average report in the 7-point Likert scale) for males and females, conditional on the actual population septiles in the experiment, as computed from the number of correctly solved additions (compare with Figure 2(b)). As in Experiment 1, we observe that females tend to underestimate performance compared to males. This is confirmed by a Mann-Whitney-Wilcoxon test on the whole sample (males, $N = 135$, average expected septile 4.304; females, $N = 135$, average 3.859; $z = 2.843$, $p = 0.0045$).¹⁶

4.4 Regression Analysis

We now turn to a regression analysis to confirm the results of our non-parametric tests while controlling for expected ranking and for the various other individual variables measured in the experiment. In view of the results above, instead of treatment dummies we introduce two dummies representing what kind of information is revealed. The first dummy, *Hidden Info*, captures all cases when some information is hidden in the equal

¹⁶As in Experiment 1, women actually performed worse than men, with an average 17.2 correct additions for males and 15.1 for females (MWW test, $z = 2.294$, $p = 0.0218$). Recall Footnote 13.

Table 4: Experiment 2, Probit Regressions on Performance-based Choices

	Model 1	Model 2	Model 3
Hidden Info	0.860*** (0.1961)	0.875*** (0.1978)	0.874*** (0.2208)
Only First	-0.439** (0.1967)	-0.446** (0.1980)	-0.305 (0.2216)
Female		-0.387** (0.1592)	-0.219 (0.1759)
Expected Performance in Septiles			0.595*** (0.0809)
Constant	-0.168 (0.1328)	0.0220 (0.1547)	-0.0634 (0.1694)
LogLikelihood	-172.205	-169.224	-136.021
Wald Test	19.261***	24.540***	68.681***
Linear Combination Test Hidden Info + Only First	0.421** (0.1884)	0.430** (0.1899)	0.569*** (0.2077)

Notes. Standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

split, that is, both the No Info and the Only Winner treatment. The second dummy, *Only First*, further differentiates the cases where only information regarding the winner is revealed when the equal split is implemented, that is, identifies the Only Winner treatment. This way we can better analyze the differential effects of the informational treatments and emphasize the comparisons between the Only Winner and the No Info treatments (captured directly by the Only First dummy) and between the No Info and the control treatments (captured by the Hidden Info dummy).

Model 1 (see Table 4) captures the basic treatment effects without additional controls and reproduces the insights from the non-parametric tests as illustrated in Figure 3(a). The Hidden Information dummy is positive and highly significant, showing that choosing the performance-based split was more likely in the No Info treatment than in the control treatment. The Only First dummy is significant and negative, showing that performance-based choices were less likely in the Only Winner treatment than in the No Info treatment. A linear combination test (bottom of Table 4) shows that performance-based choices were more likely in the Only Winner treatment than in the control treatment.

In Model 2, we further control for gender. Females are less likely to choose the performance-based split than men, as reflected by the significantly negative female dummy. Adding interactions between gender and treatments does not change the results qualitatively. Otherwise, the results of Model 1 are not affected by controlling for gender.

Model 3 incorporates the expected performance (in septiles and centered at 0), which, as in Experiment 1, is shown to have a positive and highly significant effect on the likelihood of choosing the performance-based split. That is, subjects are naturally more

likely to choose the performance-based split when they expect to be highly ranked. As expected, the female dummy becomes insignificant controlling for expectations, confirming that the apparent gender effect is fully driven by the fact that females have lower expectations than males. The key treatment effect remains significant, showing that even when controlling for expectations the performance-based choice is more likely in the No Info treatment than in the control treatment (as in Experiment 1). However, the difference between the No Info and the Only Winner treatments becomes insignificant ($p = 0.1686$) when we control for expectations. Accordingly, the difference between the control and the Only Winner treatments becomes *more* pronounced and highly significant ($p = 0.0061$), suggesting that, after controlling for expectations, both the Only Winner and the No Info treatments actually have similar effects. Recall that if partial information were already enough to satisfy performance curiosity, we would expect no effect at all in the Only Winner treatment, as this information is available independently of the distribution choice. Hence, it seems that the driver of behavior is the actual curiosity to know the own performance, independently of whether one is the winner or not.

5 Discussion

We have shown that performance curiosity, a form of intrinsic motivation related to self-image, can overcome preferences for egalitarian allocations of jointly generated proceeds. In our experiments, a subtle manipulation in the structure of information regarding how well each participant performed in a real-effort task was enough to produce large shifts away from equal-split choices and toward performance-based schemes. This effect subsists even when one controls for expected relative performance, which of course correlates with the choice of the performance-based allocation, as the latter maximizes expected income if one expects to perform better than average.

As we expected, the effect of performance curiosity appears to be much stronger for females than for males, with the consequence that the manipulation greatly helps reduce gender differences in egalitarian choices. The gender difference in our results is fully explained by differences in performance beliefs. Apparent gender differences regarding the proportion of equal-split choices (and hence, inequality aversion) are caused by differences in beliefs regarding the own performance, namely the fact that, compared to women, men are typically more overconfident. The difference in beliefs creates an apparent gender effect in the control treatment, with females choosing the egalitarian allocation more often than males. This difference (and the effect of beliefs) disappears in the presence of performance curiosity. Our analysis hence identifies a factor which helps overcome gender differences arising from relative underconfidence.

Appendix A analyzes and discards a number of alternative explanations for our results, on the basis of supplementary data (questionnaires) collected in the two experiments. Some other alternative interpretations of our results and possible extensions

are as follows. First, it is well-known that different factors as entitlement, merits, etc, can be used as justifications in order to depart from “fair” allocations and switch to more favorable ones (in terms of income). It could be argued that the presence of additional information attached to the performance-based split could have been used as precisely such a justification even if the participants did not care about the information itself. However, this explanation is unlikely because our data shows a larger proportion of performance-based distribution choices in the No Info treatment than in the control treatment even for subjects who expected to perform exactly on or below average. For those participants, by design the income-maximizing decision would have been the equal split. Hence, the additional information has no value as an excuse to pursue an income-maximizing allocation.

Second, a possible alternative interpretation would be that the effect is simply due to the fact that, in the No info treatment, the performance-based distribution comes with an additional informational attribute than the equal split distribution. This explanation would imply that general curiosity is far stronger than assumed, in the sense that the particular content of the additional information is not relevant. By its very nature, a performance-based split will always reveal some information on performance, hence it is not possible to shut down that informational attribute and replace it with an alternative, less informative one. However, the explanation could be tested with an additional design where the equal split is endowed with some non-informative attribute (with respect to performance; e.g., the weather forecast of a remote country) but the performance-based split is still endowed with performance information. Hence, the number of informational attributes would be kept constant, and any difference would be due to the actual nature of the information. This design goes beyond the scope of the present work.

A related point is that in the No Info treatment it was made salient that information on performance was being left out, which might have increased the focus on performance. This is unavoidable, since the point of our design is to make the tradeoff between information and an equal-split allocation apparent. However, the real-effort task was run before the actual choices were presented. There was no information whatsoever about the fact that performance information might or might not be revealed depending on later choices (in the No Info treatment, the initial instructions mentioned that choices might differ in attached information, but the nature of such information was not explained until after the real-effort task). That is, the fact that there was a tradeoff with respect to information on performance was only apparent right before the actual choice of distribution, and hence any possible salience effects would have been kept to a minimum. Also, we remark that, as shown in the regression analysis, the effect subsists when controlling for expected performance.

Some extensions of the present research would be natural. We mention here just two of them. First, Loewenstein (1994) argued that curiosity may be an impulsive hedonic drive that easily wanes after being satisfied. If this is the case, prosocial decision makers might regret having chosen feedback over equality, leading to different decisions in the

future. Since our objective was to show that performance curiosity can trump inequality aversion, we adopted a purely one-shot design, and hence we cannot test this additional possibility. It would be possible to expand the design incorporating repeated decisions in order to test for this hypothesis, but if performance was revealed between decisions, it would be necessary to appropriately control for it.

A second natural avenue for further research would be to elicit the actual willingness to pay for information on performance. In view of well-known behavioral phenomena pointing at discrepancies between valuations and actual choices and the noisy character of evaluations (e.g., Delquié, 1993; Alós-Ferrer et al., 2016), we aimed to establish that the postulated effect had consequences on actual behavior (choice data), but once this effect has been established, evaluating the willingness to pay for information is a natural second step. Although we leave this task for future research, we can offer some preliminary evidence because, as mentioned above, participants who believed to be exactly on average or below would have maximized expected payoffs by choosing the equal-split allocation, but did not choose that allocation more often. Hence they actually revealed the willingness to incur monetary costs in order to obtain the information.

We view our research as a contribution to the general investigation into the causes of and motivations behind preferences among different distributional allocations, especially in frameworks where the resources that are distributed can be traced back to individual contributions. This is potentially important for reward schemes in firms and organizations, and for a deeper understanding of attitudes with respect to fairness and redistribution at the societal level. Our results, however, can be read in two different ways. If egalitarian allocations are viewed as fair and hence as a worthy policy objective, the results point out that support for egalitarian principles might be diminished in society if information on individual performance is restricted, for instance in order to protect privacy or to avoid making interpersonal comparisons prominent. While people might be willing to accept egalitarian redistributions independently of whether they profit from them themselves or not, they still have a strong preference for receiving information on their relative individual contribution. On the other hand, if rewarding performance through incentives is viewed as fair and desirable, linking information to rewards might result in increased social support of the corresponding distribution schemes.

Appendix A: Robustness Analysis

A.1 Experiment 1

In this subsection, we report on a few additional controls which help discard possible alternative interpretations. First, it is natural to speculate that some participants might choose the egalitarian distribution in the No Info treatment in order to protect their self-image, especially if they fear being below average, because in this way they avoid being exposed to negative feedback. However, this would create a tendency leading to

less performance-based splits, which we do not observe, even among participants who expected to be below average. In any case, we anticipated this alternative explanation and we introduced a final question for participants in the No Info treatment, after all other decisions have been made but before the distribution was implemented. In this question, we asked them whether they wanted to be privately informed about their own performance (both absolute and relative, that is, number of correctly solved additions and ranking within the group), independently of which distribution was finally implemented. Recall that even if a participant chose the performance-based split, it was not guaranteed that the information would be revealed, since the random mechanism could select a different participant, which might have chosen the egalitarian allocation instead. Hence the additional question made sense to participants. All participants but one (that is, 89 out of 90) chose to see their own performance, indicating that the possible desire to protect their self-image against feedback was not a factor.

Second, we measured risk attitudes by means of the lottery questionnaire of Holt and Laury (2002). There were no differences in risk attitudes by gender (males: mean relative risk aversion parameter 0.497; females: mean 0.558; MWW test, $z = -0.405$, $p = 0.6851$), treatment (control treatment: mean 0.526; No Info treatment: mean 0.530; $z = -0.203$, $p = 0.8388$), or choice (egalitarian choice, $N = 82$, mean 0.530; performance-based choice, $N = 98$, mean 0.526; $z = -0.062$, $p = 0.9508$).¹⁷

Third, we also used the reduced Achievement Motive Scale of Lang and Fries (2006) to elicit achievement motives. That scale is divided in questions eliciting two components, aptly named “hope of success” and “fear of failure.” Each subscale is made out of 5 items on a 4-point Likert scale, and the (added) scores of the subscale range from 4 to 20. We found the gender differences typically described in the literature, namely that men score significantly higher on the hope of success scale (males: mean 17.122; females: mean 15.944; MWW test, $z = 3.518$, $p = 0.0004$) and women score significantly higher on the fear of failure scale (males: mean 10.344; females: mean 12.400; $z = -4.346$, $p < 0.0001$). There were no differences across treatments, neither for hope of success (control treatment, mean 16.300; No Info treatment, mean 16.767; $z = -1.379$, $p = 0.1680$) nor for fear of failure (control treatment, mean 11.400; No Info treatment, mean 11.344; $z = 0.200$, $p = 0.8417$).

There was no difference by choices in the scores of the fear of failure scale (egalitarian choice: $N = 82$, mean 11.780; performance-based choice, $N = 98$, mean 11.031; $z = 1.470$, $p = 0.1415$). However, participants who chose the performance-based split scored higher on the hope of success scale (egalitarian choice: $N = 82$, mean 15.841; performance-based choice, $N = 98$, mean 17.112; $z = -3.689$, $p = 0.0002$). This difference is also significant if we look separately at the control treatment (egalitarian choice: $N = 55$, mean 15.872; performance-based choice, $N = 35$, mean 16.971; $z = -2.914$, $p = 0.0283$) and the No Info treatment (egalitarian choice, $N = 27$, mean 15.778;

¹⁷Pooling together a large number of experimental studies, Filippin and Crosetto (2016) find significant but very small gender differences in risk attitudes, as measured by the lottery-questionnaire method.

performance-based choice, $N = 63$, mean 17.190, $z = -2.776$, $p = 0.0055$). This makes sense as hope of success should be associated with higher expectations, hence a higher likelihood of choosing the performance-based split. In other words, we do not gain any new insights from the achievement motivation scales.

To further examine the effects of achievement motivation and risk attitudes, and also to study the robustness of the basic effects, we added them as controls in a further regression model (see Table A.1). This model also includes interactions with the No Info treatment, which should reflect any possible explanation of the basic effect based on the individual differences captured by the new measures. This has the problem that the interpretation of the treatment dummy becomes more complex. To test for the treatment effect of the manipulation, we calculated the average marginal effects of the treatment dummy on the performance-based split choice.¹⁸ As shown in the bottom line of Table A.1, the average marginal effects of the treatment dummy are highly significant and quantify a 25% increase in the probability of choosing the performance-based split after including the manipulation in the structure of the information.

The regression model, however, did not detect any significant effects of risk attitudes, hope of success, or fear of failure. Hence, we conclude that these additional variables do not hide any alternative explanation and our basic effects, as captured in Model 3 (Table 2), are robust.

A.2 Experiment 2

As in Experiment 1, subjects in the No Info and Only Winner treatments were asked whether they wanted to be informed about the own absolute and relative performance after making the distribution choice. Only two subjects (of 180) declined to receive feedback. Hence, we conclude again that self-image protection does not appear to be a factor in our experiments.

In Experiment 2, we further investigated the possible relation of performance curiosity to existing constructs capturing curiosity. We focused on *epistemic curiosity* (Loewenstein, 1994; Litman et al., 2005; recall Section 2) and included the epistemic curiosity questionnaire by Litman and Mussel (2013). Further, since our explanation for the relevance of performance curiosity hinges on the value of feedback for the own self-image, we explored whether this link was captured by differences in standard *self-motives* (Sedikides and Strube, 1995, 1997; recall Section 2) and included the questionnaire by Gregg et al. (2011).

Model 4 (see Table A.2) includes the two measures for epistemic curiosity (EC) and the four measures of self-motives (SM) elicited through questionnaires, and the interaction of each of the six measures with the two information dummies. None of the new measures is significant, and among the twelve interactions only one reaches significance,

¹⁸The average marginal effect calculates the average difference in the predicted probability of choosing the performance-based split if all observations would have been in the No Info treatment, compared to if all observations would have been in the control treatment.

Table A.1: Experiment 1, Additional Probit Regression on Performance-based Choices

	Model 4	Interactions with No Info Treatment
No Info Treatment	-0.588 (1.7291)	
Female	-0.262 (0.3286)	
No Info X Female	0.227 (0.4726)	
Expected Performance in Septiles	0.536*** (0.0995)	
Risk Attitude	0.019 (0.3769)	-0.054 (0.5074)
Hope of Success	0.047 (0.0666)	0.002 (0.0865)
Fear of Failure	-0.068 (0.0513)	0.122 (0.0757)
Constant	-0.325 (1.3718)	
LogLikelihood	-88.304	
Wald Test	50.598***	
Average Marginal Effects No Info Treatment	0.247*** (0.0517)	

Notes. Standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

namely the interaction between self-improvement and the Hidden Information dummy. There is hence, no convincing basis to speculate that the performance curiosity concept might be captured or closely related to existing measures of curiosity or self-motives.

As in Model 3 (Table 4), there is no significant difference between the Only Winner and the No Info treatments. Hence, Model 4 shows the robustness of Model 3. To examine the robustness of the main results to the introduction of the new 18 coefficients, we again calculated the average marginal effects to test for the treatment effect. Those are shown at the bottom of Table A.2. The effect of the No Info treatment is positive and highly significant, showing that the probability of choosing the performance-based split is increased by approximately 24% by the basic manipulation. This almost exactly replicates the conclusion from Experiment 1 (recall Table A.1)

Table A.2: Experiment 2, Additional Probit Regression: Epistemic Curiosity (EC) and Self-Motives (SM)

	Model 4	Interactions with Hidden Info	Interactions with Only First
Hidden Info	0.328 (0.4407)		
Only First	0.257 (0.4783)		
Female	-0.155 (0.1901)		
Expected Performance in Septiles	0.628*** (0.0894)		
EC-Interest	-0.012 (0.0683)	0.034 (0.0970)	-0.020 (0.1038)
EC-Deprived	-0.010 (0.0460)	-0.008 (0.0856)	0.051 (0.0941)
SM-Enhance	-0.012 (0.1388)	0.027 (0.2061)	0.005 (0.2038)
SM-Assessment	-0.074 (0.1962)	-0.016 (0.2896)	-0.110 (0.2708)
SM-Verification	-0.029 (0.1397)	0.171 (0.2301)	-0.255 (0.2394)
SM-Improvement	-0.236 (0.1810)	0.574** (0.2601)	-0.232 (0.2357)
Constant	0.204 (0.2889)		
LogLikelihood	-130.392		
Wald Test	72.795***		
Average Marginal Effects			
Hidden Info	0.235*** (0.0561)		
Only First	-0.075 (0.0618)		

Notes. Standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix B: Translated Instructions

(Original instructions were in German)

General Instructions

The experiment consists of **two parts** in which **you and four other participants** solve addition problems and make a decision. After the completion of these two parts, a questionnaire will follow. In this experiment you will earn points which will be exchanged into Euros at end of the experiment. The exchange rate is:

$$1 \text{ point} = 5 \text{ Eurocents.}$$

Independently of your decision, you will additionally receive 2.50 EUR for your participation. The total amount will be paid **in cash** and **anonymously** at the end of the experiment. On the following pages you will receive all further information which you need for the experiment.

Instructions of the Experiment

General Sequence: For the duration of the whole experiment you belong to a group consisting of **5 players (including you)**. This group will not change during the experiment. The experiment consists of **two different parts**. In the first part you will solve a series of **addition problems**. In the second part you will make a **decision**.

Addition Problems: In the **first part** you will make a series of addition problems. Each addition problem consists of **5 two-digit numbers**. Your task is to add up these numbers, enter the sum into the appropriate field, and click “OK.” You can use scratch paper to take notes. You have **8 minutes** in total to solve addition problems. Independently of the correctness of your input a new addition problem will be presented each time you click “OK.” For each **correctly** solved addition problem you will earn **10 points** for the whole group. Incorrectly solved addition problem do not earn any points for the group.

After this part, a ranking within your group will be established. Your rank will be determined by the number of correctly solved addition problems. Rank 1 has solved the most and rank 5 has solved the least amount of addition problems. In case several players correctly solved the same amount of addition problems a random device decides which player is placed before the other one.

Your Decision:*[Control Treatment]*

The **second part** is about the **distribution** of the joint points of your group. You choose between **two alternatives**. The two options differ in how the earned points are allocated among the group members.

Your Decision:*[No Info Treatment, Only Winner Treatment]*

The **second part** is about the **distribution** of the joint points of your group. You choose between **two alternatives**. The two options differ how the earned points are allocated among the group members and what information is revealed to all players.

You choose one option by clicking the button below the distribution. Every player in your group (including you) will choose among the two options. After all players have decided one player of your group (including you) is chosen **randomly** and her/his decision is implemented.

References

- Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2015). Money, Depletion, and Prosociality in the Dictator Game. *Journal of Neuroscience, Psychology, and Economics*, 8(1):1–14.
- Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2016). The Impact of Self-Control Depletion on Social Preferences in the Ultimatum Game. *Journal of Economic Psychology*, 53:1–16.
- Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2018). Social Preferences and Self-Control. *Journal of Behavioral and Experimental Economics*, 74:161–166.
- Alós-Ferrer, C. and Ania, A. B. (2005). The Evolutionary Stability of Perfectly Competitive Behavior. *Economic Theory*, 26:179–197.
- Alós-Ferrer, C., Granić, D.-G., Kern, J., and Wagner, A. K. (2016). Preference Reversals: Time and Again. *Journal of Risk and Uncertainty*, 52(1):65–97.
- Apesteuguía, J., Azmat, G., and Iriberry, N. (2012). The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field. *Management Science*, 58(1):78–93.
- Azmat, G., Calsamiglia, C., and Iriberry, N. (2016). Gender Differences in Response to Big Stakes. *Journal of the European Economic Association*, 14(6).
- Azmat, G. and Iriberry, N. (2010). The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students. *Journal of Public Economics*, 94(7):435–452.
- Azmat, G. and Iriberry, N. (2016). The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction. *Journal of Economics & Management Strategy*, 25(1):77–110.
- Bandiera, O., Barankay, I., and Rasul, I. (2005). Social Preferences and the Response to Incentives: Evidence from Personnel Data. *Quarterly Journal of Economics*, 120(3):917–962.
- Bardsley, N. (2008). Dictator Game Giving: Altruism or Artifact? *Experimental Economics*, 11(2):122–133.
- Bénabou, R. and Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5):1652–1678.
- Bénabou, R. and Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets. *Quarterly Journal of Economics*, 126(2):805–855.
- Berger, J., Harbring, C., and Sliwka, D. (2013). Performance appraisals and the impact of forced distribution—an experimental investigation. *Management Science*, 59(1):54–68.
- Blanes i Vidal, J. and Nossol, M. (2011). Tournaments without Prizes: Evidence from Personnel Records. *Management Science*, 57(10):1721–1736.

- Bolton, G. E. and Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1):166–193.
- Brunstein, J. C. and Heckhausen, H. (2008). Achievement Motivation. In Heckhausen, J. and Heckhausen, H., editors, *Motivation and Action*, pages 137–183. Cambridge University Press, New York.
- Cappelen, A. W., Moene, K. O., Sørensen, E. Ø., and Tungodden, B. (2013). Needs versus Entitlements—An International Fairness Experiment. *Journal of the European Economic Association*, 11(3):574–598.
- Charness, G. and Gneezy, U. (2008). What’s in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games. *Journal of Economic Behavior and Organization*, 68(1):29–35.
- Charness, G. and Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, 117(3):817–869.
- Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.
- Dahme, G., Jungnickel, D., and Rathje, H. (1993). Psychometric Properties of a German Translation of the Achievement Motives Scale (AMS): Comparison of Results from Norwegian and German Samples. *Diagnostica*, 39:257–270.
- Delquié, P. (1993). Inconsistent Trade-Offs Between Attributes: New Evidence in Preference Assessment Biases. *Management Science*, 39(11):1382–1395.
- Fatas, E., Morales, A., and Jaramillo-Gutiérrez, A. (2015). Social Competition in Low-Information Cournot Markets. CBESS Discussion Paper 15-15, School of Economics, University of East Anglia.
- Fehr, E., Naef, M., and Schmidt, K. M. (2006). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. *American Economic Review*, 96(5):1912–1917.
- Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3):817–868.
- Fershtman, C., Gneezy, U., and List, J. A. (2012). Equity Aversion: Social Norms and the Desire to Be Ahead. *American Economic Journal: Microeconomics*, 4(4):131–144.
- Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7(2):117–140.
- Filippin, A. and Crosetto, P. (2016). A Reconsideration of Gender Differences in Risk Attitudes. *Management Science*, 62(11):3138–3160.
- Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, 10(2):171–178.
- Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3):347–369.
- Franzen, A. and Pointner, S. (2012). Anonymity in the Dictator Game Revisited. *Journal of Economic Behavior and Organization*, 81(1):74–81.

- Gilligan, C. (1982). *In a Different Voice*. Harvard University Press.
- Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica*, 77(5):1637–1664.
- Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics*, 118(3):1049–1074.
- Gregg, A. P., Hepper, E. G., and Sedikides, C. (2011). Quantifying Self-Motives: Functional Links Between Dispositional Desires. *European Journal of Social Psychology*, 41(7):840–852.
- Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1:114–125.
- Hannan, R. L., Krishnan, R., and Newman, A. H. (2008). The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *The Accounting Review*, 83(4):893–913.
- Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior*, 7(3):346–380.
- Holt, C. A. and Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5):1644–1655.
- Huck, S., Normann, H.-T., and Oechssler, J. (1999). Learning in Cournot Oligopoly - An Experiment. *Economic Journal*, 109:C80–C95.
- Hyde, J. S., Fennema, E., and Lamon, S. J. (1990). Gender Differences in Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, 107(2):139–155.
- Konow, J. (2003). Which is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature*, 41(4):1188–1239.
- Köszegi, B. (2006). Ego Utility, Overconfidence, and Task Choice. *Journal of the European Economic Association*, 4(4):673–707.
- Kruglanski, A. W. (1989). *Lay Epistemics and Human Knowledge: Cognitive and Motivational Bases*. Plenum, New York.
- Kuhnen, C. M. and Tymula, A. (2012). Feedback, Self-Esteem, and Performance in Organizations. *Management Science*, 58(1):94–113.
- Kuziemko, I., Buell, R. W., Reich, T., and Norton, M. I. (2014). “Last-Place Aversion”: Evidence and Redistributive Implications. *Quarterly Journal of Economics*, 129(1):105–149.
- Lang, J. W. and Fries, S. (2006). A Revised 10-Item Version of the Achievement Motives Scale: Psychometric Properties in German-Speaking Samples. *European Journal of Psychological Assessment*, 22(3):216–224.
- Litman, J., Hutchins, T., and Russon, R. (2005). Epistemic Curiosity, Feeling-of-Knowing, and Exploratory Behaviour. *Cognition & Emotion*, 19(4):559–582.

- Litman, J. A. (2008). Interest and Deprivation Factors of Epistemic Curiosity. *Personality and Individual Differences*, 44(7):1585–1595.
- Litman, J. A. and Mussel, P. (2013). Validity of the Interest-and Deprivation-Type Epistemic Curiosity Model in Germany. *Journal of Individual Differences*, 34(2):59–68.
- Loewenstein, G. (1994). The Psychology of Curiosity: A Review and Reinterpretation. *Psychological Bulletin*, 116(1):75–98.
- Mazar, N., Amir, O., and Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6):633–644.
- McClelland, D. C., Koestner, R., and Weinberger, J. (1989). How Do Self-Attributed and Implicit Motives Differ? *Psychological Review*, 96(4):690–702.
- Moore, D. A. and Klein, W. M. (2008). Use of Absolute and Comparative Performance Feedback in Absolute and Comparative Judgments and Decisions. *Organizational Behavior and Human Decision Processes*, 107(1):60–74.
- Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3):1067–1101.
- Offerman, T., Potters, J., and Sonnemans, J. (2002). Imitation and Belief Learning in an Oligopoly Experiment. *Review of Economic Studies*, 69(4):973–997.
- Roberts, T.-A. and Nolen-Hoeksema, S. (1989). Sex Differences in Reactions to Evaluative Feedback. *Sex Roles*, 21(11-12):725–747.
- Schoenberg, E. J. and Haruvy, E. (2012). Relative Performance Information in Asset Markets: An Experimental Approach. *Journal of Economic Psychology*, 33(6):1143–1155.
- Sedikides, C. and Strube, M. J. (1995). The Multiply Motivated Self. *Personality and Social Psychology Bulletin*, 21(12):1330–1335.
- Sedikides, C. and Strube, M. J. (1997). Self-Evaluation: To Thine Own Self Be Good, To Thine Own Self Be Sure, To Thine Own Self Be True, and To Thine Own Self Be Better. *Advances in Experimental Social Psychology*, 29:209–269.
- Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype Threat and Women’s Math Performance. *Journal of Experimental Social Psychology*, 35(1):4–28.
- Trope, Y. (1980). Self-Assessment, Self-Enhancement, and Task Preference. *Journal of Experimental Social Psychology*, 16(2):116–129.
- Van de Ven, N., Zeelenberg, M., and Van Dijk, E. (2005). Buying and Selling Exchange Goods: Outcome Information, Curiosity and the Endowment Effect. *Journal of Economic Psychology*, 26(3):459–468.
- Vega-Redondo, F. (1997). The evolution of Walrasian behavior. *Econometrica*, 65(2):375–384.