RATIONALITY & COMPETITION
CRC TRR 190

Show What You Risk - Norms for Risk Taking

Stefan Grimm (LMU Munich)

Discussion Paper No. 119

October 11, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Show what you risk - Norms for risk taking

Stefan Grimm*

May 23, 2018

Abstract

Most economic decisions are embedded in a specific social context. In many such contexts, individual choices are influenced by their observability due to underlying social norms and social image concerns. This study investigates the impact of choices being observed, compared to anonymity of choices, on risk taking in a laboratory experiment. I relate participants' investments in a risky asset directly to social norms for risk taking that are elicited in an incentivized procedure. I find that risk taking is not affected by the choice being observed by a matched participant. Nor do investments follow elicited norms for risk taking more closely when observed. This holds when considering males and females separately. However, I provide strong evidence for gender-specific norms in risk taking. While these explain part of the existing gender gap in risk taking, males still "overshoot" by investing more than the norm dictates. This is particularly true for males being matched with a female participant.

JEL-classification: C91, D01, D81, D91, G11
Keywords: Risk taking, observability, social image, norms, gender

*LMU Munich, Department of Economics, Geschwister-Scholl-Platz 1, 80539 Munich, Germany. E-mail: stefan.grimm@econ.lmu.de. Phone: +498921809787.
I would like to thank Vojtěch Bartoš, Francesco D'Acunto, Uri Gneezy, Martin Kocher, Julia Rose, Simeon Schudy, Jeroen van de Ven and Joachim Weimann as well as participants at seminars at the University of Munich and at the 2018 ESA Asia Pacific Meeting for helpful comments. Financial support of the German Research Foundation (DFG) through GRK 1928 and CRC TRR 190 is gratefully acknowledged.

1 Introduction

Decision making under risk is ubiquitous. Almost all — also economic — decisions involve some consideration of possible states of the future. Likewise, all of our risky decisions are embedded in a certain context, mostly including social features: decisions are made jointly, individual decisions affect other people, the decision maker observes others before deciding or her decision is observed by other people.

One elementary part of many, if not most, social contexts is that the choice is observed by other people. This is true in team decision making where team members get to know each individual's choice (live or ex post), in household decision-making when family members might learn about decisions made, or in many seemingly individual decisions where at one point others find out about one's choices (e.g. smoking, sports, investing). The decision maker might care about her choice being revealed if she cares about the signal the decision might send and her social image the decision might affect.[1] Such social image concerns have been shown to have an effect on revealed social preferences when strong behavioral norms exist (Andreoni and Bernheim, 2009).[2] In the present paper, I shed light on this issue by analyzing pure observability effects and eliciting norms in risky decision making.

Even though choosing safe or risky options can both be rational depending on risk preferences, there is some evidence that norms for risk taking actually exist and that these are gender-specific. Bem (1974) initiated research investigating desirability scales for personality traits (see also Auster and Ohm, 2000; Harris, 1994; Holt and Ellis, 1998; Prentice and Carranza, 2002). Subjects rate traits as desirable or not desirable — for females and males separately. The ratings are then compared to define a trait as female or male in relative desirability. Supported by the follow-up studies, Bem (1974) finds that 'willingness to take risks' is among the masculine characteristics. That is, this characteristic is significantly more desirable for males.[3] Further, there is ample evidence for actual gender differences in risk preferences with males being less risk averse than females (for example Charness and Gneezy, 2012; Eckel and

[1]See Brennan and Pettit (2004) for a detailed representation of how people care about how they are perceived. Other seminal work in economics by Akerlof (1980) and Holländer (1990) discusses theoretical models incorporating social image and reputation concerns into utility functions.

[2]Andreoni and Bernheim (2009) as well as Dana et al. (2007) provide evidence that dictator giving might stem from social image concerns instead of pure altruism. Bohnet and Frey (1999), Dufwenberg and Muren (2006), Filiz-Ozbay and Ozbay (2014), Gächter and Fehr (1999) and Rege and Telle (2004) show that identification can have strong effects on behavior in dictator and public good games. Ariely and Levav (2000) and Ratner and Kahn (2002) show similarly strong effects for variety seeking in consumption.

[3]This assessment does not depend on the gender of the rating person. Farthing (2005) and Wilke et al. (2006) report somewhat different results. In Farthing (2005) only heroic risk taking is generally deemed desirable. Non-heroic risk takers are only preferred by males in same-sex friends. Wilke et al. (2006) report that social and recreational risk taking was rated attractive in a potential partner, while for example risk taking in investment was rated neutrally. They do not find pronounced gender differences in these ratings.

Grossman, 2008, for reviews).[4] Hence, if descriptive norms (what people actually do) and injunctive norms (what people should do) are linked (as suggested, e.g., by Rudman and Phelan, 2008, p.63), we would expect to see a gender difference in injunctive norms for risk taking as well. Finally, Prentice and Carranza (2002) denote risk taking as a "gender-relaxed prescription", i.e. it is generally desirable — including females —, but only more so for males.

With people wanting to adhere to social norms (see, e.g., Elster, 1989; López-Pérez, 2008) and with observability of choices further increasing norm adherence preferences through reputation and social image concerns (see, e.g., Akerlof, 1980; Holländer, 1990), this evidence on social norms in risk taking implies that observability of risk choices should increase risk taking. This is particularly true for males, for whom desirability of risk taking seems much more pronounced.

To the best of my knowledge, I am the first to cleanly investigate the effect of observability of risk choices alone and hence to analyze a social image effect of the revealed risk preferences. With respect to potential channels, I provide first evidence on social norms in risk taking from an incentivized elicitation procedure.

I implement a laboratory experiment in which participants are matched with another participant and make a risky investment choice (Gneezy and Potters, 1997). The matching includes visual identification. While choices for participants in the control condition are anonymous, participants in the treatment condition know that the matched participant will learn about their risk choices at the end of the experiment. To account for potential gender differences in treatment effects, I balance the sample on gender. This further allows me to analyze matched participant gender effects. After investment decisions are made, I elicit beliefs about the choice of the matched participants (descriptive norms) and behavior deemed appropriate (injunctive norms) using a procedure similar to Krupka and Weber (2013).[5] I consequently link these norms to actual risk taking.

Overall — and for both males and females separately — I do not find an effect of choices being observed on risk taking. However, both descriptive and injunctive norms strongly differ between males and females. This helps to understand the gender gap in risk taking, that is also prevalent in my data. While females on average do not strongly deviate from injunctive norms, males clearly "overshoot": They invest more than what they think other people deem appropriate. This pattern is driven by participants that indicate to care little about norm conformity.

Understanding pure observability effects and existing norms in risky choices

[4]Note that Filippin and Crosetto (2016) indicate that these gender differences might depend on the specific task used. They suggest that, e.g., the availability of a salient safe option is one element in risk tasks that induces gender differences (Crosetto and Filippin, 2017).

[5]Descriptive norms relate to "what most others do", while injunctive norms define "what most others approve or disapprove" (see, e.g., Cialdini et al., 1990). These concepts closely correspond to descriptive and prescriptive stereotypes (Gill, 2004).

is important for both modeling economic behavior and to comprehend biased measurement. First, policy makers, firms or other agents can better estimate individual and group decision making depending on the specificities of the social context. For example, financial industry firms can set up policies limiting or easing direct information flows between proprietary traders to affect signaling or social image effects on investment decisions. Policy makers can influence the observability of customer investment or insurance decisions in financial consulting procedures, but similarly in related domains (e.g. preventive health care or treatment choice). Furthermore, group decision making strongly hinges on norms and social image concerns. When forming teams for sensitive functions, supervisors need to know whether signaling concerns lead for example specific gender constellations to produce very different risky behaviors. Second, an "observer effect" in stated and revealed risk preferences — if existent — needs to be considered in survey designs when interviewers observe responses. Otherwise, measurement error in response behavior systematically impedes high data quality.

In my design I exclude any potential channels that might confound the measurement of mere observability effects: learning by the observer, outcome-based preferences via observing outcomes as well, and signaling skill or superior information by the choice. Basically only information on the curvature of the utility function is observed. My setup further allows a systematic analysis of gender pairing effects and can directly link incentivized risk taking to incentivized beliefs about norms in risk taking. These norms have so far only been elicited in non-incentivized procedures and the evidence mostly relates to relative, not absolute, desirability by gender.

Other studies have looked at aspects of observability of risky choices, where other confounding concerns are present, too. Yechiam et al. (2008) run an experiment with choices between a risky lottery and a safe option and pairs of participants both observe the other's choice and outcome live on their screen. The authors find that the social exposure increases risk taking compared to a purely individual control group in one out of two tasks used.[6] Tymula and Whitehair (2017), however, find no effect of live observation on risky choices in the laboratory. While these studies allow to discuss observer behavior as well, I can disentangle the effects of merely being observed (social image concerns and norms) from the effects of possibly affecting the choice of the matched participant (expecting learning from the other) and mere consistency preferences between tasks. Further, by observing choices *and* outcomes, outcome-based social preferences might play a role in the findings of Yechiam et al. (2008). Curley et al. (1986) find that ambiguous choices are made less often if experimental participants were observed by a group of other participants. Other

[6] In a small second study including 32 participants, only one participant observes the matched participant's choices and outcomes. In this study those participants observing the other's choice and outcome choose the risky option more often.

evidence comes from accountability studies. In these studies, participants have to explain and justify their choices to the experimenter after the experiment. Vieider (2009) finds that participants behave less loss averse when "held accountable". For choices under risk, Weigold and Schlenker (1991) report that experimental participants become more extreme in their revealed risk preference. With subjects facing rather complex lotteries represented as histograms this might stem from these more extreme but consistent choices being easier to justify in front of the experimenter.[7]

Lastly, there is a strand of literature predominately in psychology that provides mostly correlational evidence on the effects of being observed in various forms of risky behavior. Hamed (2001), Himanen and Kulmala (1988) and Pawlowski et al. (2008) assess road crossing behavior depending on group composition and bystanders, Chen et al. (2000), Ebbesen and Haney (1973), Jackson and Gray (1976) and Nuyts and Vesentini (2005) take car driving behavior in combination with proximity of other cars and passenger characteristics, Ronay and Hippel (2010) report from skateboard tricks, and Frankenhuis et al. (2010) take bridge crossing time in virtual reality as risk measure. These studies mostly indicate that males increase risk taking in the presence of females. While this is suggestive evidence for an observability effect, it remains unclear to what degree these findings arise from endogenous assignment to treatment, from subjects being able to signal more than mere risk preferences and being able to affect others with their choices, from the lack of incentives, or indeed from social image concerns regarding risk preferences. Baker and Maner (2008, 2009) and Frankenhuis and Karremans (2012) further explicitly relate risk taking in males to mating preferences and relationship status, arguing that single males use risky behaviors to attract attractive females. My design allows me to test these observer gender and observer characteristics effects, too. While I indeed see that males take more risk and "overshoot" norms more strongly when matched with a female, this is independent of the choice being observed and hence does not relate to social image concerns. However, I find some evidence for the attractiveness of the matched participant being important for the effects of observability. Participants react differently (more risk taking) to the choice being observed if matched with an attractive participant compared to being matched with a less attractive participant.

The remainder of the paper is structured as follows. Section 2 describes the experimental design in detail. Section 3 presents the risk-taking results and Section 4 discusses gender-specific norms and norm following behavior. I discuss the results in Section 5 and conclude in Section 6.

[7] As reviewed by Patil et al. (2014), depending on whether there is a normatively correct choice or not, accountability can lead subjects to make choices that are simply more easily justified or to exert more cognitive effort to find the correct answer (see also Simonson, 1989; Simonson and Nye, 1992).

2 Experimental Design and Procedures

2.1 Experimental Design

At the beginning of the experiment, subjects were randomly assigned seats in the laboratory. After reading of the instructions, subjects were informed on-screen that they were matched with the subject seated in the seat vis-à-vis their own (facing each other). For that purpose, the wooden wall of the cubicle usually shielding subjects facing each other was removed before the experiment. Hence, besides the screen blocking most of the view, matched subjects were able to see each other.[8] Further, the first experimental screen showed the picture of the matched participant and all following decision screens showed a small picture of the matched participant at the bottom of the screen. In a between-subjects design, one subject in a pair was assigned to the treatment condition and the other to the control condition.

The main and first part of the experiment was an investment task (Gneezy and Potters, 1997). Subjects received 100 Taler worth €5 and could invest any integer amount in a risky asset. The asset paid off 2.5 times the invested amount with a 50% probability. With the remaining 50% the investment was lost, implying an expected return of 25%. The amount not invested was kept with certainty, but did not pay any interest.

To investigate the effect of the choice being observed, before making their decision in the investment game, half of the subjects were told that their choice would be shown to their matched participant at the end of the experiment. Hence, choices of these subjects in *Treatment* were not anonymous to the matched partner. The other half of the subjects (*Control*) was told that their choice was anonymous. Revealing the choice at the very end of the experiment excludes any type of learning by the observer. Not revealing the outcome excludes an impact of outcome-based social preferences. Showing matched participant pictures and allowing visual identification of the matched participant in both *Treatment* and *Control* holds matched participant identification effects constant across treatments and enables me to measure a clean effect of only the choice itself being observed.

In the second part of the experiment, all subjects answered a non-incentivized risk questionnaire on stated willingness to take risks in general and in the domains of driving, finance, sport, trust, health and career (see, e.g., Dohmen et al., 2011). Here too, subjects in *Treatment* were informed that their choices would be shown to their matched participant at the end of the experiment. This allows me to provide some evidence on whether observability of choices affects decision making differently in the different domains. Comparing the results from these non-incentivized questions to incentivized investment behavior can further speak to the interaction between

[8] See Figure A.8 in Appendix D for a picture of the seat arrangement.

signaling and signaling costs.

In the third part, subjects' beliefs about choices of others and different types of norms were elicited. This part contained five elicitation procedures of which one was at the end of the experiment randomly chosen to be paid. However, before starting part three, subjects were debriefed about the treatment conditions. They were informed that half of the participants had made anonymous choices while the other half of the participants had made choices that would be observed at the end of the experiment. It was also announced that always one subject from each treatment formed a pair. This is important for comparing elicited beliefs and norms across treatments. It further allows me to test whether subjects in fact expected treatment effects.[9]

In the first belief eliciation procedure, subjects were asked for beliefs about the matched participant's investment (*guess partner*). Subjects could earn €5 if they did not deviate by more than 10 Taler from the true value. This 10-Taler-deviation-based incentive scheme was the same for all following elicitation procedures. In the second procedure, subjects stated beliefs about the average investment in the experimental session (*guess all*).

The third, fourth, and fifth elicitation procedures measure injunctive norms for investment and are inspired by the procedure first used by Krupka and Weber (2013). Prior to that, all subjects were informed that their picture would be shown to four other participants at a later stage. These four participants would then have to indicate the appropriate amount that the person in the picture "should have invested". I label the average of these four statements as the injunctive norm for that person.

Consequently, in the third elicitation, each subject (e.g. subject *A*) had to anticipate this average norm (*perceived norm*), that was to be indicated by the four (unknown) other participants (when seeing *A*'s picture). Subjects were then told that one of the four participants that would see their picture would be their matched participant.[10] For the fourth elicitation, subjects had to anticipate what appropriate amount the matched participant would indicate when seeing their picture. I denote this anticipation *perceived norm partner*. In the fifth and final elicitation procedure all subjects then actually saw — one by one — four pictures and indicated what they thought were appropriate investment amounts for the respective participants (*stated norm 1 - stated norm 4*). The last picture rated (*stated norm 4*) always showed the matched participant. Hence, this fifth elicitation procedure consisted of four choices: Every subject indicates this norm for four participants, and hence every participant's picture is seen by four other participants.[11] If this fifth elicitation procedure in the

[9]Further, without debriefing it would have been very difficult to elicit beliefs about behavior of the entire group and the matched participant without deceiving subjects (by omission of information). Subjects would have wrongly expected the matched participant to have seen the same instructions.

[10]The first three pictures were actually chosen randomly — by randomly assigned seat numbers.

[11]The average of the four statements of participants seeing the same picture is what I labeled as injunctive norm above. This again is the value that the subject in the picture had to anticipate in *perceived*

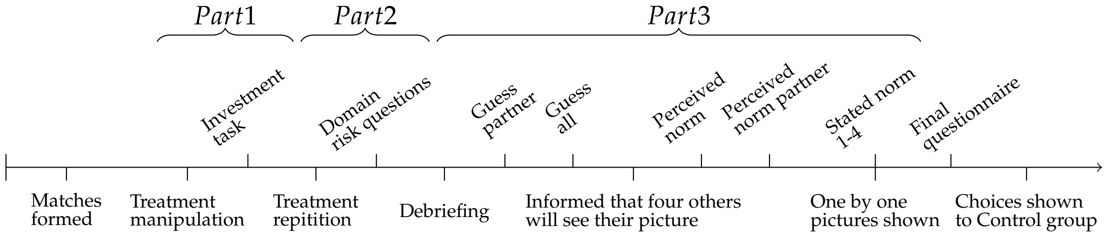


Figure 1: Timeline of the experiment

end was by the computer chosen to be payoff relevant for a participant, one of the four stated norms was randomly selected for payment. The subject was then paid if she did not deviate by more than 10 Taler from the average answer of the three other participants rating the respective picture.[12] Figure 1 displays the timeline of the experiment.

After the elicitation tasks, all subjects answered a final questionnaire. Next to standard questions related to age, gender, field of study and mother tongue, the questionnaire also consisted of open-end questions that allowed subjects to make general comments regarding the experiment and to explain what they considered during the decision process. Further, Likert-scale questions asked for *norm conformity* ("How much do you usually conform to norms?"), *rule breaking* ("How much do you like to break rules?") and *social image* ("How much do you care about other people's perception of you?") preferences, as well as how risk-loving (risk-avoiding) subjects would want to be perceived (*ideal perception*). While I can control for standard observables when estimating the treatment effect, the survey questions allow me to analyze heterogeneity in the treatment effect.

Lastly, I collected data on picture characteristics. Four research assistants (RAs) independently coded each picture on whether the individual made "eye contact" with the camera and looked friendly, and rated attractiveness on a scale from one to ten.[13] This allows me to check whether treatment or matched gender effects depend on visual cues.

2.2 Procedural Details

I programmed and conducted the experiment with "z-Tree" (Fischbacher, 2007) and 428 subjects, 215 males and 213 females, were recruited using the online recruiting

norm.

[12]This coordination game induced by the incentive scheme is used to identify beliefs regarding group perceptions of what people ought to do, i.e. injunctive norms. See Krupka and Weber (2013) for details.

[13]For the data analysis I use the average attractiveness rating of all four RAs. The binary variables *eye contact* and *friendly face* are one if more than two out of four RAs indicated so. RAs further guessed age, gender and ethnicity to account for subjects' looks potentially diverging from facts. No meaningful differences emerged (e.g. none at all for the assignment of participants to sex).

system "ORSEE" (Greiner, 2015).[14]

All 28 experimental sessions took place at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) between May and October 2017. To ensure a gender balance for in expectation same-sized gender pairing cells, half of the planned number of subjects per session were of one gender. Upon arrival at the laboratory, subjects first had to sign a consent form that allowed pictures to be taken during the experiment.[15] Subjects were then randomly assigned seats in the laboratory and (portrait) pictures were taken individually upon entering the laboratory. While instructions were read out aloud, RAs copied the pictures via remote access from the camera to the local drives of the subjects, so that pictures could later be displayed onscreen.

All subjects were paid privately after the experiment and earned €12.68 on average (including a fix payment of €5 for showing up on time), ranging from €5 to €22.50. While the investment task was always paid, only one of the elicitation procedures in part 3 was at the end randomly and individually chosen for payment by the computer. The sessions lasted slightly less than 45 minutes on average.

3 Risk Taking

3.1 Manipulation Check and Sample Balance

To make sure that subjects indeed perceived the treatments differently — a necessary condition to observe a causal treatment effect — I asked subjects at the end of the experiment how much they felt being watched when making the investment decision (on a scale from one to ten). Figure 2 shows average answers to the question by treatment condition including 95% confidence intervals.

Subjects in *Treatment* clearly indicate that they felt being watched more strongly than subjects in *Control* (p-value < 0.01, Mann-Whitney test). Hence, the treatment effectively changed the decision environment of subjects.[16]

Table A.1 in the Appendix reports a randomization table between *Control* and *Treatment*. I do not observe any significant differences in socio-economic background variables and picture ratings between the treatment conditions.

[14]The final questionnaire included a question whether subjects had heard of the experiment before participating. Since subjects were debriefed and future subjects could have been made aware of the treatment manipulation and research interest of the experiment, I exlude 12 (out of 440) subjects that had indeed heard of the experiment. This exclusion does not affect my results.

[15]See Appendix E for the exact wording. The email invitation to sign in for an experimental session made clear that pictures would be taken during the experiment. Nobody objected to pictures being taken. Appendix F further includes the exact wording of the instructions.

[16]While power naturally decreases, this is directionally true for all gender and gender pairing subsamples.

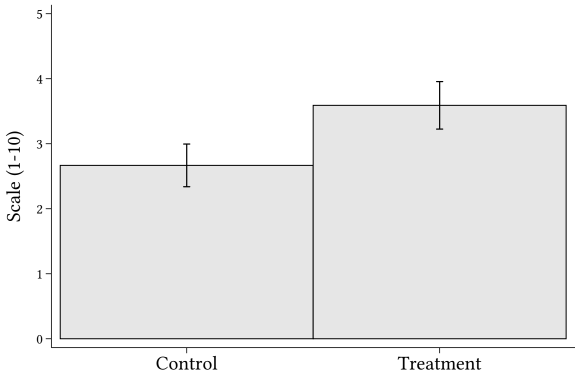


Figure 2: Manipulation check: Stated feeling of being watched

3.2 Full Sample Results

Based on the literature discussed above, I expect higher investments when being observed for males, but possibly also a (weaker) positive treatment effect for females (see the weaker, but existing, norms for females in Prentice and Carranza, 2002). If willingness to take risks indeed is a desirable trait, social image concerns in *Treatment* should push subjects to riskier choices compared to *Control*.

In contrast to that, I do not see an overall difference in investment by treatment. Subjects in *Treatment* invest 51.14 Taler on average, while *Control* subjects' average investment is 52.29. This small difference is clearly insignificant (p-value = 0.55, Mann-Whitney test). Statistical power to detect treatment effects is not the reason for this null-finding. To detect the measured effect size as being significant, I would need roughly 16,000 observations and with the 428 observations and given the standard deviation in investment in my sample I would be able to detect an effect size of roughly 7 Taler (0.27 standard deviations) with a power of 80% (two-sided test, $\alpha = 0.05$).

For a complete representation of investment amounts, Figure 3 displays the cumulative distribution functions of investment by treatment condition. With the functions crossing multiple times and never strongly diverging, I clearly do not see large differences in the distributions (p-value = 0.61, Kolmogorov-Smirnov test).[17] Further, the distribution shows that I have sufficient variation in investments to potentially observe treatment effects, such that an overly strong focal point at 50 is not responsible for the null effect.

Even though the sample was balanced on observables across treatments, I check the robustness of the overall non-parametric null finding in a regression framework with additional controls. Table 1 displays Tobit regressions on investment. Model (1) explains investment solely with the treatment indicator and therefore is the parametric equivalent to the non-parametric test. Model (2) and (3) add gender and the gender of

[17] There seem to be some differences by treatment in the fraction of subjects choosing round numbers (multiples of 10; i.e., 0, 10, 20,...). Table A.2 in the Appendix gives a more detailed overview of these patterns.

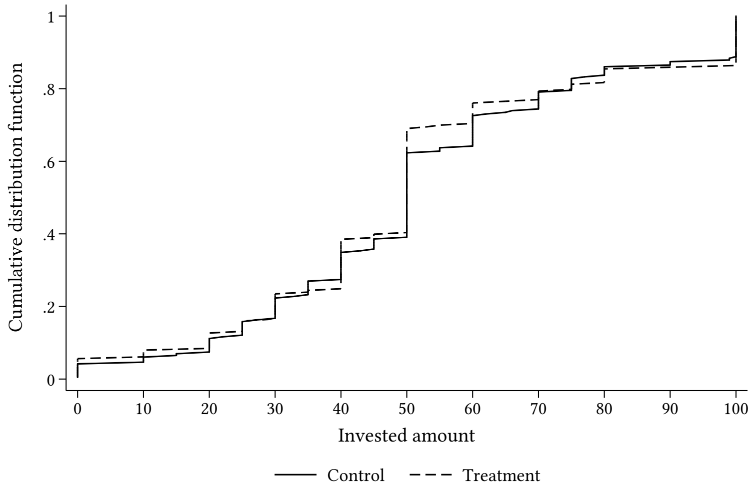


Figure 3: Cumulative distribution function of investment by treatment

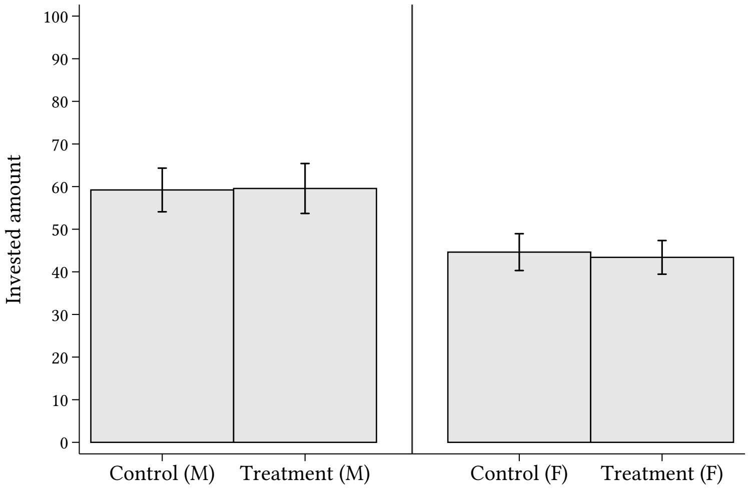
the matched participant as controls, respectively. The effect of *female* clearly shows a large gender gap in investments. This is in line with much of the literature suggesting gender differences in risk taking, particularly in this type of task. The gender of the matched participant has no significant overall effect. Model (4) adds standard observables, model (5) further includes information from the individual's picture and model (6) further incorporates survey responses on *norm conformity*, *rule breaking*, *social image* and *ideal perception*.

Table 1: Tobit regressions on investment

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	−1.033	−0.117	0.044	0.362	0.204	1.597
	(3.046)	(2.661)	(2.648)	(2.695)	(2.596)	(2.437)
Female		−18.493***	−18.065***	-16.972***	-17.199***	-16.197***
		(2.580)	(2.550)	(2.535)	(2.581)	(2.509)
Female Partner			3.280	3.166	2.982	2.571
			(3.249)	(3.237)	(3.312)	(3.133)
Constant	53.365***	62.169***	60.246***	39.648***	28.039	2.597
	(2.088)	(2.548)	(3.076)	(15.170)	(19.243)	(19.486)
Standard observables	No	No	No	Yes	Yes	Yes
Picture characteristics	No	No	No	No	Yes	Yes
Survey responses	No	No	No	No	No	Yes
Observations	428	428	428	428	428	428
Pseudo R^2	0.00	0.01	0.01	0.01	0.02	0.02

Notes: Two-limit (0-100) Tobit regressions on invested amount. Clustered (on experimental session level) standard errors in parentheses. Stars indicate significant coefficients, and *** p<0.01, ** p<0.05, * p<0.1. Standard observables include: age, last math grade in school (indicator variables), nationality (indicator for German, European, Non-European), relationship status, whether the subject studies economics or business and time spent on the first experimental screen. Picture characteristics were rated by RAs: friendly look, attractiveness and eyecontact. Survey responses include *norm conformity*, *rule breaking*, *social image* and *ideal perception* from the questions at the end of the experiment.

The regression table shows that the null effect of *Treatment* is very robust to controlling for all available information. Even when adding the survey measures in



Notes: Error bars indicate 95% confidence intervals. "M" and "F" refer to data for males and females, respectively.

Figure 4: Treatment effect for males and females separately

model (6), the treatment effect remains very small and clearly insignificant. This is despite these measures potentially being endogenous to treatment and most likely biasing the treatment coefficient upwards.[18] Apart from the significant effect of gender, better math grades (for all specifications), *rule breaking* and *ideal perception* are significantly positively linked to investment.

3.3 Gender and Matched Gender

Looking at overall treatment effects possibly obscures differences by gender. While there is evidence that risk taking is desirable for females, too, the vast majority of papers considering desirability of risk taking highlights a strong asymmetry in desirability by gender. With such an asymmetry, the overall null effect might be the result of a negative treatment effect for females canceling out a positive treatment effect for males, for example. Similarly, gender pairing is a prime candidate for heterogeneity. The literature in psychology on mating preferences and risk taking (Baker and Maner, 2008, 2009; Frankenhuis and Karremans, 2012) for example is an indication for an asymmetric treatment effect on males. Males should react particularly strong and positive when being matched with a female. In this subsection, I will consider these sources of treatment effect heterogeneity.

Contrary to expectations, I do not find a treatment effect for neither gender. Figure 4 shows that males and females both do not react to their choices being observed

[18]Both *social image* (p-value = 0.081, Mann-Whitney test) and *ideal perception* (p-value = 0.086, Mann-Whitney test) are weakly significantly lower in *Treatment* compared to *Control*. They are also both overall positively related to investment amounts. Hence, controlling for these measures wrongly estimates lower predicted investments for subjects in *Treatment*, leading to a higher coefficient on *Treatment* to compensate for that effect.

Table 2: Treatment effects overall, by gender and by gender pairs

	Control (n)	Treatment (n)	Treatment effect	p-value
All Subjects	52.29 (215)	51.14 (213)	−1.15	0.55
Females	44.63 (102)	43.40 (111)	−1.23	0.51
Males	59.21 (113)	59.56 (102)	0.35	0.80
Females matched with females	38.40 (47)	46.11 (45)	7.71	0.14
Females matched with males	49.95 (55)	41.55 (66)	−8.40	0.02**
Males matched with females	61.48 (66)	64.36 (55)	2.88	0.54
Males matched with males	56.02 (47)	53.94 (47)	−2.08	0.94

Notes: Invested amounts for all subjects, males and females separately, and all four gender pairs separately, by treatment condition. Values in parenthesis denote the number of observations in a given cell. P-values for the treatment differences are based on Mann-Whitney tests, and *** p<0.01, ** p<0.05, * p<0.1.

(p-value for females = 0.51; for males = 0.80; Mann-Whitney tests).[19] I only clearly see the overall gender differences in risk taking already seen in the regression results. Men on average (independent of treatment) invest 59.38, while females only invest on average 43.99 (p-value < 0.01, Mann-Whitney test).

Besides considering gender separately and showing again the overall treatment effect, Table 2 displays subgroup treatment effects for the four possible gender pairs. This shows whether potentially opposing treatment effects by matched gender cancel each other out when ignoring matched gender.

It indeed seems as if the effects within gender pairs cancel out for females, obscuring existing treatment effects. Females invest clearly less in *Treatment* when matched with males, while they directionally invest more in *Treatment* when matched with another female. The difference between these differences is large and significant (diff-in-diff of 16.11, p-value = 0.01, two-sided t-test) suggesting that females react differently to *Treatment* depending on the matched participant gender.[20] While the treatment effect for females matched with males gets (weakly) insignificant when correcting for multiple hypothesis testing (conservative Bonferroni correction for four hypotheses tested, $\hat{p} = 0.09$), the difference in differences is pronounced and robust ($\hat{p} = 0.014$, correction for two hypotheses tested).

However, this does not seem to be a treatment effect per se: The difference arises from within *Control*, where investment depends on the gender of the matched participant. The average investment of females in *Treatment* does not depend on the matched gender (p-value = 0.55, Mann-Whitney test) and for both matched genders does not differ significantly from the average investment of females overall. In *Control*

[19] Similarly, there is no change of the distribution from *Control* to *Treatment* for neither females nor males considered separately. See Figure A.1 in the Appendix for cumulative distribution functions of investment by treatment conditional on gender.

[20] I can compare two empirical distributions non-parametrically. However, for difference in differences tests in my between-subjects design, I can only calculate the treatment effect on means and have to rely on parametric assumptions for testing using the t-test.

however, females' investment is much higher if they are matched with a male as compared to when being matched with a female (p-value < 0.01, Mann-Whitney test). Further, there is some evidence that these treatment effects on the gender pair level arise from failed randomization in some subsamples. I will discuss these results in Section 4.2 when relating investment behavior to norms.

3.4 Other Dimensions of Treatment Heterogeneity

Apart from the main subsamples by gender, I further check heterogeneity of treatment effects by personality traits. *Social image* and *ideal perception* could strongly affect the treatment effect. Social image concerns are important, because if a person does not care about how she is perceived, then behavior should be independent of treatment. Likewise, conditional on having social image concerns, the effect of *Treatment* should crucially depend on how subjects want to be seen. If subjects do not want to be perceived as willing to take risks, they should lower risk taking in *Treatment*. If they do want to be perceived as willing to take risks, they should increase risk taking. While I find *ideal perception* to have a positive and significant impact on investment overall, there is no significant interaction with the treatment condition. This holds for controlling for above median social image concerns.

The only weak heterogeneity in treatment effects relates to the attractiveness of the matched partner. While for both — being matched with an attractive or non-attractive participant (measured by below or above median rated attractiveness) — the treatment effect is insignificant, they clearly go in opposite directions: Those matched with an attractive partner increase risk taking in *Treatment* (insignificantly) and those matched with an unattractive partner reduce risk taking in *Treatment* (again insignificantly). This difference in treatment effects (TED) is significant (TED = 11, p-value = 0.04, two-sided t-test). I can look at specific subsamples separately. It seems that males (TED = 14, p-value = 0.08) and those matched with female partners (TED = 13, p-value = 0.09) show this pattern more strongly. The difference gets clearly larger (TED = 33; p-value = 0.01; n = 66) only considering single males matched with females. Further — among these — only considering those with above median social image concerns and above median ideal risk perception shoots up the TED to 72 Taler (p-value = 0.08; n = 12).[21]

3.5 Non-Incentivized Domain-Specific Risk Questions

The reaction to observability might generally depend on how choices are incentivized. On the one hand, if subjects want to signal a specific type or trait with their risky choice when being observed, this comes at a signaling cost if the choice is incentivized. This

[21] See Figure A.2 in the Appendix for a graphical illustration of these patterns.

signaling is costless if the choice is not incentivized. On the other hand, there might be a relationship between the (perceived) informativeness of a signal and signaling costs. Signaling a specific type might only be effective if the signal itself is credible, i.e. when deviation from truth-telling is costly.

Further, there might be domain-specific effects of the risk choice being visible, since people may want to be perceived differently depending on what type of risk taking is considered.

I use the domain-specific risk questionnaire from the German Socio-Economic Panel (SOEP) that elicits willingness to take risks in general, in car driving, in personal finance, in sports, in trusting other people, in health and in one's career. With the main task in the experiment relating to financial risk taking, this can also shed light on how the null effect measured in the main part might translate into other domains.

Figure 5 shows basically no differences between the treatment effects in the different domains. All domain-specific treatment effects are small and insignificant. This again is no power issue. The 95% confidence intervals span only slightly more than a 0.5 treatment effect size allowing me to detect small effects on the questionnaire scale from 1-10 (with a power of 80%, $\alpha = 0.05$, and two-sided tests, I would be able to detect an effect size of roughly 0.25 for all domains). Figure A.3 in Appendix B shows treatment effects by domain for all four gender pairs separately. Also there, none of the effects is significant (and none large).

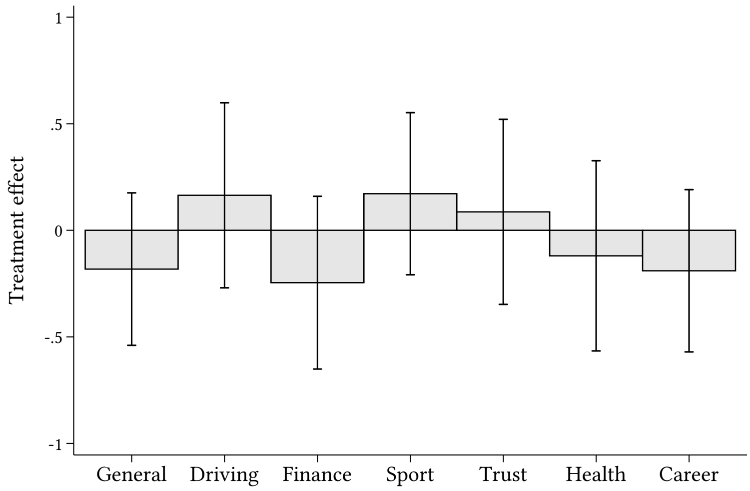


Figure 5: Treatment effect on non-incentivized domain-specific risk taking

4 Norms

With much of the experimental design focusing on different types of norms in the investment decision, I next discuss the overall gender-specific patterns in these norms and then relate these elicited norms to actual investment behavior. If not stated

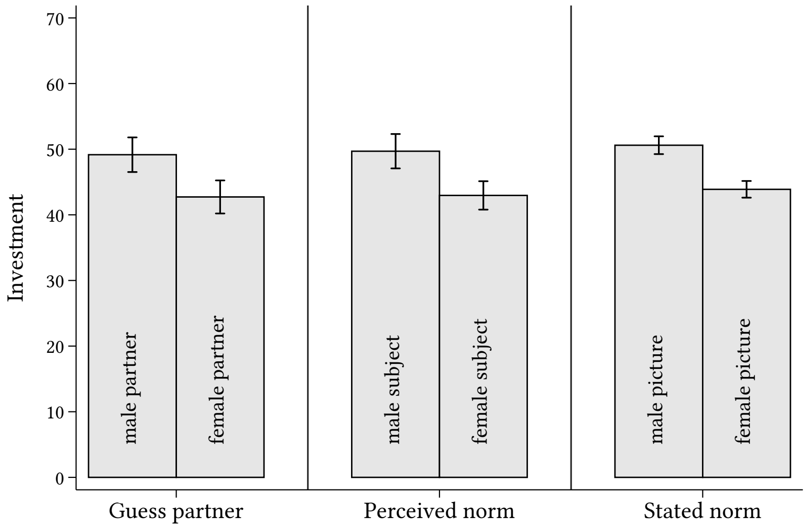


Figure 6: *Guess partner* by matched participant gender, *perceived norm* by subject gender, and *stated norm* by picture gender

otherwise, I refer to overall norms — independent of treatment — and only distinguish between the treatment conditions where informative.

4.1 Gender-Specific Norms

In light of the large gender differences in actual risk taking, it is interesting to see whether subjects expect these gender differences. This is indeed the case. Subjects think their matched participant invested more if the matched participant was male (p-value < 0.01, Mann-Whitney test) providing strong evidence for gender differences in descriptive norms in risk taking. The left panel of Figure 6 shows the effect of matched participant gender on *guess partner* (including 95% confidence intervals). This difference in beliefs is stronger for females (p-value < 0.01, Mann-Whitney test), but directionally similar for males only (p-value $= 0.16$). The finding is robust to only using data from subjects that indicated to be at least somewhat confident in their guess. After every belief statement, I asked subjects to indicate confidence on a scale from one to five. Excluding subjects that stated one ("I am not at all sure about my answer — I basically guessed randomly") does not change the result — if anything, the (matched) gender differences become more pronounced. This is true for all of the following other norm statements.

While *guess all* does not allow to differentiate between norms for females and males, I observe that, despite there being large gender differences in actual investment, the average guess of females regarding the average session investment does not at all differ from the average guess of males. Subjects generally underestimate average investment by slightly more than 4.5 Taler (p-value < 0.01, Wilcoxon signrank test).

As depicted in the middle panel of Figure 6, not only descriptive norms differ by gender, but *perceived norm* clearly depends on gender, too (p-value < 0.01, Mann-Whitney test). Males think they should have invested 49.69, while females think they should have invested only 42.96.[22] This is strong support for the non-incentivized survey evidence on gender-specific desirability of risk attitudes (e.g. Bem, 1974) and shows that the difference is robust to incentivizing subjects for normative statements.

Lastly, also stated norms for investment are higher for male pictures (right panel of Figure 6). That is, subjects agree on males being supposed to invest more. This difference by gender of the rated picture is highly significant over data from all four pictures rated combined (average of *stated norm 1* to *stated norm 4* by picture gender as displayed in Figure 6, p-value < 0.01, Mann-Whitney test), but also when I compare ratings for male and female pictures for *stated norm 1* to *stated norm 4* separately (p-value < 0.01 for all four pictures, Mann-Whitney test).[23] Both males and females hold these gender specific norms.

For a detailed overview of all indicated norms (including *perceived norm partner* and *guess all* which are not shown in Figure 6) by gender, matched participant gender and treatment cell, see Table A.3 in Appendix C.

In the next subsection, I discuss the relationship between norms and investment, both overall and by treatment. Independent of this relationship it is interesting to note that in line with there being no treatment effect, subjects did not expect investment differences based on treatment. *Guess partner* is independent of treatment.

4.2 Norms and Investment Behavior

When people care about social image that in turn depends on norm adherence, one should expect individuals' investment behavior to more closely track perceived norms in *Treatment* compared to *Control*. However, patterns of norm adherence in my experiment are independent of treatment condition, i.e. of the choice being observed or not. This is true overall and when looking at behavior of females or males separately. Consequently, when discussing the relationship between norms and investments in the following, I will abstract from the treatment condition and report results over both treatments combined.

[22] This is very similar for *perceived norm partner*, even though the difference is only weakly significant (p-value = 0.06, Mann-Whitney test). This difference becomes larger and significant at the 1%-level if I consider only subjects that indicated to not having basically randomly guessed the value (one fourth excluded). *Perceived norm partner* was elicited to detect potential differences in the perceived norm by gender of the rating person. This is not the case, neither overall nor for females or males separately. Since *perceived norm* corresponds to the more general notion of norms and is much more meaningful for subjects in *Control*, I refer to *perceived norm* in the main analyses. Results generally are very similar when using *perceived norm partner* and I indicate any difference where applicable.

[23] Having to state norms for different people might make subjects think there should be a difference in their assessment - even when there originally is not. This is not a concern when only considering the first picture. While subjects could in principle have inferred that they would have to rate more participants (they knew that their own norm would be based on four other participants) only 10% of the subjects in the follow-up questionnaire indicated that they expected to rate more than one picture.

Section 4.1 demonstrates large norm differences between males and females. These can explain (at least part of) the gender gap in choices under risk in my experiment. However, they do not explain the entire gap in investment. While descriptive and injunctive norms for males are roughly at 50 Taler, average actual investment of males lies at almost 60 Taler. That is, males clearly "overshoot" beyond their perceived norms leading to an even larger gender gap in actual investments.

The difference between actual investment and *perceived norm* (what should affect individual behavior) — which I use to describe "norm following" — is highly significant for males (p-value < 0.01, Wilcoxon signrank test). This is not the case for females. Figure 7 shows average norm following by gender and matched gender cells and plots 95%-confidence intervals. While the difference for males is significant for either matched participant gender, it is especially pronounced for males being matched with females. This difference in norm following by matched gender is weakly significant (p-value $= 0.05$, Mann-Whitney test).

The norm following results for males inform an interesting pattern. Ignoring treatment, men invest on average 62.79 when matched with a female and 54.98 when matched with males. This investment difference by the gender of the matched partner is significant at the 5% level (p-value $= 0.04$, Mann-Whitney test). As indicated in Figure 7, this cannot be explained by a difference in perceived norms depending on matched gender. It rather is indeed a (matched gender-specific) "overshooting" beyond perceived norms.

In contrast to males, females' deviation from norms is much less pronounced and overall insignificant. Only for those matched with males, the deviation is weakly significant (p-value $= 0.09$, Wilcoxon signrank test).[24] As for males, the difference in norm following by matched gender is weakly significant for females, too (p-value $=$ 0.09, Mann-Whitney test).[25]

Lastly, *rule breaking* and *norm conformity* should naturally be linked to norm following behavior. Those with strong norm conformity preferences can be assumed to follow perceived norms more closely (vice versa for rule breaking preference). This is borne out by the data, at least for males. For those with above median norm conformity preferences the norm deviations are not significant, and small in magnitude. In contrast to that, the deviations are very large and highly significant for males with below median norm conformity preferences. As expected, just the reverse — only

[24] Inference based on the confidence intervals (calculated with t-statistics) can lead to slightly different results (p-values) than when using the (correct) non-parametric Wilcoxon signrank test.

[25] These findings are similar if I use *perceived norm partner* instead of *perceived norm*. The diff-in-diff for males is then significant at the 1%-level. For females matched with males the deviation becomes insignficant, while the "undershooting" when matched with females becomes significant (p-value $<$ 0.01, Wilcoxon signrank test). The diff-in-diff for females is not significant. The "undershooting" of females matched with females, however, is entirely driven by females in *Control*, for which *perceived norm partner* is less applicable. This difference to Figure 7 using *perceived norm* should therefore not be overweighted. For norm following by treatment condition see Figure A.5 in the Appendix.

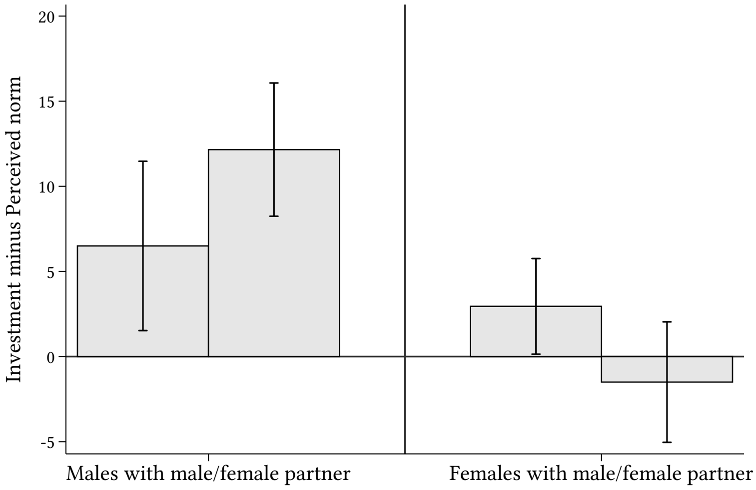


Figure 7: Norm following by gender pairs

less pronounced — is true for rule breaking preferences. For females not much of a difference emerges for *norm conformity*, while *rule breaking* is positively linked to norm defiance. See Figures A.6 and A.7 in the Appendix for details.

Explaining the Treatment Effect for Females Matched with Males

I use answers to belief and norm questions to look into and understand the significant treatment effect for investment of females matched with males. As discussed above, female investments do not strongly deviate from norms in neither treatment when matched with males. This shows that the strong treatment effects for these females can to a large extend be explained by differences in perceived norms. The question then is, where these norm differences by matched gender and treatment come from.

One can put the norm differences by treatment for females differently: Given being in *Control* or *Treatment*, females' *perceived norm* depends on the gender of the matched participant. In *Control*, females perceive higher investments as the norm when matched with males compared to when matched with females (p-value = 0.05, Mann-Whitney test). In *Treatment*, however, females perceive lower investments as the norm when matched with males compared to when matched with females (p-value = 0.04, Mann-Whitney test). This is something I should not observe. Subjects know that four other — not mentioned — participants will state the appropriate investment amount. As far as subjects are concerned, they cannot infer anything from the matched participant about the four selected participants. Hence, *perceived norm* should be independent of the matched participant.

For stated norms something very similar applies. Independent of what picture they rated (random for pictures one to three), stated norms of females matched with males are on average lower in *Treatment* than in *Control* (p-value = 0.06, Mann-Whitney test).[26] If they indeed perceived different norms based on treatment, why would they also on average state different norms for random other people (some in *Treatment*, some in *Control*; some female, some male)? The pattern for descriptive norms is similarly unintuitive.[27]

This casts doubt on the notion that investment and norm differences for females matched with males are indeed induced by the treatment per se. The evidence speaks rather in favor of an unfortunate randomization leading females in *Treatment* and matched with a male to invest less and at the same time indicate lower descriptive and injunctive norms compared to those in the control group. Evidence from balancing tests supports this. Females matched with males have a significantly lower *ideal perception* in *Treatment* (p-value = 0.02, Mann-Whitney test), which can possibly explain the treatment difference for these females.[28]

This overall asymmetry in response behavior can help to explain the observed treatment effect for females matched with males arising from the large investment difference for females in *Control* depending on the matched participant gender. The remaining unexplained difference in (opposing) norm deviations in *Control* between those matched with males and those matched with females is only weakly significant (p-value = 0.09, Mann-Whitney test). That is, in *Control*, when matched with a male, females invest relatively more than their perceived norm compared to when matched with a female. The difference in the deviation from norms is insignificant for females in *Treatment*. This again points towards — if anything — *Control* inducing these behavioral differences.

5 Discussion

In this paper I demonstrate a clear overall null effect of observability and hence signaling opportunities on choice under risk. That is, merely having somebody knowing the choice does not affect decision making. The experimental design eliminates other channels that could potentially affect decision making. As such, I exclude concerns regarding the influence one might have on others, outcome-based social preferences, the mere psychological pressure to decide with "live" audiences, opportunities to explain or justify choices and the chance to provide more than merely a signal regarding the curvature of the utility function. While the manipulation check

[26] I only take average stated norms for pictures 1 to 3, since including picture 4 (the matched participant) would make the average again depend on the matched partner. Including the matched participant does not change the result.

[27] As can be seen in Table A.3, for descriptive norms, too, females matched with males in *Control* consistently state higher values than those in *Treatment*.

[28] Note that *ideal perception* of course can itself be endogenous to treatment. It seems rather unlikely, however, that *ideal perception* should only be depending on treatment for females matched with males. For all other gender pairs no differences exist.

demonstrates that subjects indeed were affected by the treatment, they overall did not change risk taking out of social image concerns when choices were observed.

This is a surprising null effect based on the literature. If willingness to take risks is deemed desirable — particularly for males — the opportunity to signal a risky type should lead subjects to invest more. Interestingly, however, injunctive norms are generally not very high in my experiment. Norms averaging at an investment of 50 Taler stand somewhat in contrast to the notion that willingness to take risks is generally deemed appropriate.[29] This could be one potential explanation for the null effect of the treatment manipulation. If the absolute norm level is not high, why should people — when observed and when caring about social image — increase risk taking? The norm levels do not explain norm following behavior though, which surprisingly does not depend on the treatment condition either. If people care about their social image, they should have a much stronger incentive to behave according to prescriptions when observed compared to when making anonymous, purely individual decisions.

Gender and gender pairing are the most natural subsamples in my setting to consider in terms of heterogeneity in treatment effects. On these dimensions I find very little evidence for treatment effects, and the evidence on treatment effects for females matched with males seems to be mainly driven by randomization issues.

The attractiveness of the matched participant, however, seems to interact with the treatment effect. While the overall effect for the entire sample is weak, this interaction becomes very large for some subsamples. Considering the cell sizes of these ever smaller subsamples, I urge the reader to interpret these patterns cautiously. Nevertheless, these sometimes very pronounced asymmetries are striking and relate to the literature on mating preference induced behavior in psychology. Baker and Maner (2008) relatedly indicate that the mere exposure of males to pictures of attractive females leads to a positive relationship between "mating preferences" and risk taking, which was not observable for any other group. Similarly, males in Baker and Maner (2009) that expected to meet a female participant at the end of the experiment selected riskier experimental choices when that female participant was single, interested in seeing somebody and would learn about the outcome. Frankenhuis and Karremans (2012) show contrasting results for males in a relationship. They seem to not adjust own behavior to what they think females consider attractive, contrary to behavior of single males.[30]

Moving away from the incentivized investment task, I also do not find treatment

[29]Similar to investments, the average of 50 for perceived norms does not merely arise from a focal point at 50. Instead, there is large variation in these norms. See Figure A.4 in Appendix C for the distribution of perceived norms and investment.

[30]Evolutionary theory suggests some mechanisms why risk taking of males might indeed be perceived as attractive by females (see, e.g., Kelly and Dunbar, 2001, for a discussion of the arguments). Mate choice theory highlights resource availability and protection as elementary factors for female survival. These might be better provided by brave and risk tolerant males. Signaling good genes by risky behavior makes risk taking attractive based on sexual selection theory.

effects on non-incentivized risk attitude statements in any domain. This is maybe even more surprising than the null effect in the investment task, since the signal here is basically free (if we abstract from truth-telling preferences). However, it is possible that signals have to be costly to be credible.

Lastly, and besides treatment effects, males — independent of treatment condition — clearly invest more when matched with females. Since this holds also for *Control* and cannot be explained by differences in injunctive norms, this is a remarkable effect. It relates to the findings by Carr and Steele (2010) and D'Acunto (2015) showing that males increase risk taking after a stereotype threat or gender identity priming, respectively. A similar effect might drive behavior in my experiment: Sitting vis-à-vis a female participant and seeing her picture could already prime males on their gender identity and induce more risk taking, clearly beyond the perceived individual norm.

6 Conclusion

This paper provides first clean experimental evidence that observability of the choice alone in a decision under risk does not affect overall risk taking. That is, in my setting, risk taking is not strategically used as a signal to affect social image.

This directly relates to many settings of individual decision making without strong relationships between the decision maker and the observer. Considering survey interview responses, but also decisions for example in front of doctors or financial advisors have very much in common with the controlled environment of the experiment. In many other and related domains, next to the mere observability of the choice and the opportunity for signaling, other elements of social contexts are relevant. Disentangling the effect of this one basic element is crucial to understand these more complex environments.

One prominent setting in which knowledge of these effects is especially important is group decision making under risk where signals are immediate and oftentimes important. Understanding the signaling values of revealed risk preferences can potentially help to explain inconclusive findings regarding the transmission of individual risk preferences into group risk preferences and decision making (see, e.g., Kugler et al., 2012). The evidence on gender-specific effects of observability depending on the attractiveness of observers (i.e. for example team colleagues in a group setting) further highlights that the gender distribution in teams might have very specific effects. While very recent papers (e.g. Lamiraud and Vranceanu, 2017; Lima de Miranda et al., 2017) discuss the effects of gender composition per se on risky group decision making, more research is needed to understand the mechanisms behind gender and possibly attractiveness specific effects. The finding that males generally increase risk taking when being matched with a female further highlights the importance of understanding the effect of gender identities and matching.

Besides the analysis of treatment effects, my findings clearly establish a large gender difference in norms for risk taking. This closely relates to and helps to explain the usually observed gender differences in actual risk taking — with males taking more risk than females. At the same time, I clearly show that norms do not explain the entire difference in actual risk taking. Rather, males "overshoot" in their risky choices clearly beyond norms. Importantly, this pronounced asymmetry in revealed norm conformity — which surprisingly is independent of observability of choices — is robust to controlling for self-assessed norm conformity preferences. While I measure and establish endogenously emerged norms, it would be interesting to look at norm following behavior with respect to exogenously established norms. Exogenous variation allows to directly measure behavior as a function of different norms.

Interestingly, while the general finding of gender-specific norms in risk taking is very robust, the absolute norm levels provide an insight into the general desirability of risk taking. Norms in the experimental setting describe intermediate levels of risk taking and do not fully support the idea that risk taking overall is desirable. Further research in different domains of risk taking is needed to assess the robustness of this finding.

References

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics*, 94(4):749–775.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Ariely, D. and Levav, J. (2000). Sequential choice in group settings: Taking the road less traveled and less enjoyed. *Journal of Consumer Research*, 27(3):279–290.

Auster, C. J. and Ohm, S. C. (2000). Masculinity and femininity in contemporary american society: A reevaluation using the Bem Sex-Role Inventory. *Sex Roles*, 43(7):499–528.

Baker, M. D. and Maner, J. K. (2008). Risk-taking as a situationally sensitive male mating strategy. *Evolution and Human Behavior*, 29(6):391–395.

Baker, M. D. and Maner, J. K. (2009). Male risk-taking as a context-sensitive signaling device. *Journal of Experimental Social Psychology*, 45(5):1136–1139.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42(2):155–162.

Bohnet, I. and Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, 89(1):335–339.

Brennan, G. and Pettit, P. (2004). *The economy of esteem: An essay on civil and political society*. Oxford University Press, Oxford.

Carr, P. B. and Steele, C. M. (2010). Stereotype threat affects financial decision making. *Psychological Science*, 21(10):1411–1416.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1):50–58.

Chen, L.-H., Baker, S. P., Braver, E. R., and Li, G. (2000). Carrying passengers as a risk factor for crashes fatal to 16-and 17-year-old drivers. *Journal of the American Medical Association*, 283(12):1578–1582.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.

Crosetto, P. and Filippin, A. (2017). Safe options induce gender differences in risk attitudes. *Working Paper*.

Curley, S. P., Yates, J. F., and Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes*, 38(2):230–256.

D'Acunto, F. (2015). Identity, overconfidence, and investment decisions. *Working Paper*.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dufwenberg, M. and Muren, A. (2006). Generosity, anonymity, gender. *Journal of Economic Behavior & Organization*, 61(1):42–49.

Ebbesen, E. B. and Haney, M. (1973). Flirting with death: Variables affecting risk taking at intersections. *Journal of Applied Social Psychology*, 3(4):303–324.

Eckel, C. C. and Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, 1:1061–1073.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, 3(4):99–117.

Farthing, G. W. (2005). Attitudes toward heroic and nonheroic physical risk takers as mates and as friends. *Evolution and Human Behavior*, 26(2):171–185.

Filippin, A. and Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science*, 62(11):3138–3160.

Filiz-Ozbay, E. and Ozbay, E. Y. (2014). Effect of an audience in public goods provision. *Experimental Economics*, 17(2):200–214.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Frankenhuis, W. E., Dotsch, R., Karremans, J. C., and Wigboldus, D. H. (2010). Male physical risk taking in a virtual environment. *Journal of Evolutionary Psychology*, 8(1):75–86.

Frankenhuis, W. E. and Karremans, J. C. (2012). Uncommitted men match their risk taking to female preferences, while committed men do the opposite. *Journal of Experimental Social Psychology*, 48(1):428–431.

Gächter, S. and Fehr, E. (1999). Collective action as a social exchange. *Journal of Economic Behavior & Organization*, 39(4):341–369.

Gill, M. J. (2004). When information does not deter stereotyping: Prescriptive stereotyping can foster bias under conditions that deter descriptive stereotyping. *Journal of Experimental Social Psychology*, 40(5):619–632.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, 112(2):631–645.

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

Hamed, M. M. (2001). Analysis of pedestrians' behavior at pedestrian crossings. *Safety Science*, 38(1):63–82.

Harris, A. C. (1994). Ethnicity as a determinant of sex role identity: A replication study of item selection for the Bem Sex Role Inventory. *Sex Roles*, 31(3-4):241–273.

Himanen, V. and Kulmala, R. (1988). An application of logit models in analysing the behaviour of pedestrians and car drivers on pedestrian crossings. *Accident Analysis & Prevention*, 20(3):187–197.

Holländer, H. (1990). A social exchange approach to voluntary cooperation. *American Economic Review*, 80(5):1157–1167.

Holt, C. L. and Ellis, J. B. (1998). Assessing the current validity of the Bem Sex-Role Inventory. *Sex Roles*, 39(11):929–941.

Jackson, T. T. and Gray, M. C. (1976). Field study of risk-taking behavior of automobile drivers. *Perceptual and Motor Skills*, 43(2):471–474.

Kelly, S. and Dunbar, R. I. (2001). Who dares, wins. *Human Nature*, 12(2):89–105.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Kugler, T., Kausel, E. E., and Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4):471–482.

Lamiraud, K. and Vranceanu, R. (2017). Group gender composition and economic decision-making. *Journal of Economic Behavior & Organization*, forthcoming.

Lima de Miranda, K., Detlefsen, L., and Schmidt, U. (2017). Can gender quotas prevent excessive risk taking? The effect of gender composition on group decisions under risk. *mimeo*.

López-Pérez, R. (2008). Aversion to norm-breaking: A model. *Games and Economic Behavior*, 64(1):237–267.

Nuyts, E. and Vesentini, L. (2005). The relation between seat belt use of drivers and passengers. In de Waard, D., Brookhuis, K., van Egmond, R., and Boersema, T., editors, *Human Factors in Design, Safety, and Management*, pages 81–92. Shaker Publishing, Maastricht.

Patil, S. V., Vieider, F., and Tetlock, P. E. (2014). Process versus outcome accountability. *Oxford Handbook of Public Accountability*, pages 69–89.

Pawlowski, B., Atwal, R., and Dunbar, R. (2008). Sex differences in everyday risk-taking behavior in humans. *Evolutionary Psychology*, 6(1):29–42.

Prentice, D. A. and Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, 26(4):269–281.

Ratner, R. K. and Kahn, B. E. (2002). The impact of private versus public consumption on variety-seeking behavior. *Journal of Consumer Research*, 29(2):246–257.

Rege, M. and Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, 88(7):1625–1644.

Ronay, R. and Hippel, W. v. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Social Psychological and Personality Science*, 1(1):57–64.

Rudman, L. A. and Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, 28:61–79.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16(2):158–174.

Simonson, I. and Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, 51(3):416–446.

Tymula, A. and Whitehair, J. (2017). Young adults gamble less when observed by peers. *Working Paper*.

Vieider, F. M. (2009). The effect of accountability on loss aversion. *Acta Psychologica*, 132(1):96–101.

Weigold, M. F. and Schlenker, B. R. (1991). Accountability and risk taking. *Personality and Social Psychology Bulletin*, 17(1):25–29.

Wilke, A., Hutchinson, J. M., Todd, P. M., and Kruger, D. J. (2006). Is risk taking used as a cue in mate choice? *Evolutionary Psychology*, 4(1):367–393.

Yechiam, E., Druyan, M., and Ert, E. (2008). Observing others' behavior and risk taking in decisions from experience. *Judgment and Decision Making*, 3(7):493.

Appendix

A Randomization

Table A.1 reports results from a randomization test between *Control* and *Treatment*. It lists average values by treatment for the subject characteristics. These include individual observables (age, gender, nationality, mother tongue, math grade, relationship status, as well as decision — i.e. reading — time for the first screen which was the same for both treatments) and information obtained from the pictures (whether they made eye contact, looked friendly and how attractive they were rated).

Table A.1: Balance table by treatment conditions

Variable	Control	Treatment	p-value of difference
Age	22.935	23.521	0.354
Female	0.474	0.521	0.334
Geman nationality	0.795	0.822	0.490
German mother tongue	0.791	0.779	0.775
Math grade	2.101	2.072	0.621
Relationship status	0.460	0.484	0.632
Time left after first screen	21.181	21.329	0.962
Eye contact with picture	0.926	0.944	0.449
Friendly face in picture	0.474	0.469	0.919
Attractive (0-10)	4.836	4.891	0.554
Total earnings	12.429	12.942	0.222

Notes: Math grade refers to the final (last) math grade in high school. *Time left after first screen* is seconds left upon reading the instructions onscreen and can serve as proxy for reading and comprehension speed. Measures regarding the photo taken were rated independently by 4 RAs and the average was taken for *attractive*, while for the binary measures regarding eye contact and facial expression the dummy is coded as one if at least three out of four RAs indicated one. The test of difference between the treatments used is either Chi^2 or Mann-Whitney, depending on whether variables only have categorical values or distributions.

B Details on Investment Choices

B.1 Investment Patterns for Round Numbers

Partitioning investment choices in specific focal and non-focal investment amounts gives some insights into behavior depending on treatment. While there is neither an overall treatment effect nor a general change in the distribution, it seems that focal (salient) investment amounts become somewhat more important in *Treatment*. Table A.2 displays these patterns. Directionally, more *Treatment* subjects choose to invest all, nothing as well as exactly half of their endowment. Combining these investment amounts into one group results even in a weakly significantly higher share investing either 0, 50 or 100 in *Treatment* (p-value = 0.05, Chi^2-test). While not more subjects in *Treatment* invest 25 or 75 (one fourth or three fourth), the difference between the treatments when only considering subjects investing multiples of ten (or zero) becomes even more apparent: More than 90% in *Treatment* invest in such a manner, while only 78% do so in *Control* (p-value < 0.01, Chi^2-test).

Table A.2: Fraction of subjects investing focal amounts

	Control	Treatment	p-value
Fraction of subjects investing 100:	0.12	0.14	0.45
Fraction of subjects investing 0:	0.04	0.06	0.49
Fraction of subjects investing 50:	0.24	0.29	0.21
Fraction of subjects investing 0, 50, or 100:	0.40	0.49	0.05*
Fraction of subjects investing 0, 25, 50, 75 or 100:	0.47	0.54	0.18
Fraction of subjects investing multiples of 10:	0.78	0.91	0.00***

Notes: Fractions of subjects, by treatment condition, making investment choices based on certain patterns: Investing the entire endowment, investing nothing, investing exactly have of the endowment, investing one of either of these focal points, investing any amount represented by 25 point increments or investing any amount represented by 10 point increments. The p-value of the test for difference between the treatments is based on Chi^2-tests (expected cells size > 50, results robust to using Fisher exact test). Stars indicate significance, with *** p<0.01, ** p<0.05, * p<0.1.

While the effects are not very large in magnitude and hence should be interpreted with caution (i.e. multiple testing), they are in line with evidence from accountability studies mentioned in Section 1. These studies indicate that subjects who need to justify their choices to the experimenter after the experiment, choose more easily justifiable options. Investing round numbers can be interpreted in a similar way. Even though subjects do not have to explicitly justify behavior in front of their matched participant, they might still expect uneven investments to be a very specific signal. This indicates that if researchers or marketing departments are indeed interested in exact (and sometimes "weird") values, they should consider these accountability and observability effects in study designs.

B.2 Distribution of Investment Choices by Gender

Figure A.1 displays cumulative distribution functions by treatment for males and females separately. This clearly shows that considering average investments in Figure 4 does not obscure any more subtle treatment effects on the distribution of choices. Kolmogorov-Smirnov tests confirm this assessment (p-value $= 0.98$ for males, p-value $= 0.66$ for females).

B.3 Heterogeneity in Treatment Effects

Figure A.2 displays the difference in the treatment effects between subjects matched with an attractive partner and those matched with an unattractive partner.

Every bar relates to a specific sample considered and the positive values almost always arise from both (directionally) negative treatment effects for those matched

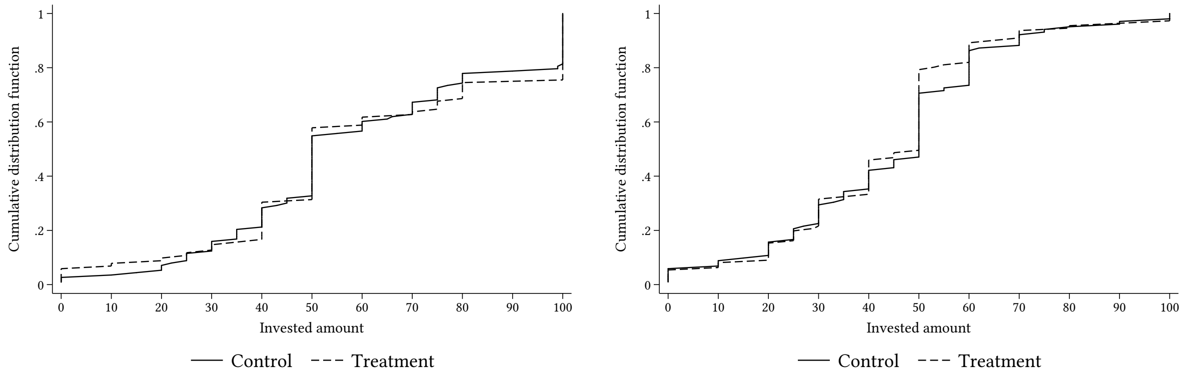
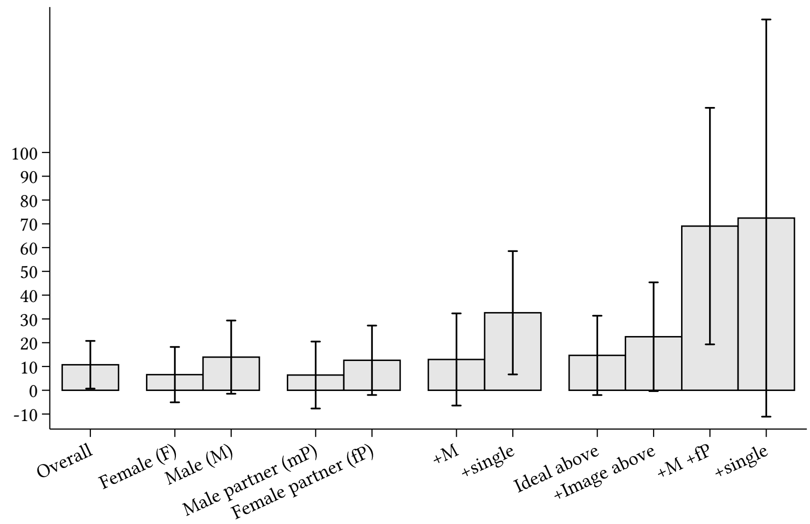


Figure A.1: Cumulative distribution function of investment by treatment for males (left) and females (right) separately

with an unattractive participant and (directionally) positive treatment effects for those matched with an attractive participant. Confidence intervals are based on t-tests. For some subsamples, these differences become very large. Interstingly, the further I move to the right in the figure and the higher the intuitively expected treatment effect differences should get, indeed the more pronounced effects I observe.

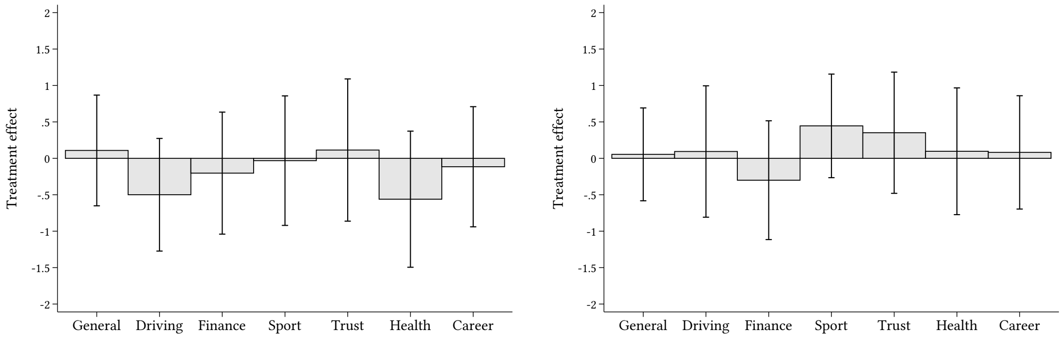


Notes: X-axis labels with a "+" indicate that the respective subsample constraint is put on top of the subsample definition of the bar to the left.

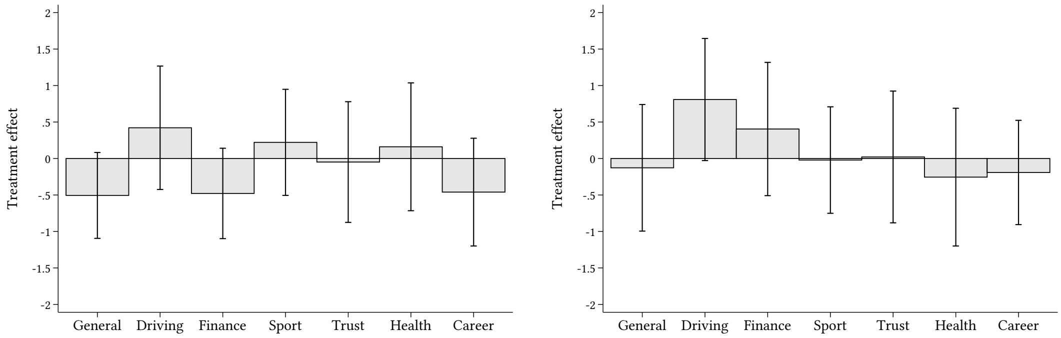
Figure A.2: Treatment effect difference between matched with an attractive vs. unattractive participant by subsamples

B.4 Non-Incentivized Domain-Specific Stated Willingness to Take Risks

Just as for overall treatment effects on domain-specific non-incentivized risk taking, I do not observe strong patterns when considering gender pairs separately. See Figure A.3 for the equivalents of Figure 5 by gender pairs.



Answers by females (left) and males (right) matched with females



Answers by females (left) and males (right) matched with males

Figure A.3: Treatment effect on non-incentivized domain-specific risk taking by gender pairs

The largest effect observed is the treatment effect on willingness to take risks in driving for males matched with males. When being observed, these subjects state higher risk attitudes than in *Control*. The effect size, however, is still below one and only weakly significant (p-value = 0.08, Mann-Whitney test; p-value = 0.06, two-sided t-test as displayed in Figure A.3). Apart from this difference, based on t-tests, only females' general risk assessment is weakly higher in *Treatment* compared to *Control* when matched with males (p-value = 0.09; p-value = 0.11, Mann-Whitney test). One statistic that does not show up in the figure (and t-tests) is the difference in sports for males matched with females. Based on a non-parametric test response behavior is clearly different in *Treatment* (p-value = 0.02, Mann-Whitney test). This is insignificant with a t-test since the treatment means are not too far apart. However, answers are much more dispersed in *Treatment* such that the Mann-Whitney test results in a significant statistic. For no other gender pairing or domain a similar effect could be observed.

While for these subsamples I do not have sufficient power to detect small effects, I'd still be able to detect economically important differences. Importantly, the statistically small effects would become even less meaningful once corrected for multiple testing.

C More Details on Norms

C.1 Norm Choices

For completeness, Table A.3 displays all elicited beliefs (i.e. norms) by gender, matched gender and treatment cells. To allow for a comparison to actual investment choices, average invested amounts by cell are included.

Table A.3: All norms by gender, matched gender and treatment cells

	MMC	MMT	MFC	MFT	FMC	FMT	FFC	FFT
Guess partner	49.60	45.53	44.44	43.84	52.22	48.88	40.23	41.44
Guess all	46.19	42.28	50.74	48.55	49.87	44.68	45.40	47.44
Perceived norm	52.19	44.77	49.79	51.65	46.13	39.32	41.83	45.60
Perceived norm partner	54.32	47.32	49.45	50.04	50.47	42.03	46.00	46.29
Stated norm (1-3)	47.99	44.13	47.23	49.89	50.11	44.94	45.82	49.81
Stated norm (4)	50.30	44.96	43.70	45.11	52.82	47.55	42.77	45.87
Investment	56.02	53.94	61.48	64.36	49.95	41.55	38.40	46.11

Notes: All choices in the belief elicitation part of the experiment (and investment) by gender, matched partner gender and treatment with MMC (males in *Control* matched with males), MMT (males in *Treatment* matched with males), MFC (males in *Control* matched with females), MFT (males in *Treatment* matched with females), FMC, FMT, FFC and FFT (the four groups equivalent to before only for female decision makers) denoting the different treatment combination cells. *Guess partner* refers to the average guess for the investment of the matched partner for subjects in the respective cell. *Guess all* equivalently refers to the guessed overall investment in the given session. *Perceived norm* is the subject's belief regarding the average stated norm of the four people being shown the subject's picture. *Perceived norm partner* denotes the beliefs about that stated norm by the matched partner. *Stated norm (1-3)* and *stated norm (4)* refer to average stated norms by subjects in the respective cell (for the first three pictures seen, and for the fourth picture — the picture of the matched partner).

Figure A.4 shows kernel densities for *perceived norm* and *investment*. The distributions show that a strong focal point at half the investment amount can neither per se explain the null effect of the treatment on investment, nor the average perceived norm level of 50. For both outcomes there is sufficient variation in participant answers.

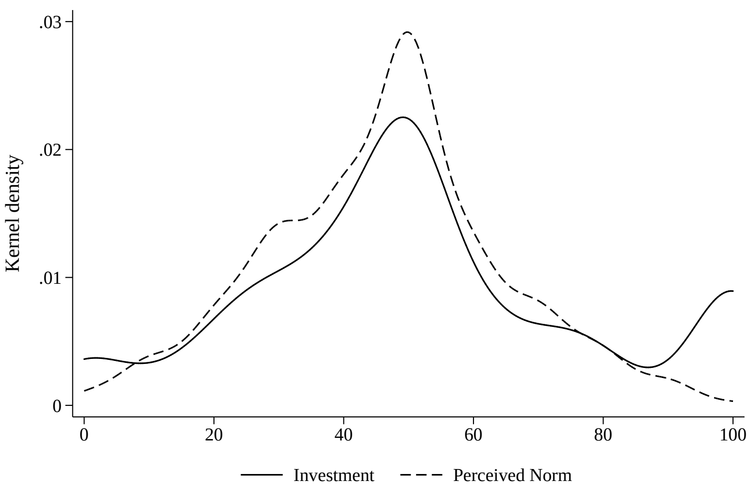
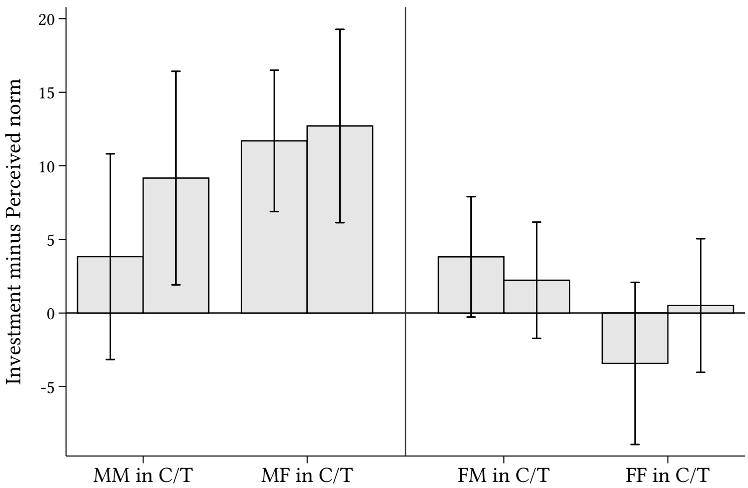


Figure A.4: Kernel density functions for *perceived norm* and *investment*

C.2 Norm Following

As indicated in Section 4.2 norm following overall does not depend on treatment condition. Figure A.5 shows norm following not only by gender pairs, but also by treatment condition. The first letter of the x-axis labels refers to the gender of the decision maker and the second letter to the gender of the matched participant. "C" and "T" denote *Control* and *Treatment*, respectively.



Notes: MM (males matched with males), MF (males matched with females), FM (females matched with males) and FF (females matched with females) refer to the four gender pairing cells. Within a gender pair, norm following behavior is split by treatment condition with behavior in *Control* (C) displayed always on the left and *Treatment* always shown on the right.

Figure A.5: Norm following by treatment, gender and partner gender cells

All differences, including the differences for males matched with males and females matched with females, by treatment are not significant. Hence, norm following does neither overall nor for the different gender pairs separately depend on the treatment.

Figure A.6 and A.7 are equivalent to Figure 7 in the main text, but split the sample by median norm conformity and rule breaking preferences, respectively. This clearly indicates that the "overshooting" observed for males is almost entirely driven by and very strong for males with below median *norm conformity*. In line with expectations, subjects with high norm conformity preferences do not significantly deviate from perceived norms.

I expect a similar relationship between rule-breaking preferences and norm following, just in the reverse direction. The concept measures — to a large extent — very similar aspects as *norm conformity*. This is what I find. Again, "overshooting" by males is mainly driven by above median "rule breakers". However, for *rule breaking*, also below median males do deviate from perceived norms. Further, also for females matched with males "rule breakers" significantly deviate from norms (p-value = 0.01, Wicoxon signrank test).

Not relying on median splits, but rather using the entire distribution of *rule breaking* and *norm conformity* in a linear regression model leads to the same inference. Overall, *norm conformity* is negatively (p-value = 0.02) and *rule breaking* positively (p-value < 0.01) linked to individual norm deviations. Also here, the is no significant interaction with *Treatment*.

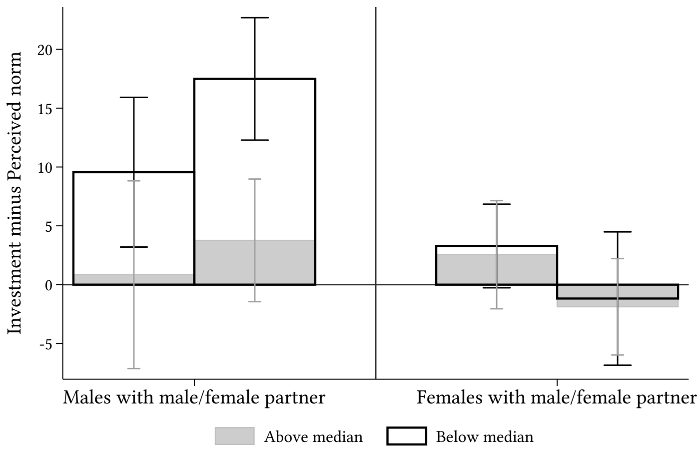


Figure A.6: Norm following by gender and matched gender for above and below median norm conformists

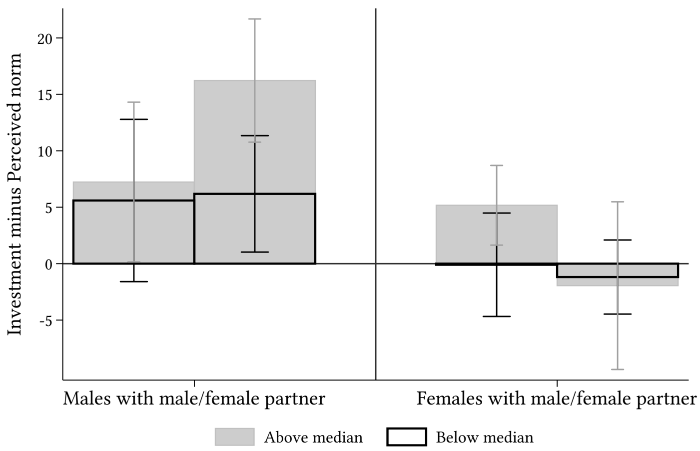


Figure A.7: Norm following by gender and matched gender for above and below rule breakers

D Seating Arrangement - *ONLINE APPENDIX ONLY*

The separating wooden walls between opposing seats in the laboratory were taken out before the start of the experiment to allow participants to identify their matched partner. Separating walls to the left and right remained (see Figure A.8).

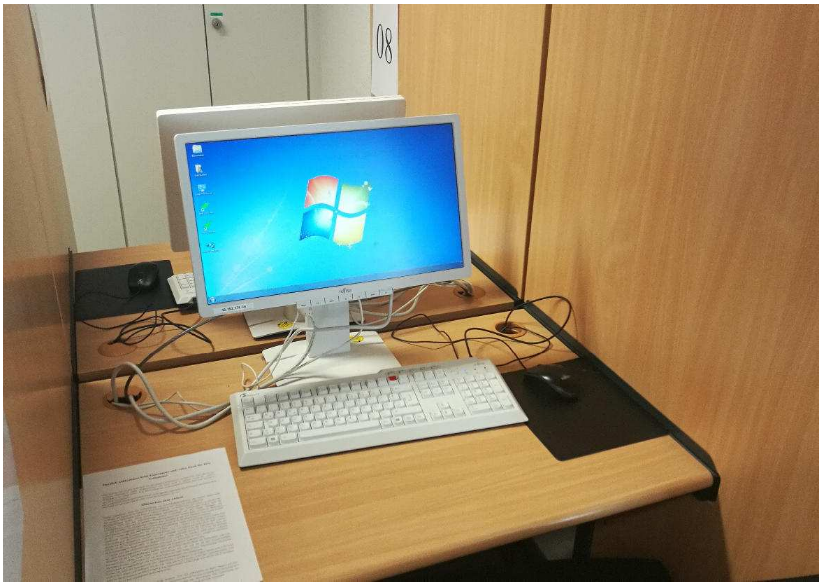


Figure A.8: Seat arrangement for matched participants

E Picture Consent Form - *ONLINE APPENDIX ONLY*

At the beginning of the following experiment, a picture will be taken of all participants. The anonymous picture will be saved on servers of MELESSA, can only be accessed by the research team and will be deleted after the end of the study.

The picture taken of you might be shown to other participants during the experiment.

I hereby state that I read and understood above-standing information and that I agree to the generation and use of my personal data as stated in this form.

F Instructions - *ONLINE APPENDIX ONLY*

The following passages are excerpts of the instructions that participants read on-screen. Text in italics denotes treatment manipulations and text in brackets denotes self-explaining comments. These accentuations are added for illustrative reasons and were not part of the original instructions.

[first screen]

Before the start of the experiment

In this experiment, you are matched with another participant. This participant sits in the seat vis-à-vis your own. Below you see the picture of your matched participant. During the experiment, the matching will become relevant. We will point out once it is relevant. Independent of the relevance of the matching, the picture of your matched participant will be shown on all decision screens.

[Picture of matched participant here]

[third screen]

Part 1

You receive 100 Taler for this decision. 100 Taler corresponds to €5. The exchange rate is 1 Taler=€0.05. You can now invest any integer amount between 0 and 100 Taler in a risky option. You will keep the amount that you do not invest.

With a probability of 50% the investment in the risky option will be successful. If it is successful, you receive 2.5 times the invested amount. If it is not successful, you lose the invested amount.

Your earnings from this part of the experiment are made up of the amount not invested, and (potentially) the earnings from your investment in the risky option. Your earnings will be converted to Euros at the end.

[for participants in Control:] *Your decision is anonymous.*

[for participants in Treatment:] *Your decision is not anonymous. Your matched participant will be shown your choice (not your earnings) on-screen at the end of the experiment.*

Please indicate now the amount in Taler (0-100) that you want to invest in the risky option.

[Small picture of matched participant here]

[eigth screen]

Part 2 — General information

In all following decision tasks in part 2 you will have to provide assessments and estimates regarding the behavior of other participants in part 1.

[for participants in Control:] *In part 1, you and in total 50% of the participants made anonymous choices. Your matched participant and the other half of the participants made non-anonymous choices. That is, these participants were told in part 1 that their choices in part 1 would be shown to their matched participants at the end of the experiment. That means that you will be shown the decisions of your matched participant at the end of the experiment (you will not see the earnings of your matched participant). Your decisions remain anonymous.*

[for participants in Treatment:] *In part 1, you and in total 50% of the participants made choices that will be shown to the matched participants at the end of the experiment. This was pointed out to you before making your decisions. Your matched participant and the other half of the participants made anonymous choices. Since their choices remain anonymous, they also*

were not told that their choices were non-anonymous. Hence, you will not be shown the decision of your matched participant at the end of the experiment.
In part 2, however, all participants make the same decisions. All these decisions will be anonymous for all participants.

[ninth screen]
Part 2 — Block 1
In block 1 of part 2 you are supposed to provide an estimate for the choice of your matched participant in part 1. In part 1 participants could invest any integer amount between 0 and 100 Taler (corresponds to between €0 and €5) in a risky option.
Please indicate now in Taler your estimate regarding your matched participant's invested amount.
In case this block will be payoff relevant, your earnings for this block will depend on your answer and the invested amount of your matched participant in part 1. If you deviate by 10 Taler or less from the actual invested amount of your matched participant, you will receive €5 for this block. If you deviate by more than 10 Taler, you will not receive any payment for this block.
What do you think your matched participant invested in the risky option (amount in Taler between 0 and 100)?
[Small picture of matched participant here]

[tenth screen]
Part 2 — Block 1
How confident are you in your answer from the screen before?
[Scale from 1 ("Not at all confident. I basically guessed randomly.") to 5 ("I am completely convinced that I gave the correct answer."). Other options were 2 ("I did not guess randomly, but I am still very uncertain."), 3 ("I am somewhat uncertain, but I had some idea of the correct answer."), and 4 ("I am rather certain that I gave the correct answer.").]

[This question on confidence was used after all following decision screens in exactly the same way. Therefore, hereinafter, these screens will not be shown again.]

[eleventh screen]
Part 2 — Block 2
In block 2 of part 2 you are supposed to provide an estimate for the choices of all participants in part 1. In part 1 participants could invest any integer amount between 0 and 100 Taler (corresponds to between €0 and €5) in a risky option.
Please indicate now in Taler your estimate regarding the average invested amount of all participants.
[Payoff information - similar to screen before]

What do you think did participants on average invest in the risky option (amount in Taler between 0 and 100)?
[Small picture of matched participant here]

[thirteenth screen]
Part 2 — Block 3
In block 3 of part 2 you are again supposed to provide an estimate regarding the choices of other participants. These will be different assessments though.
Later, four other participants will be shown your picture (they will also be told whether your choice in part 1 was anonymous or not). These participants will then, based on your picture, indicate what answer in part 1 would have been appropriate for you to make. They will not know your actual investment when making that assessment. Each of these four participants will be paid for his/her assessment if it does not deviate by more than 10 Taler from the average assessment of the other three participants seeing your picture.
You will receive €5 for this block, if you do not deviate by more than 10 Taler from the average assessment of the four other participants. If you deviate by more than 10 Taler, you will not receive any payment for this block.
What investment in Taler (between 0 and 100) do you think the other four participants deem appropriate for you; i.e., what do they think you should have invested?
[Small picture of matched participant here]

[fifteenth screen]
Part 2 — Block 4
One of the four other participants that will be shown your picture and that will indicate the appropriate investment for you will be your matched participant from part 1.
In block 4 of part 2 you are supposed to indicate the following: What do think your matched participant thinks would have been the appropriate investment that you should have made?
[Payoff information - similar to screens before]
What does your matched participant think would have been the appropriate investment that you should have made (in Taler between 0 and 100)?
[Small picture of matched participant here]

[eighteenth screen]
Part 2 — Block 5 — Decision 1
In the first decision you see the picture of another participant ([Here, the treatment condition of the participant in the picture was indicated. If "picture 1 participant" was in *Control* it said: *anonymous decision in part 1*; if "picture 1 participant" was in *Treatment* it said: *non-anonymous decision in part 1*]). You are supposed to indicate, what you think, what would have been the appropriate investment (in Taler between 0 and 100) that

the person in the picture should have invested in part 1.
Three other participants see the same picture and answer the same question that you will answer.
In case this decision becomes payoff relevant, you will earn €5 if your answer does not deviate by more than 10 Taler from the average answer of the other three participants. If your answer deviates by more than 10 Taler from the average answer of the three other participants, you will not receive any payment for this decision.
What investment in Taler (0 to 100) would have been the appropriate investment for the person in the picture; i.e., what amount should that person have invested?
[Small picture of picture 1 participant here]

[Screens for decision 2, 3, and 4 in block 5 of part 2 were equivalent to the screen for decision 1. They only referred to and showed the picture of picture 2 participant, picture 3 participant, and picture 4 participant (note that picture 4 participant always was the matched participant). Decision 4 in block 5 of part 2 concluded the main part of the experiment.]