STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2017–2018

Try your best or risk it all?

An experimental study of competition, effort, and risk-taking

Alice Hedda Nielsen 41047
Viking Waldén 23194

Abstract

While competition is praised as a means to increase effort, it may also give rise to negative externalities or trade-offs from increased risk-taking. Examining these effects, this thesis studies whether competition affects risk-taking and effort. To do so, we identify the impact of competition by isolating the effect of simultaneous actions aimed to gain a rival good (*competitive incentives*), from that of the simple presentation of a situation as competitive (*competitive framing*). We employ an experimental design where subjects' choices of effort and risk-levels are elicited in three randomised treatment groups where we manipulate the presence of each aspect of competition. In contrast to theoretical predictions and much of the experimental literature, we find no evidence that either incentives or framing impact risk-taking or effort, and no evidence of any correlation between the two. We also examine potential heterogeneity in responses to competition in three potential channels – gender, overconfidence, and risk-aversion – and find competitive incentives to increase effort among overconfident individuals. We identify no further evidence on differential responses to competition, for either effort or risk-taking. Overall, our results suggest choices of effort and risk-taking are on average no different in or outside competition.

Keywords: competition, risk-taking, effort, incentives, framing, experiment
JEL: C91, D91, D80, M52

Supervisor: Chloé Le Coq
Date submitted: 13 May 2018
Date examined: 29 May 2018
Discussant: Francesco Caslini
Examiner: Maria Perrotta Berlin

Acknowledgements

First and foremost, we thank our supervisor, Chloé Le Coq, for her valuable advice throughout this process. In addition, we thank Anna Dreber Almenberg for feedback on our experimental design, and for warning us about the low replication rate in economics. Like most students in our program, we would also like to thank Mark Voorneveld for taking the time to help us with simple mathematics. We also thank family, friends, and classmates for their honest feedback and continuous support.

Lastly, we recognise that without self-financing our experiment, this thesis would not have been possible. We thank the National Board of Study Support (CSN) for this possibility.

Table of contents

1 Introduction

Competition is ubiquitous in our lives and features in everything from adults competing for job offers and marriage partners, to children playing hide-and-seek in the school yard. Generally, the outcome of a competition is determined by both effort and risk-taking. A company's success is determined by the effort exerted by its employees as well as the riskiness of its investments in, for example, research and development. Similarly, politicians exert effort to craft good policies but sometimes pursue very risky strategies, such as going to war, in order to win an upcoming election (Downs and Rocke, 1994). These trade-offs between effort and risk-taking are seen everywhere from financial investors (Brown et al., 1996) to alpine skiers and race car drivers (Becker and Huselid, 1992; Föllmi et al., 2016).

More often than not, economists argue competition leads to efficient outcomes with high levels of effort. However, if competition also leads to high levels of risk-taking, this may create negative externalities. Excessive risk-taking by a company implies an increased likelihood of default with adverse impacts on employees as well as society, costs which are not necessarily borne by the company itself. Similarly, the politician that enters a war to win the election generates enormous negative consequences which affect both soldiers and civilians.

The outcome of a competition often also serves as a basis for remuneration in the workplace. Portfolio managers, whose work essentially concerns risk allocation, are often ranked by their returns and receive bonuses accordingly (Brown et al., 1996). Contracts like these are hailed in the principal-agent literature, which argues that competition for bonuses may prove efficient as workers are incentivised to exert optimal effort (Lazear and Rosen, 1981; Holmström, 1982). In particular, Relative Performance Evaluation (RPE) proposes bonus payments which depend on a worker's performance relative to that of her peers. This filters out the impact of common shocks, which, at least in theory, incentivises higher effort. In practice, however, RPE remains uncommon in observed compensation schemes (Jensen and Murphy, 1990; Frydman and Jenter, 2010). A possible explanation is the impact competition may have on agents' risk-taking. If competition leads agents to choose higher risk-levels than principals prefer, this creates a trade-off between principals' wish to limit risk-taking and agents' wish to minimise effort but retain a high probability of winning the bonus. This discrepancy may in turn explain why principals favour other incentive structures to reach preferred outcomes.

Given that risk-taking and effort are important for economic outcomes, and competition has an effect on economic decisions, it is surprising that the impact of competition on both risk-taking and effort has not been more extensively studied in the behavioural or experimental literature. To contribute to the analysis of these impacts and their implications, we ask the following research question:

Does competition between individuals affect risk-taking and effort when the outcome of the competition is determined by both?

While competition as a concept encompasses varied settings and actors, it is in essence a situation in which an individual strive against another individual to gain a rival prize. As such, competition consists of two aspects which may affect risk-taking and effort, together and separately. First, the rivalry of competition gives rise to *competitive incentives* as individuals simultaneously compete for a sole prize. Hvide (2002) and Gilpatric (2009), amongst other, theoretically show that competition incentivises agents to use effort and risk as substitutes, which leads to an equilibrium with decreased effort and increased risk-taking. In contrast, situations in which individuals perform against a predetermined threshold and without rivalry over the prize do not give rise to competitive incentives. Second, competition provides a *competitive frame* which in itself can affect effort and risk-taking through behavioural or cognitive factors, or by inducing social comparison. Economists often consider preferences to be stable (Stigler and Becker,

1977), but presentation of a situation as competitive is found to increase risk-taking even when there are no strict strategic incentives for it, i.e. when there is no rivalry over the good (e.g. Eriksen and Kvaløy, 2017; Kirchler et al., forthcoming). Empirical evidence also suggests a positive impact of competitive framing on effort (Falk and Ichino, 2006; Mas and Moretti, 2009, e.g.). Hence, in order to understand the potential impact of competition one must examine the potentially divergent effects of each aspect.

To disentangle the impact of competitive incentives from that of competitive framing, we employ an experimental design where subjects' choices of effort and risk-levels are elicited in three randomised treatment groups. Subjects gain a bonus payment if they outperform a treatment-specific target through which we manipulate the presence of the two aspects of competition. Our experimental design, presented in Section 2, constitutes an extension of present research. In particular, we expand upon previous studies by letting outcomes depend upon simultaneous choices of real effort and risk, by isolating the effects of competitive incentives and competitive framing, as well as by including a non-competitive setting. Additionally, in Section 3 we develop hypotheses for the impact of competitive incentives on effort and risk-taking. To do so, we compare the predictions of the competitive model in Hvide (2002) with our original expansion of Hvide's model to a non-competitive decision problem. In Section 4, we discuss theoretical and empirical evidence to develop hypotheses for competitive framing. Furthermore, Section 5 provides hypotheses regarding additive effects and substitution between risk-taking and effort.

While numerous studies have explored impacts of competition on effort (see review by Dechenaux et al., 2015) and on risk-taking (e.g. Eriksen and Kvaløy, 2017; Kirchler et al., forthcoming), to our knowledge only two experimental studies examine effects on both simultaneously. On the one hand, Andersson et al. (2017) find increased competitive incentive structures to increase risk-taking and to some extent also effort. On the other hand, Nieken (2010) find choices of risk-taking and effort to be negatively correlated under competition, in line with theoretical predictions by Hvide (2002). Importantly, both Andersson et al. and Nieken differ from our experiment as they neither include the effect of competitive framing nor employ a real effort task. We also go further by examining potential channels for the responses to competition beyond average impacts.

For our experiment, 417 subjects were recruited on Amazon Mechanical Turk (MTurk). The experiment was analysed using the pre-registered empirical strategy presented in Section 6. The results from our experiment, presented in Section 7, show that neither competitive incentives nor competitive framing affects effort exerted or risk taken by subjects on average. Our results are consistently robust to alterations, yet contrast both our predictions and much of the previous literature. As discussed in Section 8, our results are rather in line with recent studies which find no effect of competition on subsequent risk-taking (Filippin and Gioia, 2017), of competitive framing on effort (Gächter et al., 2017), or on substitution between effort and risk (Andersson et al., 2017).

Furthermore, while our results show no impact of competition on average, we find some evidence of within-group heterogeneity in responses to competition to impact effort choices. In Section 9 we explore three potential "channels" through which competition may affect risk-taking or effort – gender, overconfidence, or risk-aversion. We find individuals who overestimate their winning probability to exert significantly more effort under competitive incentives. However, no further evidence is found on any interaction between the impacts of competition on effort or risk-taking for gender or risk-aversion, and no further evidence for overconfidence. Finally, we critically discuss our combined results and their validity in Section 10 and provide a final conclusion in Section 11.

2 Experimental design

Using an experimental approach allows us to eliminate potential confounding factors, and focus on the causal relation of interest, i.e. does competition affect effort and risk-taking. Our design, related hypotheses and empirical strategy were preregistered at the Open Science Framework (OSF) prior to our experiment.[1]

We design a simple, one-round, experiment where subjects can exert effort and take risks in order to gain a bonus payment. The task was performed on MTurk, an online marketplace for crowdsourcing workers. Choosing a between-subjects method with three treatment groups, our experiment enables us to identify the additive as well as the separate impact of competitive incentives and of competitive framing. To increase the salience of competition, we limit competitive treatments to two players: one direct opponent with whom to compare oneself and only two possible outcomes – you win or you lose. We combine a widely used real-effort task, designed by Gill and Prowse (2012), and a risk-taking choice, building on Gneezy and Potters's (1997) commonly employed risk preference elicitation method.

In this section we present our experimental tasks, the experimental procedure, the treatments, the subject population and exclusion criteria, as well as the dependent and independent variables. Full experimental instructions can be found in Appendix A.

2.1 Elicitation of effort and risk-taking

In our experiment, the goal for subjects is to win a bonus payment. To do so, subjects need to collect as many points as possible through exerting real effort and investing in a lottery. Each subject's total number of points, y_i, determines if the subject meets a target, which varies across treatments, and thus receives a bonus payment of \$0.75 at the end of the experiment.

To elicit effort (e_i), subjects are asked to perform the slider task designed by Gill and Prowse (2012). Mimicking a costly activity which requires focus and persistence, the task consists of correctly placing sliders on a computer screen to predesignated values, as in Figure 2.1 below. Each subject is shown 60 sliders on-screen and given two minutes to place as many sliders as possible.[2] Each correctly placed slider (e_i) adds nine points to the subject's total. The position and number of sliders do not differ between subjects, and the predesignated values are varied between sliders to hinder subjects from using previous sliders as reference points. The sliders can be readjusted unlimited times within the time span.

Figure 2.1: Slider task

Importantly, the simplicity of our effort task allows us to capture effort rather than other motivations. As performance in the slider task is arguably not impacted by having completed similar tasks before, ability or pre-existing knowledge should play minor roles. For comparisons between treatments, any triggering of latent ability when in competition is also unlikely. In line with Gill and Prowse (2012), we argue that the task has a number of other desirable attributes: it is simple to communicate, it is identical across

[1] The pre-analysis plan, as attached in Appendix D, our data, as well as code for replicating regressions, tables and graphs can be found at `https://osf.io/8wy5b/`

[2] 60 sliders is well above what any subject completed in e.g. Gill and Prowse (2012).

individuals, and there is no scope for guessing or randomness in measured responses. The simplicity also circumvents intrinsic motivation; the task is tedious with no general meaning for society or contribution to something meaningful. Instead, it is effort exerted through focus and movement of the computer mouse or track pad that determines how many sliders one can place – and thus our measure of effort.

To elicit risk-taking, (r_i), we follow the method provided in Gneezy and Potters (1997). Subjects are given the opportunity to invest $r_i \leq 9$ points from each correctly placed slider in a lottery, prior to performing the slider task. The lottery has a probability p of yielding a positive outcome, $k \times r_i$ points, and a probability $(1-p)$ of yielding a negative outcome, 0 points. Investing in the lottery gives a higher possible total points, but also increases variance in total points. Hence, our measure of risk captures willingness to increase the variance in the final outcome and subjects can only make losses relative to their expectations and not actual losses.

To promote easy understanding of the lottery, we set $p = 0.5$ for all subjects and we set $k = 2.5$, as in e.g. Charness and Gneezy (2012). Note, this differs from some related studies, e.g. Andersson et al. (2017), where the task has no positive (or instead negative) returns to increased risk-taking. However, it is shown that individuals take risk even under negative incentives to risk-taking (Eriksen and Kvaløy, 2014), and setting $k = 2.5$ is thus unlikely to majorly impact directions of our results. As we do not seek to identify precise estimates for effort and risk levels, but rather potential differences between treatments, possible impacts on exact magnitudes are of lesser importance.

Together, the expected value of investing in the lottery is thus higher than that of not investing, as $0.5 \times 2.5 > 1$. In this scenario, a risk-seeking or risk-neutral person invests the full nine points, whereas a risk-averse invests less. While taking lottery risks does not imply any direct cost to the individual, it may imply a non-monetary disutility as higher risk is associated with increased uncertainty of outcome. As such, the expected total points of a subject i is:

$$E(y_i) = (pkr_i + (1-p) \times 0 \times r_i) \times e_i + (9 - r_i)e_i = (0.25r_i + 9)e_i$$

A subject gains the bonus if total points, y_i, is sufficiently large. As such, by combining a real effort and a lottery risk choice, we expand upon existing experimental tasks for eliciting the nature of the trade-off between effort and risk.

2.2 Experimental procedure

While the previous section outlined core experimental tasks, the procedure itself consists of three stages: an information phase, four experimental steps and finally a submission and subsequent payment phase, as illustrated in the timeline in Figure 2.2.

Subjects were recruited on MTurk and subsequently taken to an external website (Qualtrics) where the experiment was performed. Subjects were randomised into one of three treatments and received full information on eligibility requirements, experimental tasks, and potential payments. Following instructions, subjects were familiarised with the main experimental tasks through examples and by testing the slider task during an unlimited time period.

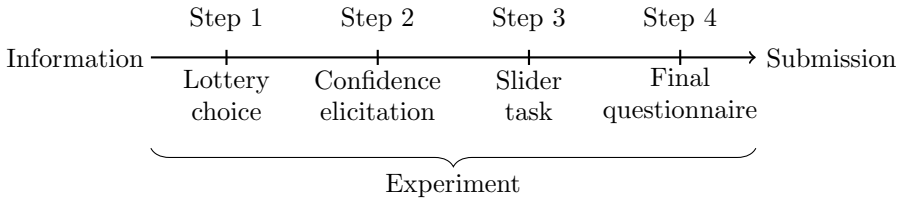


Figure 2.2: Timeline of experimental procedure

After having completed the trial task correctly, subjects proceeded to Step 1 of the experimental phase. Here, subjects were asked to choose how many points they wish to invest in the lottery. In Step 2 subjects were asked three questions to elicit their level of over- or underconfidence; expectations of their own and the average performance on the slider task, and of whether or not they will receive the bonus. Placing confidence elicitation between the lottery choice and the slider task avoids potential influencing of the risk-choice, while estimating subjects' beliefs prior to completion of slider task.

In Step 3, subjects performed the slider task during two minutes. In the final experimental step, Step 4, subjects answered a questionnaire about personal characteristics. The questionnaire is placed last in the experimental procedure in order to avoid biasing any measures of dependent variables by invoking stereotypical behaviors or beliefs. The questionnaire covered risk preferences, gender, age, and country of residence. For risk preferences, we used a 10-point Likert-scale where we asked subjects to rate their willingness to take risks on a scale from 0 ("Not willing to take risks") to 10 ("Very willing to take risks"). Beyond general risk preference we also asked for experience with starting an own company and with gambling in the past month, as motivated by findings in Dohmen et al. (2011) and Dreber et al. (2011). The full questionnaire can be found in Appendix A.2.

Finally, subjects received a unique answer code from Qualtrics which they filled in to MTurk. After submitting the answer code, subjects' responses were accepted or rejected, and subjects' total points were calculated and compared against treatment-specific targets. Subjects were subsequently paid \$0.5 if their response was accepted, or \$1.25 if it was accepted and they also met their target.

2.3 Treatments

In order to isolate the impacts of competitive incentives and of competitive framing through our experiment, we manipulate the nature of the value needed to win a bonus payment of \$0.75. As such, the sole factor which differs between treatments is the verbal information on the value, which takes three different forms to invoke competitive incentives and framing, only competitive framing, or neither aspect.[3] Table 2.1 outlines which aspects of competition are included in each treatment and Table 2.2 exemplifies how treatment information differs between groups.

In the first treatment, *Neutral threshold* (NT), subjects perform the task against a set threshold and are informed that they need to gain more points than the threshold value to get the bonus. As such, neither competitive incentives nor competitive framing is included – essentially, there is no aspect of competition. The value of the threshold, 273 points, is given by the median subject in the pilot study, see Section 2.4.3. Subjects in *Neutral threshold* are simply informed of the numeric value but not the origin of the threshold.

Table 2.1: Treatment incentives

	Neutral threshold	Direct competition	Competitive threshold
Competitive incentives	No	Yes	No
Competitive frame	No	Yes	Yes

[3] As such, all other information, all tasks, and all questions, apart from the key treatment information and related examples, remain identical across treatments.

Subjects in the second treatment, *Direct competition* (DC), face a simultaneous contest where the counter-party is an unnamed player and the prize is rival, i.e. a "true" competition. Concretely, in order to win the bonus, subjects need to gain more points than their counter-party. As such, the treatment invokes competitive framing, as subjects are told they need to win over an opponent, as well as competitive incentives, as there is another actor competing for the same prize. *Direct competition* is thus representative of the competition one encounters in everyday life, and for this reason similar treatments are widely used to test competition in experimental studies, such as Niederle and Vesterlund (2007) and Buser and Dreber (2016).

Table 2.2: Key Treatment Information

	Target specification
Neutral threshold (NT)	"You will perform against a **threshold**. If you collect **more points** than the threshold which is 273 points, you are paid a **bonus payment** of \$0.75, in addition to a completion payment of \$0.50"
Direct competition (DC)	"You will compete against **another participant**. If you collect **more points** than your opponent's points, you are paid a **bonus payment** of \$0.75, in addition to a completion payment of \$0.50."
Competitive threshold (CT)	"You will compete against **another participant**. If you collect **more points** than your opponent who got 273 points, you are paid a **bonus payment** of \$0.75, in addition to a completion payment of \$0.50."

While the first two treatments allow us to isolate the differential impacts on risk-taking and effort of no competition versus competition, a third treatment is needed in order to separate the effects of incentives from those of framing. As such, subjects in the *Competitive threshold* (CT) treatment face competitive framing but not competitive incentives, i.e. the subject performs the task against a counter-party, but uncertainty over the value required to win is removed by including a threshold.

Concretely, in *Competitive threshold*, subjects are informed the threshold to beat is 273 points and that it comes from another, unnamed subject. Practically, the threshold is drawn from the same pilot study as in *Neutral threshold*. As such, all information is truthful but as more information is given, competitive framing is also invoked. Competitive performance against a previous subject's score has been utilised to induce impacts of competition (see e.g. Apicella et al., 2011; Amir et al., 2012) or for asynchronous experiments (Straub, 2017). As such, by including both the *Direct competition* and the *Competitive threshold* treatments we not only can disentangle the impact of competitive incentives, but we can also test for potential differences between two common methods of creating competition in experiments.

As such, to analyse relative impacts of competition without adjusting the general goal of subjects to gain a bonus payment, we manipulate the presence of competitive incentives and framing but retain a performance target, whether this is a set threshold or the unknown output of another. As such, our conclusions are silent on effort and risk-taking in tasks that are not commonly used for bonus payments, e.g. a piece-rate task where also the target is removed.

2.4 Subjects

As mentioned, subjects were recruited on MTurk, a crowdsourcing platform for a wide range of simpler assignments with low payments. An assignment on MTurk is known as a "human intelligence task" (HIT) and presented with a title and a short description.

When viewing our HIT, subjects were shown an introductory page inviting them to take part in the study. After reading the general information about total time of task, potential payment, and requirements for completion, they accepted the invite and were taken to Qualtrics where the experimental procedure began.

Platforms like MTurk have been used by companies for temporary use of human skills, such as tagging images or writing reviews. Increasingly researchers have recognised its potential as a forum for running incentivised social science experiments (see e.g. Horton et al., 2011). In particular, MTurk workers are commonly recruited for short time periods and thus paid relatively low amounts, enabling a substantially higher statistical power than laboratory experiments with equal budgets. In 2017, average and median hourly wages were $3.18 and $6.19, respectively (Hara et al., 2018), whereas a normal experimental task requires less than 10 minutes. Additionally, MTurk gives access to a large subject pool, with an average of 7,300 workers available to sample at any given moment (Stewart et al., 2015).

2.4.1 Exclusion criteria

MTurk allows for approval or rejection of each submitted assignment, permitting us to impose three main exclusion restrictions for the subject pool. As detailed in our pre-analysis plan, as attached in Appendix D, we exclude the following groups of observations from payment and from our results:

1. **"Attention Checks"**: With an online workforce, and in particular when basing treatment on verbal formulations, capturing whether a subject pays attention or not is of key importance (Straub, 2017). Subjects were therefore asked three simple attention questions to which correct answers were found in nearby text. Subjects who failed two or more questions were excluded from the sample and not paid for their participation. The questions were clearly marked and subjects were informed of the exclusion criteria when accepting the HIT. The attention checks also allow us to screen subjects' ability to understand English.
2. **Multiple participation**: To avoid contamination between treatments, subjects were told they may only participate once. However, for those who after all performed the task multiple times, all but the first submission were rejected. Matching of submissions to subjects was done using IP addresses collected by Qualtrics, and the exclusion restriction was also applied to responses recorded as "in progress" rather than finally submitted. As such, subjects who started the task, stopped it, restarted and then submitted were excluded.
3. **Answer code**: In order to pay subjects for their participation we provided each subject with a unique answer code in the submission phase on Qualtrics. Using this code, subjects' answers on Qualtrics were matched with their accounts on MTurk for payment. As such, submissions to MTurk which indicated an invalid answer code were rejected.

2.4.2 Power analysis and sample size

In light of the concerns about the lack of the replicability in experimental literature (Open Science Collaboration, 2015; Camerer et al., 2016), we design our experiment to have an 80% statistical power. Drawing upon effect sizes found in three of our closest related studies, which, unlike most, also made needed measures available in their papers, we used power calculations to identify required sample size for our experiment.[4] As such, we estimated a minimum sample size of around 130 subjects in each treatment based on the effect size in Eriksen and Kvaløy (2017), 70 based on the effect size in Filippin and

[4] Calculations were carried out with the power calculator provided by HyLown Consulting (2018).

Gioia (2017), and 65 based on the effect size in Andersson et al. (2017). Consequently, we chose a sample size of approximately 140 subjects in each treatment group, to leave some margins of error.

2.4.3 Pilot study

In addition to the main study conducted on 7th to 9th of April, 2018, a pilot study was run on 31st of March to 2nd of April, 2018 to generate threshold values. 30 subjects were recruited on MTurk and performed the same experimental procedure, but which included only the *Direct competition* treatment. From this, median total points was chosen to represent the threshold value in the main study.[5] This way, our chosen value provides a threshold that is both plausible to reach and with an expected 50% winning rate in the two threshold treatments, thus mimicking a real-world, two-person competition. The responses of pilot subjects are not included in the main study as they were not randomised into treatments. Additionally, the pilot study allowed us to test the design and phrasings, with subsequent minor changes which are outlined in the pre-analysis plan in Appendix D.

3 The impact of competitive incentives

Through our experimental design we isolate the impact of competitive incentives by comparing effort and risk-taking in the *Direct competition* treatment and in the *Competitive threshold* treatment. While both treatments are presented as a competition against another subject, i.e. competitive framing, only *Direct competition* entails simultaneous decisions by two subjects aimed to gain a rival good, i.e. competitive incentives. These incentives are different from the *Competitive threshold* treatment, where gaining the prize is not exclusive to one of the two subjects and there is also no uncertainty over the value required to gain it. To derive predictions for the impact of competitive incentives we lean on theoretical models for competition, risk-taking, and effort.

The study of competitive incentives builds from the seminal paper by Lazear and Rosen (1981) which derives that, for risk-neutral agents, a two-player competition can increase agents' effort and lead to optimal contracts. Most subsequent studies have focused on either the impact of competition on effort (e.g. Holmström, 1979, 1982; Nalebuff and Stiglitz, 1983) or on risk-taking (e.g. Dekel and Scotchmer, 1999; Gaba et al., 2004; Tsetlin et al., 2004). More recently, the literature has extended to models for the impact on both effort and risk-taking. Hvide (2002) models competitive incentives in winner-takes-all contests, predicting agents will prefer taking high risk and exerting no effort, as this gives a 50% chance of winning at no effort cost. Nieken (2010) shows this result to be robust to a sequential model where symmetric agents make choices first over correlated risks and then over effort levels. In connection with the model in Hvide (2002), the predictions are also shown to be robust to changes in core assumptions; to risk-averse agents,[6] to asymmetry in ability, as well as to more than two agents. The only caveat is that with large asymmetries in ability, a low-risk equilibrium strategy may be reached.[7] However, as our design considers a simple real-effort task, with little room for asymmetries in ability, this prediction is unlikely to hold.

[5] As the pilot included 30 subjects, the upper of the two median values was chosen.

[6] For changed risk preferences to yield the same predictions, monotonic utility functions of agents is the only requirement.

[7] For a strong agent, decreasing risk-taking increases the agent's probability of winning as effort differences have relatively larger impacts. However, doing so also increases equilibrium effort.

Studying a similar two-player, two-stage contest but with agents of asymmetric ability, Kräkel and Sliwka (2004) also find similarly symmetric equilibria. Here, agents coordinate on either high or low risk-taking, and equilibrium effort is increasing in risk-taking in cases of large asymmetry or large prize differentials. In another paper, Kräkel (2008) however, derives asymmetric equilibria also in two-player contests, with risk-averse agents who are asymmetric in cost of effort or in preferences for winning. The agent with higher ability can prefer both relatively higher and lower risk, depending on the type of asymmetry. However, the model has limited applicability to experimental studies such as ours, as potential asymmetry in subjects' utility functions is unobservable. It is also reasonable to assume that subjects in our experiment perform the task in the belief that their counter-party is in a similar situation, i.e. the competition is symmetric. As such, while some pairs may be asymmetrically matched, in particular regarding preferences for winning, the competition is symmetric on average.

Furthermore, Gilpatric (2009) models a choice of risk in situations where there is a cost of increasing risk, such as when workers push the limits of firms activities to pursue riskier projects. However, many other cases, such as choosing to invest in a risky or safe asset, carry no significant cost of increasing the level of risk. In the limited application of Gilpatrics' model to the case of two-players, the predictions become similar to those of Hvide (2002). In more extensive applications, Gilpatric proposes an alternative to predictions of infinite risk-taking in models with more than two agents. Here, contest organisers can set prize differentials to incentivise any combination of effort and risk-taking, yet this result is not applicable to our two-player design.

In light of the theoretical literature, we build on the work of Hvide (2002) to analyse the effect of competitive incentives in our experiment. Hvide's model provides clear, and robust theoretical foundations for deriving hypotheses for our experiment with simultaneous, two-player competition including both effort and risk-taking. Hence, we first provide the reader with Hvide's model, explain any deviations we make from it, as well as expand on his proof of the uniqueness of the equilibrium. We then expand his model to a decision problem where an agent faces a fixed output threshold for winning a bonus payment. Finally, we compare the predictions of the two models to generate our hypotheses for competitive incentives. By doing so, we show how simultaneous competition for a rival good implies different incentives to evaluation against a fixed threshold, and thus result in different choices of effort and risk-taking.

3.1 A model of direct competition

Modelling our direct competition (DC) setting we lean on the model of competition, effort and risk-taking in Hvide (2002). Here, two rational, risk-neutral and homogeneous agents i, j simultaneously compete for a prize with value $W \in (0, \infty)$. The prize W could be a bonus, a promotion, or another physical good the agent gains utility from. Our analysis focuses on the choices of agent i, but due to symmetry the equivalent analysis holds for agent j.

In order to win the prize, agent i needs to produce higher output than her competitor. To affect her output, the agent exerts effort $e_{i,DC} \in [0, \infty)$ and chooses a "risk-level" $\sigma^2_{i,DC} \in [0, \infty) \cup \{\infty\}$, which is the variance of a stochastic variable $\varepsilon_{i,DC} \sim N(0, \sigma^2_{i,DC})$. $\varepsilon_{i,DC}$ and $\varepsilon_{j,DC}$ are independently drawn.

Two things are worth noting about the risk-level, $\varepsilon_{i,DC}$. Firstly, we simplify Hvide (2002) by allowing the variance to be equal to zero, as is done by amongst other Nieken (2010). Secondly, the normal distribution is not defined for $\sigma^2_{i,DC} = \infty$. However, we include this option as well, representing a choice which gives a 50% probability of $\varepsilon_{i,DC}$

being infinitely positive and a 50 % probability of it being infinitely negative.[8] One can think of this option as an all-or-nothing bet which gives equal chances of winning and of losing, regardless of your effort choice.

Given the agent's choice of effort and risk, output $y_{i,DC}(e_{i,DC}, \sigma^2_{i,DC})$ is realised as

$$y_{i,DC} = e_{i,DC} + \varepsilon_{i,DC}\left(\sigma^2_{i,DC}\right)$$

In order to win over her opponent and thus prize W, the agent needs to reach a higher output than her competitor, i.e. $y_{i,DC} > y_{j,DC}$. If output of the agent is below her competitor's, the agent loses and receives 0.[9] If the output of agent is equal to her competitor's, i.e. $y_{i,DC} = y_{j,DC}$, the outcome of the competition will be determined by a coin flip.

As such, an agent's probability of winning over her competitor is given by

$$\begin{aligned}\mathbb{P}\left(y_{i,DC} \geq y_{j,DC}\right) &= \mathbb{P}(e_{i,DC} - e_{j,DC} > \varepsilon_{j,DC} - \varepsilon_{i,DC}) \\ &= F_{\sigma^2_{i,DC},\sigma^2_{j,DC}}(e_{i,DC} - e_{j,DC},)\end{aligned}$$

where $F(\cdot)$ is the cumulative distribution function (CDF) of $\varepsilon_{DC} = \varepsilon_{j,DC} - \varepsilon_{i,DC}$ with $\mathrm{E}(\varepsilon_{DC}) = 0$ and $\mathrm{Var}(\varepsilon_{DC}) = \sigma^2_{i,DC} + \sigma^2_{j,DC}$.

When exerting effort to increase output, the agent also experiences a cost $V(e_{i,DC})$ where $V(0) = V'(0) = 0$ and $V'(e_{i,DC}) > 0$ for $e_{i,DC} = 0$. As such, the utility function of the agent is given by her probability of reaching a higher output than her competitor, the size of the prize and the cost of exerting effort

$$\begin{aligned}U_{i,DC}(e_{i,DC}, \sigma^2_{i,DC} \mid W, e_{j,DC}, \sigma^2_{j,DC}) &= \mathbb{P}\left(y_{i,DC} \geq y_{j,DC}\right) - V(e_{i,DC}) \\ &= F_{\sigma^2_{i,DC},\sigma^2_{j,DC}}(e_{i,DC} - e_{j,DC})W - V(e_{i,DC})\end{aligned}$$

The first order condition for the optimal choice of effort then becomes:

$$\frac{\partial U_{i,DC}}{\partial e_{i,DC}} = f(e_{i,DC} - e_{j,DC}) \times W - V'(e_{i,DC}) = 0 \tag{1}$$

Following Hvide (2002), we derive the first proposition:

Proposition 1. *The unique Nash equilibrium is for both agents to exert zero effort,* $e_{i,DC}, e_{j,DC} = 0$*, and take infinite levels of risk,* $\sigma_{i,DC}, \sigma_{j,DC} = \infty$*.*

Proof. Suppose agent j chooses $e_{j,DC} = 0$ and $\sigma^2_{j,DC} = \infty$. This will make agent j win with a 50% probability and, conversely, agent i will win with 50% probability, regardless of which strategy $\{e_{i,DC}, \sigma^2_{i,DC}\}$ she chooses. Essentially, as the risk-level chosen by agent j is so high, neither the effort nor the risk-level chosen by agent i can affect her winning probability.

A best response to the strategy of agent j, is to minimise effort, choosing $e_{i,DC} = 0$ and maximise the risk-taking, choosing $\sigma_{i,DC} = \infty$. Firstly, minimising effort is optimal, as effort implies a direct disutility. Secondly, as agent j chooses infinite risk-taking, the risk choice of agent i is irrelevant as it cannot affect the distribution probability of winning when $\sigma_{j,DC} = \infty$. As risk-taking is cost-less, $\{e_{i,DC} = 0, \sigma_{i,DC} = \infty\}$ is a best response to agent j's strategy, and $\{e_{i,DC} = e_{j,DC} = 0, \sigma^2_{i,DC} = \sigma^2_{j,DC} = \infty\}$ is a Nash equilibrium.

To prove the uniqueness of this equilibrium, six cases need to be evaluated. We present here two cases which are not considered by Hvide (2002). The remaining four cases are considered in Appendix B.1 for completeness. As such:

[8] While formally incorrect, we follow Hvide (2002) and simplify notation by writing $\sigma = \infty$ for a sigma which is infinitely large. Hvide, however, is quiet about how he squares this with his assumption of a normal distribution.

[9] Hvide (2002) includes a prize for the loser, but for simplicity we set this to zero. As such, one can view our winning prize, W, also as the prize differential between winning and losing prizes.

(1) $e_{i,DC} = e_{j,DC} = 0$ and $\sigma^2_{i,DC} = \infty, \sigma^2_{j,DC} < \infty$: Even if $\sigma^2_{j,DC} < \infty$, agent i has a 50% chance of winning when choosing $\sigma^2_{i,DC} = \infty$. However, by choosing positive effort, $e_{i,DC} > 0$ and no risk $\sigma^2_{i,DC} = 0$ the probability of winning will be larger than 50%. This is, again, because the probability of winning is strictly increasing in effort if $\sigma_{i,DC}, \sigma_{j,DC} < \infty$.[10] Hence, $\{e_{i,DC} = e_{j,DC} = 0, \sigma^2_{i,DC} = \infty, \sigma^2_{j,DC} < \infty\}$ is not a Nash equilibrium.

(2) By symmetry and (1) above, $e_{i,DC} = e_{j,DC} = 0$, $\sigma^2_{i,DC} < \infty, \sigma^2_{j,DC} = \infty$ is not a Nash equilibrium.

As such, $e_{i,DC}, e_{j,DC} = 0$ and $\sigma^2_{i,DC}, \sigma^2_{j,DC} = \infty$ is the unique Nash equilibrium.

□

Intuitively, by increasing risk, differences in effort between the two agents become less important as the increase in variance implies more noise in output, regardless of effort level. In turn, as effort becomes less important to the probability of winning, the incentive to take effort becomes weaker. (Hvide, 2002, p.884) Hence, the model of direct competition as proposed by Hvide results in an equilibrium which is characterised by no effort and very high risk-taking. By introducing a choice of both effort and risk, competition is shown to not necessarily lead to efficient outcomes with high rates of productive effort, as suggested by i.a. Lazear and Rosen (1981).

3.2 A model of threshold evaluation

To isolate the impact of incentives from other aspects of competition, we model a theoretical story for performance against a fixed threshold. In reality, workers face situations where a certain level of output is required, to e.g. gain a bonus payment, but where the required output level is independent of other individuals' output. In our experiment, subjects in both the *Neutral threshold* and *Competitive threshold* treatments face such evaluation.

We generate predictions for threshold evaluation by expanding the model in Hvide (2002) to a one-agent decision-problem. Retaining his core assumptions, we now model a scenario, T where the agent faces a commonly known threshold $\mathcal{T} \in (0, \infty)$ which she must surpass to win the commonly known prize, $W \in (0, \infty)$. The agent chooses effort $e_{i,T} \in [0, \infty)$ and a risk-level, i.e. the variance $\sigma^2_{i,T} \in [0, \infty) \cup \{\infty\}$ of a stochastic variable $\varepsilon_{i,T} \sim N(0, \sigma^2_{i,T})$. As in the model of a direct competition, we allow the variance to also be zero. Also here, choosing infinite variance yields a 50% probability of the outcome reaching above the threshold and the agent winning the prize, and a 50% probability of it falling below the threshold and the agent gaining nothing.

Given the agent's choice of effort and risk-level, output $y_{i,T}(e_{i,}, \sigma^2_{i,T})$ is realised as

$$y_{i,T} = e_{i,T} + \varepsilon_{i,T}\left(\sigma^2_{i,T}\right)$$

In order to win the bonus prize W, the agent needs an output at or above the threshold $\mathcal{T}$, i.e. $y_{i,T} \geq \mathcal{T}$.[11] If instead, the output of the agent is below the threshold, i.e. $y_{i,T} < \mathcal{T}$, the agent receives 0. As such, an agent's probability of winning the prize is given by

$$\begin{aligned}\mathbb{P}(y_{i,T} \geq \mathcal{T}) &= \mathbb{P}(e_{i,T} + \varepsilon_{i,T}(\sigma^2_{i,T}) \geq \mathcal{T}) = \mathbb{P}(\varepsilon_{i,T}(\sigma^2_{i,T}) \geq \mathcal{T} - e_{i,T}) \\ &= 1 - \mathbb{P}(\varepsilon_{i,T}(\sigma^2_{i,T}) \leq \mathcal{T} - e_{i,T}) = 1 - F_{\sigma^2_{i,T}}(\mathcal{T} - e_{i,T})\end{aligned}$$

[10] This follows from the assumptions of $\mathrm{E}(\varepsilon_{i,DC}) = 0$ and of agent i winning if $y_{i,DC} > y_{j,DC}$.

[11] In the direct competition model, a tie between agents is determined by a coin flip. If this was the case here, the decision problem would lose a solution as maximisation would be over an open interval.

where $F(e_{i,T}, \sigma^2_{i,T} \mid \mathcal{T})$ is the cumulative distribution function (CDF) of $\varepsilon_{i,T}$ with $\mathrm{E}(\varepsilon_{i,T}) = 0$ and $\mathrm{Var}(\varepsilon_{i,T}) = \sigma^2_{i,T}$. Furthermore, the agent has the same utility function as in Section 3.1

$$\begin{aligned} U_{i,T}(e_{i,T}, \sigma^2_{i,T} \mid W, \mathcal{T}) &= \mathbb{P}\left(y_{i,T}(e_{i,T}, \sigma^2_{i,T}) \geq \mathcal{T}\right) W - V(e_{i,T}) \\ &= (1 - F_{\sigma^2_{i,T}}(\mathcal{T} - e_i))W - V(e_{i,T}) \end{aligned}$$

Following this, we derive our second, original, proposition:

Proposition 2. *The agent will maximise her utility by choosing* $\{e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0\}$ *if and only if* $\frac{W}{2} \geq V(\mathcal{T})$*. Otherwise, the agent will choose* $\{e_{i,T} = 0, \sigma^2_{i,T} = \infty\}$*.*

Proof. To find the optimal choice for the agent, several combinations of risk-taking and effort must be explored. With regards to effort, there are four possible cases: (i) $e_{i,T} = 0$, (ii) $e_{i,T} \in (0, \mathcal{T})$, (iii) $e_{i,T} = \mathcal{T}$ and (iv) $e_{i,T} > \mathcal{T}$. As the agent also chooses the risk-level, we need to investigate the three cases where (i) $\sigma^2_{i,T} = 0$, (ii) $\sigma^2_{i,T} \in (0, \infty)$ and (iii) $\sigma^2_{i,T} = \infty$. In total, there are twelve possible cases to study, which are illustrated in Table B.2.1.

Deriving the combination which maximises agents' utility, we can rule out all but two cases: $\left\{e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0\right\}$ and $\left\{e_{i,T} = 0, \sigma^2_{i,T} = \infty\right\}$. In essence, the agent chooses between a "safe" or a "risky" strategy. For a safe strategy, she can choose to exert effort to reach the target. If so, she would want to exert exactly as much effort as needed and take no risk, as any risk implies a negative outcome is possible. For a risky strategy, she can instead choose high risk and no effort, as high risk gives a 50% chance of winning and any effort level cannot increase this. These two options dominate all other possible strategies, as shown extensively in Appendix B.2. To find the optimal choice for agent i, we seek the option which yields the highest utility. Thus, agent i chooses $\left\{e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0\right\}$ if and only if the following condition holds:

$$\begin{aligned} & U(e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0 \mid W, \mathcal{T}) \geq U(e_{i,T} = 0, \sigma^2_{i,T} = \infty \mid W, \mathcal{T}) \\ \iff \quad & \mathbb{P}\left(y_{i,T}(e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0) \geq \mathcal{T}\right) W - V(\mathcal{T}) \geq \mathbb{P}\left(y_{i,T}(e_{i,T} = 0, \sigma^2_{i,T} = \infty) \geq \mathcal{T}\right) W - V(0) \\ \iff \quad & W - V(\mathcal{T}) \geq 0.5W - V(0) \\ \iff \quad & \frac{W}{2} \geq V(\mathcal{T}) \end{aligned}$$

Hence, the agent chooses $\{e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0\}$ if and only if $\frac{W}{2} \geq V(\mathcal{T})$. Otherwise, she chooses $\{e_{i,T} = 0, \sigma^2_{i,T} = \infty\}$.

□

The intuition behind the result is straightforward: If the marginal benefit of receiving the prize with certainty $\left(\frac{W}{2}\right)$ is large enough compared to the cost of exerting enough effort to do so $(V(\mathcal{T}))$, agent will choose $\left\{e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0\right\}$. Otherwise, the agent will settle for a 0.5 probability of winning the prize W by choosing $\left\{e_{i,T} = 0, \sigma^2_{i,T} = \infty\right\}$.

3.3 Hypotheses

By comparing Proposition 1 and 2, we derive the first set of hypotheses for the impact of one aspect of competition – competitive incentives – on subjects in our experiment. While the theoretical models do not replicate our experimental task exactly, e.g. we include a monetary incentive from investing in the lottery as discussed previously, the models provide the intuition behind the direction of impacts of competitive incentives. As such, according to Proposition 1, subjects will exert zero effort and take infinite risk if facing direct competition (DC), as doing so ensures a 50% probability of winning with certainty. Increasing effort does not impact the probability and decreasing risk-taking lowers the probability. In contrast, Proposition 2 states that agents in threshold

evaluation (T) will exert high effort and take no risk as long as the cost of effort is low enough compared to the potential gain. Doing so implies a 100% probability of winning, and increasing risk-taking would only lower the probability. Increasing effort on the other hand increases the cost but not the probability of winning. As a result, we hypothesise that:

H1 *Direct competition leads to a higher level of risk-taking than threshold incentives:* $\sigma^2_{i,DC} > \sigma^2_{i,T}$

H2 *Direct competition leads to a lower level of effort than threshold incentives:* $e_{i,DC} < e_{i,T}$

Importantly, within the threshold evaluation model we assume that it does not matter if the threshold is competitively framed – in either case, facing a fixed threshold and no rivalry in prize does not yield competitive incentives. Instead, threshold evaluation leads to lower risk-taking and higher effort, given the cost of effort is low enough compared to the potential gain. For our experiment specifically, we therefore expect a subject in our direct competition treatment, *Direct competition* (DC), to take higher risks and lower effort than a subject in either of our threshold evaluation treatments, *Competitive threshold* (CT) and *Neutral threshold* (NT).

4 The impact of competitive framing

Many of our daily choices are impacted by factors beyond strictly strategic incentives, not to mention decisions taken in competitive settings. In competitive situations, behaviour may be affected by factors which are inherent to competition, but which are distinct from the simultaneous rivalry outlined in Section 3. We define *competitive framing* as the presentation of a situation as competitive, which gives rise to factors such as cognitive or behavioural biases and social comparison. As such, subjects in the *Competitive threshold* treatment may choose different risk-taking and effort levels, relative to those in the *Neutral threshold* treatment. While our experiment does not discern between the different explanations for why competitive framing may have an effect, we examine the extent to which it does impact individuals' choices. The following section provides a review of related literature and develops hypotheses for the impact on effort and risk-taking.

4.1 Theoretical and empirical evidence

As defined, competitive framing may impact decisions either through evoking biases from previous experiences or from heuristics, or through altering a decision problem to a social action. Moreover, theoretical and empirical evidence on the two explanations jointly suggest competitive framing is separable and different to competitive incentives. Discussing the two explanations streams in turn, we argue the competition's impact extends to situations with only a frame of rivalry but no actual rivalry over the prize.

Providing a two-pronged explanation, Eriksen and Kvaløy (2017) argue competition's impact on decisions over amongst other effort and risk-taking may have both behavioural and cognitive explanations. First, behavioural science suggests a "contingency of reinforcement" (see for example chapter 6 in Skinner, 1969) may arise from previous experiences of positive outcomes from increasing one's risk-taking and effort when in competition. In subsequent situations, the simple framing of a setting as competitive, but where strategic incentives are in fact not present, leads individuals, like Pavlovian dogs, to exert higher effort and to take larger risk.

Second, a cognitive scientist, however, may posit another explanation for the same phenomenon. If winning a competition is more commonly associated with high effort and

risk-taking, it may create an "availability heuristic" for competitive environments (Tversky and Kahneman, 1973). As for "contingencies of reinforcement", facing another competitively framed situation such heuristics implies increased risk-taking and effort. However, now because it is simply the reaction that comes first to mind, and thus the simplest to act upon. While unable to disentangle the behavioural and cognitive explanations, Eriksen and Kvaløy (2017) test their joint impact in an experimental study of risk-taking in competition. Their results suggest that increased competitive framing leads to higher risk-taking, even when there are no, or even negative, incentives to take risk. As such, Eriksen and Kvaløy's findings provide a theoretical and empirical indication that risk-taking is affected by factors beyond competitive incentives.

The second aspect of competitive framing, social comparison, builds on the fact that individuals, in contrast to standard economic assumptions, do not only care about their absolute well-being. Instead, individuals are also concerned with how they compare to others (Fehr and Schmidt, 1999), which may impact their decisions over effort and risk-taking. Inherently, competition results in winners and losers – a salient experience, in particular in winner-takes-all contests such as our design. As such, competition invokes a "social reference point", where one's preferences and choices depend on the relative ranking of one's own outcome to that of another. For example, if subjects are loss averse, as individuals commonly are, comparisons against a (social) reference point can have large implications in competition (Kahneman and Tversky, 1979). In theoretical models of social comparison, Koszegi and Rabin (2007) model endogenously determined reference points, shaped by expectations over both absolute and relative wealth.

The impact of social comparison on effort is widely studied. For example, Falk and Ichino (2006) and Mas and Moretti (2009) show presence of peers affects productivity and pace of workers, in particular among the least productive workers. Falk and Ichino employ high school students to perform menial tasks, finding those who work alone in a room to exert less effort than those who work independently, but in the same room as another. Mas and Moretti find similar result in an observational study of supermarket cashiers. However, invoking social comparison does not require presence of another. Instead, laboratory studies by Gächter and Thöni (2010), Gächter et al. (2013), and Alain et al. (2014) use three-person gift-exchange games where subjects are informed about the wage of another "worker". Here, the results indicates horizontal comparison of one's wage to that of others impacts effort, perhaps as the wage difference is interpreted as a social hierarchy to which the subject conforms. In particular, individuals who receive a relatively lower wage decrease their effort, a result which holds for information on the other's effort provision (Thöni and Gächter, 2015). Together, this gives a first indication that the simple observation or knowledge of others affects individuals' effort, even without any actual rivalry.

Contrasting these results, Gill and Prowse (2012) present a case for social comparison instead leading to reduced effort as a consequence of disappointment aversion. Similar to our *Competitive threshold* treatment, Gill and Prowse's model, and experimental design, test the impact of informing subjects about a previous subject's performance which subjects need to surpass for bonus payment. Competitive framing is here predicted to lead individuals to exhibit loss aversion around a social reference point, which depends upon both own and another's performance. Testing the predictions in a sequential, multi-round application of the slider task used in this study, the results in Gill and Prowse (2012) confirm high first-mover effort discourages second-movers from exerting effort in fear of losing and experiencing only the effort cost. However, the results fail to replicate in Gächter et al. (2017).

The impact of social comparison on risk-taking is only more recently studied. In an expansion on models for social reference points, Schmidt et al. (2015) model the link between social comparison and risk-taking in a lottery task. The model predicts social comparison to increase risk-taking, a result which is confirmed in Schmidt et al.'s related class-room experiment. Also here social comparison is invoked by information

about, rather than presence of, another subject. Creating social comparison through observation instead, Gamba et al. (2017) let subjects perform a task in presence of, but independently of, another subject. After completing the task, subjects are informed about their own and the other's wage, which are randomly assigned. Subsequently, subjects are asked to choose between two lotteries. In turn, Gamba et al. (2017) find subjects make riskier choices if they received a lower relative wage prior to the risk choice, i.e. if they found themselves lower in the social hierarchy. Similar evidence of increased risk-taking to "keep up with the winners" is found in Hill and Buss (2010), Linde and Sonnemans (2012), and Fafchamps et al. (2015).

To summarise, the literature on competitive framing provides evidence on the impact of competition besides that of competitive incentives. Instead, either mental reactions to a competitive frame or the presence or the knowledge of others being evaluated alongside you impact effort and risk-taking choices. Together, competitive framing is predicted to have a positive impact on risk-taking, but predictions for effort are less clear, albeit mostly positive. However, to our knowledge no study has examined effects of competitive, relative to a neutral, framing when subjects simultaneously choose effort and risk-taking.

4.2 Hypotheses

While our models for the effect of competitive incentives made no difference between whether a threshold is framed in a neutral way (T = NT) or in a competitive way (T = CT), the related literature shows such a distinction may yield differential effort and risk levels. As such, introducing competitive framing implies that choices of risk and effort in *Neutral threshold* (NT) differ from those in *Competitive threshold* (CT). In particular, the literature suggests that by introducing a competitive frame, agents choose higher risk-levels but also increased effort, even when there is no strategic incentive of rivalry to try harder. As such, we formalise the following hypotheses:

H3 *A competitively framed threshold leads to a higher level of risk-taking than a neutrally framed threshold:* $\sigma^2_{i,CT} > \sigma^2_{i,NT}$

H4 *A competitively framed threshold leads to a higher level of effort than a neutrally framed threshold:* $e_{i,CT} > e_{i,NT}$

Based on this, we expect an agent who faces a threshold framed as the points of another competitor to increase both choice variables in the hopes of winning a bonus.

5 Hypotheses for addition and substitution

To close our hypotheses for the difference in effort and risk-taking across our treatments, we consider the additive impact of competition's two aspects. Moreover, as choices of risk-taking and effort are simultaneous, we also create hypotheses for the direction of their correlation.

5.1 Additive impact of incentives and framing

While the past two sections provide separate hypotheses for competitive incentives and competitive framing, competition may, and commonly does, consist of both aspects at once. For our experimental design particularly, the competitive frame is included in the *Direct competition* (DC) as well as the *Competitive threshold* (CT) treatments, but *Direct competition* also includes competitive incentives. In order to examine the impact of "real" competition, as found in our everyday life, we therefore hypothesise for the additive effect by relating the two aspect to one another.

For risk-taking, both Dijk et al. (2014) and Kirchler et al. (forthcoming) add to the reviewed literature by focusing on the interaction between competitive incentives and framing. In field and online experiments, Kirchler et al. (forthcoming) find competitive incentives an framing to increase risk-taking together, among bankers and students who are less well-performing, i.e. with lowest winning probability. On its own, however, a competitive frame does not increase risk-taking for students. Despite this, the result does suggest competitive incentives and framing interplay, and together create stronger impacts of competition, in particular in some groups. Similar results are found in the relative performance evaluation study by Dijk et al. (2014). As such, both find competitive framing is distinct from and additive to competitive incentives.

Building on these joint findings and the separate observations of framing and of incentives, we hypothesise that the additive effect is a convex combination, producing stronger impacts together than each aspect does on its own. As the literature suggests both competitive incentives (Andersson et al., 2017) and competitive framing (Eriksen and Kvaløy, 2017; Kirchler et al., forthcoming) increase risk-taking, we hypothesise they jointly lead to higher risk-taking than either alone:

H5 *Direct competition leads to higher risk-taking than a competitively framed threshold and than an neutrally framed threshold:* $\sigma^2_{i,DC} > \sigma^2_{i,CT} > \sigma^2_{i,NT}$

However, the additive reasoning becomes less straightforward for the impact on effort. While models of competitive incentives suggest direct competition leads to lower effort than threshold evaluation, competitive framing has been found to generate higher effort (e.g. Falk and Ichino, 2006; Mas and Moretti, 2009; Gächter et al., 2013). As the literature is unclear over which effect dominates, we argue the competitive framing aspect is not considered in the hypotheses derived directly from competitive incentives models. Rather, we hypothesise framing has an additive, positive impact on effort in *Direct competition* (DC) and *Competitive threshold* (CT), an impact which is larger than the positive impact of no competitive incentives in *Neutral threshold* (NT). As such, we formulate the following hypothesis for effort:

H6 *Direct competition leads to lower effort than competitively framed threshold but higher effort than a neutral threshold*: $e_{i,NT} < e_{i,DC} < e_{i,CT}$

5.2 Substitution between effort and risk-taking

Moreover, throughout our experimental design and all treatments, effort and risk-taking are not mutually exclusive, but subjects can affect the outcome measure, and thus their probability of winning, through both choices simultaneously. Concretely, effort and risk-taking may act as either strategic substitutes or complements. In line with our models and the evidence by Hvide (2002), Nieken (2010), and Andersson et al. (2017), we hypothesise that:

H7 *Effort and risk-taking are substitutes to one another, i.e. subjects who exert high effort exert lower risks and vice versa.*

Our predictions for separable and additive impacts of competitive incentives and competitive framing are thus complete. While our experimental treatments embody one or more of these aspects, comparing them allows us to isolate each. Comparing choices in *Direct competition* and *Competitive threshold* isolates the impact of competitive incentives; Comparing choices in *Neutral threshold* and *Competitive threshold* isolates the impact of competitive framing; Relating all three treatments to one another isolates the additive impact.

6 Empirical strategy

This section provides a description of our variables and dataset, followed by outlining our regression strategy. Building on our regression components, we derive statistical hypotheses and related tests for competitive incentives, from competitive framing, as well as from their joint effect. The hypotheses, data collection procedure and empirical strategy was preregistered at the Open Science Framework prior to data collection, and is found in Appendix D. Any deviations from our pre-analysis plan are outlined and motivated.

6.1 Data

The following section defines dependent, independent, and control variables for the main analyses to provide an understanding of our data. Variables are outlined in Table 6.1.[12] Through our experimental procedure, we collect our two dependent variables, effort and risk-taking, directly. We define effort exerted as the number of correctly placed sliders. We define risk-taking as the number of points from each correct slider the subject invests in the lottery.

Moreover, our key independent variable is a subject's treatment group. Setting the *Neutral threshold* (NT) treatment as baseline in regressions, we create dummy variables for the *Direct competition* (DC) and the *Competitive threshold* (CT) treatments.

Table 6.1: Variable definitions

Variable	Range	Description	Reference
Dependent			
Effort, e	0 to 60	Number of correctly placed sliders	Gill and Prowse (2012)
Risk-taking, r	0 to 9	Number of points per correct slider bet in lottery	Gneezy and Potters (1997)
Independent			
DC	0, 1	1 if subject was in the *Direct competition* treatment	
CT	0, 1	1 if subject was in the *Competitive threshold* treatment	
Control			
Age	0 to ∞	Age of subject in years	
Female	0, 1	1 if *Gender* = female	
USA	0, 1	1 if *Country of residence* = USA	Ipeirotis (2010), Difallah et al. (2018)
India	0, 1	1 if *Country of residence* = India	Ipeirotis (2010), Difallah et al. (2018)
Overplacement	-120 to 120	$(\mathrm{E}(e_i) - \mathrm{E}(\bar{e})) - (e_i - \bar{e})$ where e_i is the subject's effort and $\bar{e}_i$ is the effort of the average participant	Moore and Healy (2008)
Overestimation of winning	0, 1	1 if subject believed she would win and did not, 0 otherwise	Moore and Healy (2008)
General risk preferences	0 to 10	Scaled answer, 0 being "Not willing to take risks" and 10 "Very willing to take risks"	Dohmen et al. (2011)

[12] Further variable definitions can be found in connection with robustness tests and further analyses in Appendices C.2 to C.4.

We also collect information on individual attributes to understand potential heterogeneity in choices of effort and risk-taking across a population. In particular, we suspect individual attributes may evoke differential reactions to competitive incentives and framing, with two main implications. On the one hand, to capture only the variation caused by treatment, these personal characteristics should be controlled for when estimating the impacts on effort and risk-taking.[13] On the other hand, collecting this information also allow us to explore potential heterogeneity in responses to competition for gender, risk-aversion, and overconfidence in our secondary analysis in Section 9.

Focusing on five key attributes, subjects were firstly asked to provide their age, gender, and country of residence. From the latter two we build dummy variables for "Female" gender and for the two main countries of residence for MTurk workers: USA and India. Secondly, we elicited subjects' level of under- or overconfidence by asking how many sliders they think they will place correctly, how many sliders they think people on average will place correctly, and whether they believe they will beat their target. Given this, we capture overconfidence in two main measures. We define *Overplacement* as the belief that you will correctly place more sliders than the average subject, but fail to do so. Moreover, we define *Overestimation of winning* as the belief that you will win, but do not. While overplacement embodies only beliefs about effort, overestimation regards beliefs about winning – including effort, risk-taking, and uncertainty over the lottery outcome. Lastly, subjects' general risk preferences were elicited on a range from zero ("Not willing to take risks") to ten ("Very willing to take risks").

6.2 Regression analysis

To isolate the effect of competitive incentives from that of competitive framing, as well as their additive effect, we analyse differences in effort and risk-taking between our three treatment groups. In line with related experimental literature, our main statistical method to answer our research question is ordinary least squares regressions (henceforth: OLS). Two things primarily motivate our choice. Firstly, as subjects are randomised into treatment groups, differences in unobserved characteristics are zero in expectation. The assumption of exogenous independent variables therefore holds, and our estimates are thus arguably unbiased. Secondly, in comparison with simple non-parametric or parametric tests of means or medians, OLS regressions allows us to control for potentially influential observable factors. Additionally, to account for potential heteroskedasticity, we use robust standard errors throughout our analysis.

6.2.1 Regression specification and hypotheses tests

We analyse the impact of competition on effort and risk-taking in separate OLS regressions. Throughout this section, regressions are specified with risk-taking (r_i) as dependent variable, but equivalent specifications are used for regressions with effort (e_i) as dependent variable. Generally, our specification is given by

$$r_i = \alpha + \beta_2 \times \mathrm{DC}_i + \beta_3 \times \mathrm{CT}_i + \gamma_1 K_{i1} + ... + \gamma_k K_{ik} + \epsilon_i, \tag{2}$$

where DC_i and CT_i are treatment dummies for our *Direct competition* and *Competitive threshold* treatments, respectively. α is a constant and K_{i1}, ..., K_{ik} a list of k potential control variables.

We perform three versions of the specification, including different combinations of control variables in order to account for potential non-treatment drivers for effort and risk-taking, as well as variables which may relate to competition, beyond incentives and framing. First, we set $k = 0$, for a pure specification of only treatment dummies and a

[13] As in done in e.g. Buser and Dreber (2016) and Andersson et al. (2017).

constant. Thereafter, we set $k = 3$ by including key personality controls for overconfidence and risk preferences; *Overestimation of winning*, *Overplacement* and *General risk preferences*. As such, we control for situationally relevant beliefs and preferences which relate directly to exerting effort and risk-taking when performing against a target. Finally, for our so called "full model" specification, we set $k = 7$ by including, in addition to the three previous controls, also personal characteristic controls for gender (*Female*), age (*Age*), and country of residence (*USA*, *India*).

Our experimental design allows us to test our theoretical hypotheses through comparisons between treatments. In turn, we expect the following: competitive incentives lead to higher risk-taking in *Direct competition* relative to *Competitive threshold* (**H1**); competitive framing generate higher risk-taking in *Competitive threshold* relative to *Neutral threshold* (**H3**); the additive impact of incentives and framing cause higher risk-taking in both *Direct competition* and *Competitive threshold* relative to *Neutral threshold* (**H5**).[14] The resulting statistical hypotheses and predictions for the direction of coefficients for risk-taking are outlined in Table 6.2.

Table 6.2: Statistical hypotheses for risk-taking

	Competitive incentives (**H1**)	Competitive framing **(H3)**	Additive effect (**H5**)
H_0	$\beta_2 = \beta_3$	$\beta_3 = 0$	$\beta_2 = \beta_3 = 0$
H_1	$\beta_2 \neq \beta_3$	$\beta_3 \neq 0$	$\beta_2 \neq 0$ or $\beta_3 \neq 0$
Prediction	$\beta_2 > \beta_3$	$\beta_3 > 0$	$\beta_2 > \beta_3 > 0$

For the regressions with effort (e_i) as dependent variable, Table 6.3 outlines statistical hypotheses and predictions. While methods for isolating the impact of each aspect of competition remains the same as for risk-taking, predictions for directions of the coefficients are altered. For effort we expect; competitive incentives gives lower effort in *Direct competition* than in *Competitive threshold* (**H2**); competitive framing generates higher effort in *Competitive threshold* than in *Neutral threshold* (**H4**); the additive impact of incentives and framing causes higher effort in both *Direct competition* and *Competitive threshold* relative to *Neutral threshold* (**H6**).

Table 6.3: Statistical hypotheses for effort

	Competitive incentives (**H2**)	Competitive framing **(H4)**	Additive effect (**H6**)
H_0	$\beta_2 = \beta_3$	$\beta_3 = 0$	$\beta_2 = \beta_3 = 0$
H_1	$\beta_2 \neq \beta_3$	$\beta_3 \neq 0$	$\beta_2 \neq 0$ or $\beta_3 \neq 0$
Prediction	$\beta_2 < \beta_3$	$\beta_3 > 0$	$\beta_3 > \beta_2 > 0$

For tests of a single coefficient a standard t-test is used; for tests of more than one coefficient, an F-test is used. Throughout, we use two-sided tests and reject any null hypotheses at a 95% significance level.

Finally, as effort and risk-taking are joint strategic components of output in our experimental design, we analyse the potential for risk-taking and effort to function as complements or substitutes (**H7**) on average. To do so, we explore their relation using

[14] The final prediction for hypothesis **H5**, $\beta_2 > \beta_3 > 0$, follows directly from **H1** and **H3**.

OLS regressions where we include one in the regression for the other.[15] In particular, we regress risk-taking (r_i) on effort (e_i) and specify

$$e_i = \alpha + \beta_1 r_i + \beta_2 \times \mathrm{DC}_i + \beta_3 \times \mathrm{CT}_i + \gamma_1 K_{i1} + ... + \gamma_7 K_{i7} + \epsilon_i,$$

where all variables are defined as in Equation 2, apart from including risk-taking, r_i. Here, we only employ the full model list of control variables, i.e. $k = 7$. We separately perform the regression for the full sample as well as for the separate treatment groups to discern potential treatment differences in the correlation. As such, we test the following hypothesis for the direction of the relation between effort (e_i) and risk-taking (r_i) in our regressions:

$$\textbf{H7} \quad \begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

As we seek to test whether the dependent variables are substitutes; if subjects who decrease effort instead increase risk-taking, we would expect a negative relation, i.e. $\beta_1 < 0$. If they on the other hand are complements, we would see a positive relation, i.e. $\beta_1 > 0$. The magnitude of β_1 provides an understanding of not just the correlation between the two, but its relation relative to other factors which may drive effort, e.g. treatment or control variables.

6.2.2 Robustness

The robustness of our results is tested in several ways. First, we perform three versions of our regression specification. By doing so, we test the robustness of our estimates to inclusion of different observable characteristics which may explain variation in effort or risk-taking. Second, we run regressions with normal standard errors as well, as the choice of standard errors was not preregistered. Third, we test the robustness of the answer to our research question, by altering the choice of estimation strategy away from linear effects on effort and risk-taking. Instead, we employ a probit model to estimate the probability of a subject exerting high effort or taking high risk when in competition.[16] The way we phrase our variables and our regression specifications is outlined further in Appendix C.2.

Fourth, we test the robustness of our results to the specific choices in variable formulations by varying the measures of risk preferences and overconfidence we use as control variables. For risk preferences we substitute *General risk preferences* with high- versus low risk preferences (*Risk-aversion*) to explore if coefficients of interest are influenced by individuals being absolutely risk averse, rather than their specific level of risk preference. We also test robustness using domain-specific measures of risk-aversion (*Risk-aversion in entrepreneurship* and *Risk-aversion in gambling*). Individuals' risk preferences are shown to be best predicted by questions which relate to the relevant domain of choices.[17] Specifically, risk-taking in MTurk surveys generally could be linked to risk-taking in other self-employment situations and risk-taking in the lottery task specifically could be linked to propensity for gambling, respectively. For overconfidence we include combinations with *Overestimation of performance*, i.e. an overconfidence measure that relates only to own performance rather than others as compared to *Overplacement* and *Overestimation of winning*. Variable descriptions are given in Appendix C.3.

[15] As is done in e.g. Andersson et al. (2017).
[16] Similarly to what is done in Nieken and Sliwka (2010) and Kirchler et al. (forthcoming).
[17] See the comparison of risk preference elicitation methods in Dohmen et al. (2011).

7 Results and analysis

The experiment was carried out between the 7th and 9th of April, 2018 with a final sample of 417 subjects.[18] We first present a descriptive analysis[19] of the outcome of our experiment and the subject pool, followed by a causal analysis of the main results and related hypotheses tests.

Our main results can be summarised as follows; we cannot conclude that competition between individuals affects choices over effort and risk-taking in tasks where the outcome is determined by both. Neither competitive incentives nor competitive framing is found to affect choices on average – either separately or jointly. Furthermore, effort and risk-taking are not found to be either strategic substitutes or complements. Our sample is balanced between treatments and our results robust to various alterations.

7.1 Descriptive analysis

For a first insight into our dataset, variables of interest and their allocations across treatment groups are found in Table 7.1. A few observations can be made. Firstly, our treatment groups are roughly equal in size. Secondly, both variables of interest are similar across treatments. In particular, mean effort in *Neutral threshold* and *Direct competition*, as well as mean risk-taking in *Direct competition* and in *Competitive threshold*, are near identical.

On average, subjects completed 29 sliders; a slightly higher effort level than the 22-27 sliders in previous studies with the slider task (Gill and Prowse, 2012; Buser and Dreber, 2016). The standard deviation is however also large compared to related studies, yielding on average, a 95% confidence interval of approximately 28 to 30 sliders. Additionally, standard errors are relatively equal across treatments – a first indication of competition's lack of impact on effort. While the first column of Figure 7.1 indicates the distributions of effort are not identical across treatment groups, Appendix C.1.1 explores the cumulative frequencies in more detail.[20] Performing the Kolmogorov-Smirnov test of equality of distributions between each pair of treatments, we find the treatment distributions for effort are too similar to identify any relevant differences.

Table 7.1: Summary statistics: Dependent variables

	Neutral threshold	Direct competition	Competitive threshold	Total
Effort	29.364	29.239	28.439	29.014
	(8.556)	(9.072)	(9.772)	(9.133)
Risk-taking	5.350	5.797	5.763	5.635
	(3.158)	(2.844)	(3.004)	(3.005)
Observations	140	138	139	417

Mean estimates, standard deviations in parenthesis

For risk-taking, the distribution across treatments is correspondingly similar, as seen in the second column of Figure 7.1 and in Figure C.1.2. Mode risk-taking was to bet all points from each correct slider, but on average subjects bet approximately 5.6 points. As such, the majority of our subjects limit their risk-taking, in line with evidence on the

[18] Three subjects who had started but not finished the experiment, and then restarted were excluded from the original sample of 420 observations.

[19] In our descriptive analysis we extend upon our pre-specified use of descriptive analysis by including balance tests of randomisation using ANOVA as well as Kolmogorov-Smirnov tests of equality of distributions for the distributions of effort and risk-taking across treatment groups.

[20] In particular, we plot the cumulative frequency distributions in Figure C.1 and test the difference in distribution in Table C.1.2.

prevalence of risk-aversion (Holt and Laury, 2002). Moreover, in our two competitive treatments the average bet was relatively higher than in *Neutral threshold*, but the difference is not statistically significant and also here the Kolmogorov-Smirnov test of equality of distribution cannot identify any significant difference in distributions between treatments.[21]

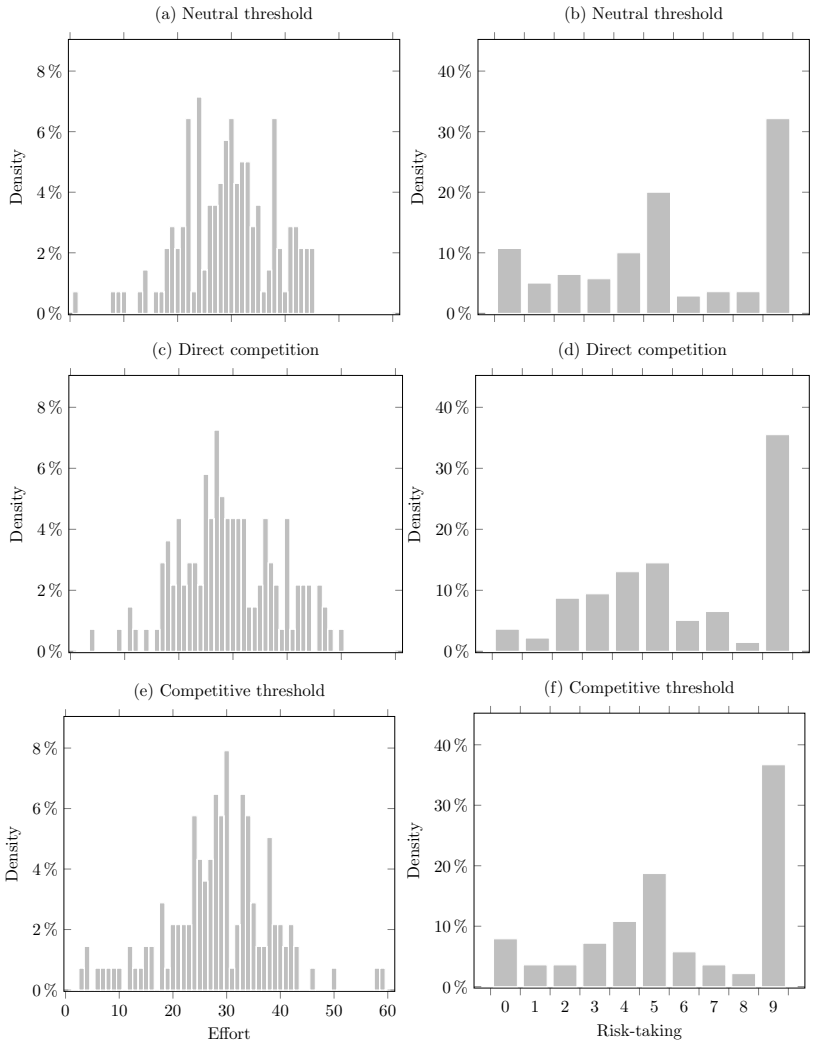


Figure 7.1: Distribution of risk-taking and effort across treatments

Interestingly, many of our subjects take excessive risks. If an individual in either threshold treatment predicts they can complete 30 sliders or more, our threshold model proposes they should take zero risk, as any positive risk only decreases the likelihood to meet the threshold.[22] Examining only those who actually went on to complete 30 sliders or more after predicting to do so, we find 22.1% of subjects in *Neutral threshold* and 25.2% in *Competitive threshold* take positive risks. Approximately half of these end up losing – simply as a consequence of their excessive risk-taking.

Beyond our dependent variables, information to construct our control variables was collected. As Table 7.2 displays, the average age of subjects was 37.7 years, with youngest and oldest at 18 and 77 years old respectively. We find 48.7% of subjects are female,

[21] Details for the cumulative frequency distributions for risk-taking are found in Appendix C.1.1.

[22] The threshold subjects in both treatments perform against is 273 points. Each slider yields nine points, hence $\frac{273}{9} \approx 30.\overline{33}$ sliders.

71.9% reside in the United States of America and 21.3% in India – leaving only 6.8% of subjects living in 17 other countries across all continents. These demographics are in line with statistics on the MTurk population (Ipeirotis, 2010; Difallah et al., 2018).

Table 7.2: Summary statistics: Control variables

	Mean	Standard deviation	Min	Max
Age	37.743	11.899	18	77
Female	0.487	0.500	0	1
USA	0.719	0.450	0	1
India	0.213	0.410	0	1
Overplacement	1.156	12.734	-39	50
Overestimation of winning	0.448	0.498	0	1
General risk preferences	5.801	2.828	0	10
Observations	417			

Our two remaining personality variables reflect that subjects' overconfidence was limited and subjects were neither extremely risk-averse nor very risk-loving. For overconfidence we find subjects on average exhibit little overplacement, predicting themselves to complete 1.2 sliders too much above the average. However, sizeable standard deviations indicate large variation in the degree of overplacement. Similarly, our second overconfidence measure, *Overestimation of winning*, suggests 44.8% of subjects are overconfident with regards to both effort and risk-taking. As such, one cannot conclude whether subjects were either over- or underconfident on average. For *General risk preferences*, Table 7.2 indicates that, on average, subjects rate themselves as 5.8, on a scale from "Not willing to take risks" to "Very willing". With a standard deviation of 2.8, the majority of subjects do not rate themselves at either extreme end of risk preferences.

While this discussion has considered individuals on average, our randomisation process implies a subject is equally likely to end up in either treatment. Thus, we also know the null hypothesis of balance in observable characteristics should hold for any variable that is unaffected by treatment (Mutz et al., 2018). In accordance with our expectations, Table C.1.1 shows no indication of unbalanced pre-treatment variables.[23] However, a few variables deserve further consideration as they are elicited post-treatment and regard treatment-related factors – the overconfidence and risk preference measures. Yet also here we cannot identify any significant difference in means between treatments. This, along with the descriptive statistics of the dependent variables, provides a strengthened indication that competition does not affect either risk-taking or effort, as well as leaving individual's overconfidence and general risk preferences unaffected by treatment.[24]

7.2 Regression results

This section presents empirical evidence from our experiment. Table 7.3 provides coefficient estimates from our OLS regressions for risk-taking as well as effort.

A few initial observations can be made. First, coefficient directions follow predictions in some, but not all regressions. While *Competitive threshold* (CT) coefficients are positive as predicted in all but one case, *Direct competition* (DC) coefficients are positive, but too small for risk-taking and negative for effort, in contrast to predictions. Second, in regressions for risk-taking, treatment coefficients remain insignificantly different from

[23] Finding one significant difference among eight variables and three groups is not unsurprising. Age is somewhat lower in the *Direct competition* treatment, but only at a three-year level, an arguably economically insignificant difference.

[24] However, as seen in Section 9, individuals' overconfidence and risk preferences are not unrelated with the impacts of treatment, but are seemingly not created by the treatment.

zero throughout all specifications. *Direct competition* and *Competitive threshold* are positive, yet consistently insignificant. Third, in regressions for effort, treatment coefficients remain similarly insignificant. While weakly negative in base specifications, the *Direct competition* coefficient turns positive in fuller models, but remains insignificant. As such, initial observations give no clear indication that competitive treatments are any different to the baseline. The persistence of insignificance also suggests that any result is not contingent upon inclusion of specific controls.

Table 7.3: OLS regressions of risk-taking and effort

	(1)	(2)	(3)	(4)	(5)	(6)
	Risk-taking	Risk-taking	Risk-taking	Effort	Effort	Effort
DC	0.413	0.262	0.235	-0.925	-0.264	-0.102
	(0.369)	(0.344)	(0.346)	(1.100)	(0.756)	(0.742)
CT	0.447	0.356	0.368	-0.125	0.569	0.620
	(0.360)	(0.337)	(0.339)	(1.058)	(0.774)	(0.742)
Overestimation		0.439	0.422		0.435	0.674
of winning		(0.292)	(0.291)		(0.615)	(0.599)
Overplacement		0.009	0.008		-0.522***	-0.520***
		(0.012)	(0.012)		(0.034)	(0.034)
General		0.348***	0.333***		0.081	0.089
risk preferences		(0.055)	(0.061)		(0.115)	(0.125)
Age			0.014			-0.111***
			(0.012)			(0.027)
Female			-0.325			-1.891**
			(0.290)			(0.623)
USA			0.247			-0.507
			(0.476)			(1.270)
India			0.465			-4.014**
			(0.524)			(1.423)
Constant	5.350***	3.205***	2.661***	29.364***	28.852***	34.944***
	(0.267)	(0.398)	(0.779)	(0.723)	(0.871)	(1.744)
Observations	417	417	417	417	417	417
R^2	0.005	0.136	0.142	0.002	0.518	0.557

Robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Moreover, the persistence of insignificant treatment coefficients with inclusion of control variables suggests lack of treatment differences is not contingent upon controlling for specific characteristics. Examining the control coefficients themselves also allows us to understand where variation in effort and risk-taking may stem from beyond treatment. In particular, general risk preferences exhibit a positive, highly significant impact on risk-taking, but not on effort. A reasonable result, suggesting individuals who see themselves as more risk-liking also invest more in the lottery, but also that liking risk does not make you exert more effort on average. For effort we instead find a negative, significant impact of overplacement, indicating that on average, individuals who falsely believe they will exert more effort than others in turn exert less effort. Overestimation of winning is however not significant across specifications, indicating overconfidence measures are not directly equivalent. Finally, individual characteristics seem to drive more of the variance in exerted effort than in risk taken. For effort, coefficients for age, female gender and residing in India are negative and statistically significant. In particular, subjects who are female or reside in India complete approximately 1.9 and 4.0 sliders less than others, respectively – an economically significant result.

Together, the regression results suggest there are factors outside of our models which explain the level of effort and risk-taking. Supporting this, the explanatory power of only treatment variables is very low; an R^2 of 5% and 2% for risk-taking and effort, respectively. However, R^2 increases in particular in effort regressions, suggesting our controls cover much of the variance in effort. Interestingly, while controlling for risk preferences and overconfidence contributes significantly, adding individual characteristics contributes to a lesser extent. On the whole, we draw two temporary conclusions: first, there is interesting heterogeneity in the impact of personal characteristics on choices of effort and risk-level in general, but perhaps also between competition and no competition, considerations we return to in Section 9. Second, lack of impact of treatments is persistent and coefficients not dependent upon controls. Yet, to fully disentangle the impact of competitive incentives and framing, the remainder of this section perform our outlined hypotheses tests. Subsequently, we discuss robustness of the results and analyse its implications.

7.2.1 Competitive incentives

For the impact of competitive incentives on risk-taking (**H1**) and effort (**H2**) we compare coefficients DC and CT across the three regression specifications. We expect relatively higher risk-taking and lower effort under competitive incentives, *Direct competition* (DC), than without competitive incentives, *Competitive threshold* (CT). However, Table 7.3 suggests this can only hold for effort and not for risk-taking, as coefficient DC remains smaller throughout both regression sets. Testing the null hypotheses of no difference for each regression, Table 7.4 lists F-statistics. Across all specifications, the large p-values indicate we cannot reject the null hypotheses, and as such, cannot conclude competitive incentives affect either risk-taking or effort.

Table 7.4: F-tests

	Risk-taking		Effort	
	H1	**H5**	**H2**	**H6**
Column 1, 4	0.009	0.914	0.499	0.396
	(0.922)	(0.402)	(0.480)	(0.673)
Column 2, 5	0.078	0.589	1.165	0.603
	(0.780)	(0.555)	(0.281)	(0.548)
Column 3, 6	0.152	0.598	0.912	0.538
	(0.697)	(0.551)	(0.340)	(0.584)

Column numbers refer to regressions in Table 7.3.
p-values in parentheses.

7.2.2 Competitive framing

As outlined, subjects in the baseline *Neutral threshold* (NT) and in *Competitive threshold* (CT), face the same threshold incentives but different frames. Hence, to examine the impact of competitive framing on risk-taking (**H3**) and effort (**H4**) we test the significance of the coefficient CT. As expected, the coefficient of interest is positive in all specifications, apart from (4) where it suggests a competitive frame yields a lower effort than a neutral frame. However, as indicated in Table 7.3, the relative size of the standard errors to the coefficient CT render it insignificantly different from zero across all models. As such, we cannot reject either **H3** or **H4** and thus cannot conclude competitive framing leads to either higher risk-taking or greater effort, on average.

7.2.3 Additive impact of incentives and framing

While neither competitive incentives nor competitive framing are found to significantly impact risk-taking or effort on their own, it remains to investigate whether the two aspects may have a joint impact – i.e. testing the impact of "real" competition. Examining the additive hypotheses for risk-taking (**H5**) and effort (**H6**) in turn, we test if treatment coefficient DC and CT are equal and separately significantly different from zero. In line with our previous results, Table 7.4 indicates neither null hypotheses can be rejected, across specifications and for both variables. In turn, we cannot conclude competitive incentives and framing have any additive impact on either risk- or effort choices, on average.

7.2.4 Substitution between effort and risk-taking

Finally, we explore the relation between risk-taking and effort in order to identify whether they are substitutes or complements, i.e. hypothesis **H7**. Table 7.5 outlines outcomes of standard, full-model OLS regressions for effort but including risk-taking as a control variable. The regressions are performed separately for each treatment sample (columns 1-3) as well as for the entire sample (column 4). Alike analyses for effort and risk-taking, both treatment coefficients remain insignificant in the complete sample specifications. The key coefficient of interest here, *Risk-taking*, also remains close to zero and insignificant across the four regressions. Hence, we cannot reject the null hypothesis that the choice of risk level is related to the choice of effort. Thus, we find no evidence that effort and risk-taking are substitutes or complements, on average.[25]

Table 7.5: OLS regressions of effort: Substitution

	(1)	(2)	(3)	(4)
	Effort	Effort	Effort	Effort
Risk-taking	0.001	0.117	0.011	0.080
	(0.182)	(0.195)	(0.210)	(0.110)
DC				-0.121
				(0.743)
CT				0.591
				(0.748)
Constant	36.846***	40.422***	29.146***	34.732***
	(2.487)	(3.359)	(2.971)	(1.758)
Sample	NT	DC	CT	Total
Full model controls	Yes	Yes	Yes	Yes
Observations	140	138	139	417
R^2	0.504	0.572	0.638	0.558

Regressions for each separate treatment groups: *Neutral threshold* in column (1), *Direct competition* in (2) and *Competitive threshold* in (3). Total sample in column (4). Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

All in all, we are unable to reject hypotheses **H1-H7**, finding neither *Direct competition* nor *Competitive threshold* to yield any differential impacts relative to each other or *Neutral threshold*. We also do not find any significant – negative or positive – correlation between effort and risk-taking.

[25] It may still be that specific subjects' choices are correlated, yet such analysis would require multiround tests. However, between subjects and groups there are no significant patterns.

7.2.5 Robustness

Our results are consistently robust to various adjustments. First, including additional control variables in the baseline specifications does not change outcomes of our hypotheses tests (Table 7.3). Second, the results are robust to using normal standard errors (Table C.3.1). Third, the results from OLS regressions remain qualitatively similar in probit regressions, as detailed in Appendix C.2. While directions of treatment coefficients change in some of the probit specifications, both DC and CT coefficients remain consistently insignificant. Likewise, we cannot reject any null hypotheses for either risk-taking or effort. As such, our results are arguably independent of estimating impacts on risk-taking and effort generally, or on propensities to take higher risks and exert more effort particularly.

Fourth, our results are robust to the specific measures chosen for key control variable (Table C.3.3 and C.3.4). Throughout, treatment coefficients remain insignificant and standard errors too large to be able to detect any difference between treatments. For explanatory power, replacing *Overplacement* with another overconfidence measure clearly reduces explanatory power for effort regressions, suggesting incorrect beliefs over effort explain a greater deal of the variation in actual effort. Similarly, for risk-taking, substituting general risk preferences for context-specific risk-liking lowers explanatory power. Alterations in overconfidence measures however have little impact, suggesting risk-taking reasonably depends more on risk preferences than overconfidence. All together, our results do not seem to depend upon the general inclusion or specific choice of control variables, its standard errors, or its estimation strategy. Further details are robustness checks are given in Appendices C.2 and C.3.

7.3 Analysis

Based on the outlined results we cannot conclude if competition has an impact on either performance choices. Rather, individuals choose a combination of both effort and risk – a combination which is unaffected by whether the task is performed against competitors or not, or versus a fixed or uncertain target. Additionally, the levels of effort and risk-taking chosen do not depend on one another, either in competition or not. Overall, the results are robust to varied specifications and estimation strategies.

Nonetheless, the results may be driven by selection bias. On the one hand, randomisation into treatments precludes self-selection into competition, yielding relevant unobservable factors, in expectation, equally represented across treatments. On the other hand, early exit from treatment may bias estimates, in particular as attrition post-treatment is pervasive in online experiments with unobserved subjects. Attrition is particularly problematic if decisions to drop out are correlated with preferences for dependent variables. For example, if subjects in either competitive treatment drop out at higher rates it could indicate aversion against competition, and in turn risk-aversion when in competition. Remaining subjects may prefer higher risk, thus biasing risk-taking estimates upward.

However, employing Fisher's exact test we find no significant differences between treatments in propensity to attrite post-treatment. Performing the test, we exclude individual who attrited pre-treatment and those who failed a majority of attention questions, as these would not qualify for the main sample had they not attrited. In total, 24 subjects attrited, but only 10 post-treatment and qualified. In turn, the contingency table in Table C.1.3 indicates attrition is evenly distributed across treatment groups and the test has a p-value of 0.610.[26] Selective attrition is thus not a problem for our results.

[26] The exclusion of drop-outs failing attention checks was not specified in our pre-analysis plan. The test was employed all 18 drop-outs and yields similar results, as can be seen in Table C.1.4.

All together, our null result for the impact of competition on either variable contrasts both our predictions and much of the previous theoretical and experimental literature. To verify our results we must, however, consider alternative explanations. Firstly, to provide an insight into the divergence to the literature, in Section 8 we review related experimental studies, exploring differences in set-up and results.

Secondly, while our null result may be true on average, there may be important heterogeneity in responses to competition across individuals. A number of studies suggest personal characteristics and preferences affect competitive behaviour in opposing directions, which could explain our null on average. Characteristics such as gender, overconfidence and risk preferences are approximately equally divided between our treatments or high versus low levels, but also have large variances (Table C.1.1). Focusing on three potential "channels" for impacts of competition, we expand upon previous hypotheses, empirical strategy and results by analysing heterogeneity in responses in Section 9.

8 Related literature

As our experimental design expands upon present research in several ways there are no directly comparable studies. Hence, in this section we review literature on competition's impact on effort, on risk-taking, as well as on both effort and risk-taking, and discuss their relation to our results. As the literature is vast, we focus on the subset closest to our research question and design, and which has not yet been discussed in the review of theoretical models for competitive incentives in Section 3 and of evidence on competitive framing in Section 4.

In summary, while studies which employ our real-effort slider task or our lottery risk-choice separately find evidence of impacts on effort and risk-taking, our results indicate that when the tasks are combined and compared to a non-competitive situation, impacts are no longer significant. Additionally, while our results cannot confirm our predictions, the closest related literature is also unable to reach a consensus on the impact of competition. Reviewing the literature, we first explore studies of both effort and risk-taking, followed by studies of the separate components.

First, studies of competition's impact on both effort and risk-taking are most closely related to us, but have, to our knowledge, primarily focused on competitive incentives. In particular, Andersson et al. (2017) were, to our knowledge, first to experimentally study competition's impact on simultaneous risk and effort choices. By manipulating the degree of competitive incentives, Andersson et al. find that greater prize differentials incentivises subjects to choose higher risk levels. For effort however, negative, yet only insignificant effects of competition are found across degrees of competitive incentives. Overall, Andersson et al. identify large heterogeneity in responses to competition. Moreover, exploring potential risk-effort substitution, Andersson et al. find a significant negative correlation between effort and risk-taking between-subjects, but only in one treatment. More importantly, within-subjects they find a strong positive correlation, suggesting effort-risk substitution is individual rather than treatment-specific.

Apart from differences in their outcomes, Andersson et al. (2017) differ from our contribution in several ways. First and foremost, they do not include a neutral treatment, and subsequently cannot isolate impacts of competitive framing. Second, to elicit effort, Andersson et al. do not use a real effort task, but rather a choice of investment into the mean of the output variable similar to our risk choice. As shown by Lezzi et al. (2015), real versus non-real effort are not equivalent predictors of behaviour. Third, both Andersson et al.'s theoretical predictions, from the previously discussed Gilpatric (2009), and design build on a three-player setting, as compared to our two-subject design. Hence, potential competitive framing effects depend on two social reference points, rather than just one as in our study.

In contrast to the simultaneous effort and risk-taking in our study and Andersson et al. (2017), Nieken (2010) and Filippin and Gioia (2017) focus on sequential tournaments. On the one hand, Nieken tests predictions of the competitive incentives model by Hvide (2002) in a two-player, laboratory experiment. While Nieken does not randomise subjects into treatments, she introduces correlated risk strategies where subjects observe each other's risk-choices before choosing effort. In turn, Nieken finds subjects who choose high risk levels subsequently exert low effort; i.e. a between-subjects substitution between effort and risk-taking under competition. This is in line with Andersson et al., but contrasts our results.

On the other hand, Filippin and Gioia (2017) explore competition's impact on effort and spill-overs effects on risk-taking. Subjects first perform a real-effort task and subsequently their risk preferences are elicited in an independent, simple risk-choice. To isolate impacts of competitive incentives, payoffs from the effort task are randomly assigned in the baseline treatment and through a tournament in the competitive treatment. In contrast to Andersson et al. (2017), Filippin and Gioia find evidence of competition leading to increased effort within the task. However, they find no evidence of spill-overs on risk-taking, a surprising result as spill-over from competition are identified in a number of other economic decisions.[27] As such, despite a common focus on competitive incentives, evidence on joint impacts on effort and risk-taking range from only significant, positive risk-taking (Andersson et al., 2017) to only significant, positive effort (Filippin and Gioia, 2017).

Second, evidence of competition's impact on effort is more limited and conflicting, but offers insights into both competitive incentives and framing. As discussed, Gill and Prowse (2012) and Gächter et al. (2017) find opposing evidence on competitive incentives and effort of second movers in sequential, slider-task tournaments. While Gill and Prowse find a discouragement effect, Gächter et al. do not. Additionally, Gächter et al. isolate effects of competitive framing by adding a tournament against nature. Like us, they find no significant impact of framing on effort. However, our two results are not exactly comparable, particularly as outcomes are a function of risk-taking in addition to effort in our *Neutral* and *Competitive threshold treatments*. Nonetheless, the results jointly suggest a threshold's level seems to not matter for subjects' choices, either when competitively framed and not. Finally, our study goes beyond Gächter et al. as we also isolate impacts of competitive incentives.

Third, effects of competition on risk-taking are studied in particular in laboratory and field applications in behavioural finance. In summary, evidence on competitive incentives suggests individuals in general take on inefficient levels of risk, and in particular, bonus incentives lead individuals to gamble more, and further more when the money belongs to someone else (Agranov et al., 2013; Andersson et al., 2013; Kleinlercher et al., 2014). Similarly, we find evidence of excessive risk-taking, but in contrast we find it is not dependent on competitive incentives.

Studying not only incentives but also framing, Eriksen and Kvaløy (2014) and (2017) find the aspects to jointly increase risk-taking. Eriksen and Kvaløy (2014) also employ the risk preference elicitation task from Gneezy and Potters (1997), finding framing and incentives to increase risk-taking among student subjects. In particular, testing myopic loss aversion,[28] they show that while more frequent evaluation and feedback on investments into the lottery predicts decreased risk-taking, introducing competitive incentives yields increased risk choices. Sustaining the result, Eriksen and Kvaløy (2017), find evidence of excessive risk-taking even when it is optimal to take no risk. As only competitive treatments are included in Eriksen and Kvaløy (2014) and (2017), they isolate if excessive risk-taking is particular to competition. Similarly, we find excessive risk-taking to be common in threshold treatments, but as risk levels are equal across

[27] For example, unethical behaviour (Charness et al., 2014), cooperation (Buser and Dreber, 2016), and cheating (Garicano and Palacios-Huerta, 2014; Gill et al., 2013).

[28] Under myopic loss aversion, individuals take less risk the more often investments are evaluated.

treatments we do not find impacts to be particular to competition. The combined evidence thus suggests risk-taking is pervasive, but also that our risk task is attuned to detect risk-taking in experiments. Instead, variation in our results is perhaps rather explained by differences in set up, e.g. in feedback or number of rounds.

Presenting contrasting results, Kirchler et al. (forthcoming) find increased risk-taking to depend upon a task's domain. As discussed in Section 4.1, Kirchler et al. disentangle motivations for risk-taking behaviours in experiments with students and financial professionals. Competitive framing and incentives are found to be complements, where a social reference point and a bonus-based investment game interact to increase risk-taking. However, impacts of framing alone are only found among professionals. For students, tournament incentives and social rankings increase risk-taking together, but social rankings alone do not. As the risk choice is finance-related, the discrepancy indicates competitive framing's impact on risk-taking is perhaps domain-specific. For our results, we would have seen domain-specificity as increased risk-taking in *Competitive threshold* if our lottery task is within our MTurk population's domain. On the one hand, our task's simplicity is unlikely to render it particularly domain-specific. On the other hand, our task is performed in our subjects' work environment, where they regularly take risks by spending time on assignments which may be rejected. Unlike Kirchler et al., we do not identify a positive, additive impact of incentives and framing. Yet, robustness of their results to online versus lab-in-the-field settings suggests our online setting cannot explain why we do not find evidence of either impact.

In conclusion, placing our results in relation to literature on the impacts of competition indicates not only that the literature varied in their methods and evidence, but our results add another layer of differential outcomes. Beyond the focus of the literature reviewed, it is possible competition interlinks with personal attributes or preferences. As seen in i.a. Kirchler et al. (forthcoming), students and financial professionals may have differential reactions to framing. Kirchler et al. however do not considered whether this is due to domain-specificity, as discussed above, or heterogeneity in underlying factors. Exploring the second explanation, we turn to examining potential impacts of underlying factors in competition.

9 Heterogeneity in responses to competition

While our main results do not provide any evidence for an impact of competition on effort and risk-taking on average, there may be important heterogeneity in responses to competitive incentives and framing which are not detected in the aggregate. Posit for example, in response to competitive incentives or framing, men increase their risk-taking whereas women decrease theirs. Estimations of average treatment effects would then not find any significant impact of competition, when in fact there may exist effects for subsets of the subject pool. For this reason, we examine some of potential underlying channels for the impact of competition on effort and risk-taking.

In particular, we explore three "channels" through which the impact of competition may run: gender, overconfidence, and risk-aversion. While we build on theoretical models and empirical evidence to provide separate hypotheses for the impact of competitive incentives and competitive framing in our main analysis, the literature on aspects of competition and our channels is more limited. As such, we hypothesise the effects of both aspects are in the same direction. Naturally, this may not be the case, but as our experimental design allows us to separate the two, we provide some initial analysis of their potentially divergent directions.

In the following sections we derive hypotheses for each channel from theoretical and empirical literature, followed by our empirical strategy, results, and discussion.

9.1 Hypotheses

To develop hypotheses for heterogeneous responses to competition we consider *gender*, *overconfidence*, and *risk-aversion* in turn. Firstly, the literature on gender differences in competitiveness, risk-taking and performance is vast. Summarising experimental evidence, Croson and Gneezy (2009) concludes women are generally found to be more averse to competition and risk-taking than men, in particular opting out of competition when possible. In another approach, Niederle and Vesterlund (2007) and Buser and Yuan (2016) find women are less likely to enter into competition than men. Additionally, in mixed gender settings, men's performance (here: effort) increases more than women's as a result of competition (see review by Niederle and Vesterlund, 2011). Men are also found to be more risk-taking than women (Charness et al., 2013). Additionally, for competitive framing, Schmidt et al. (2015) formulate theoretical predictions for how, and experimentally show that, social comparison to increases risk-taking in particular for men. Weighing these results together we hypothesise:

H8 *Women increase risk-taking less than men as a result of competition*

H9 *Women increase effort less than men as a result of competition*

Secondly, overconfidence is an important, and potentially devastating, factor in economic decision-making. Theoretically, overconfidence has been shown to increase effort if individuals overestimate their skill-level and thus the marginal product of their effort. In turn, the increased cost of effort seems smaller than the benefit of expected higher likelihood of getting the bonus, in turn leading to increased effort (Gervais and Goldstein, 2007). These results have been confirmed in experimental studies of spillovers from overconfidence on real effort, ranging from finance to search theory (see e.g. Falk et al., 2006; Prokudina et al., 2015). For risk-taking, theoretical models predicts a positive linkage between overconfidence and risk-taking. Modelling investor behaviour, Odean (1998) shows overestimation of the precision of own valuations lead to excessive trading, which in turn increases risks. Model predictions have been confirmed in observational studies of investors (Barber and Odean, 2001, 2002), but also extend to e.g. results from observational studies of agricultural risk-taking among Ethiopian farmers (Barsbai et al., forthcoming). Lastly, combining choices of risk and effort, Everett and Fairchild (2015) model entrepreneurial behaviour and in turn, finding both risk-taking and effort to increase with overconfidence in competitive environment. We hypothesise:

H10 *When in competition, overconfident individuals increase risk-taking more than others*

H11 *When in competition, overconfident individuals increase effort more than others*

Lastly, uncertainty and risk are both inherent to competition and, as documented by Holt and Laury (2002), the majority of individuals are risk-averse. In turn, risk-averse individuals may react differently to competition than risk-neutral or risk-seeking individuals. Naturally, we argue individuals who prefer to not take risks in general will also take relatively less risk when in competition. For effort, the prediction is however less obvious. The general theoretical literature on correlations between risk-aversion and effort in contests argues for both positive and negative links (Konrad and Schlesinger, 1997). In a theoretical model, however, Millner and Pratt (1991) show that in a symmetric two-player competition, the nature of the utility function determines the sign of the correlation.[29] In Millner and Pratt's experimental application, risk-averse subjects are shown to exert less effort in competition than risk-neutral or risk-seeking. This result replicates across experimental studies Dechenaux et al. (2015), despite the ambiguous theoretical predictions. Essentially, the evidence suggests individuals who are risk-averse

[29] Essentially, if the individual exhibits "prudence", i.e. a positive third derivative of the utility function, one can expect effort to decrease with risk-aversion (Millner and Pratt, 1991).

decrease their risk-taking, and then it is only rational to decrease effort to not incur a loss. As such, we develop the following hypotheses:

H12 *Risk-averse participants increase risk-taking less than others as a result of competition*

H13 *Risk-averse participants increase effort less than others as a result of competition*

Applying the predictions to our experiment, we compare choices made across our treatments and channels. Overall, we expect male, overconfident or risk-liking individuals to choose relatively more effort and higher risk when in *Direct competition* or *Competitive threshold*.

9.2 Empirical strategy

Similarly to the analysis of the main hypotheses, we analyse our channel hypotheses using OLS regressions. To capture heterogeneous effects we introduce two interaction terms between our competitive treatments and our dummy variables for the factor of interest; gender, overconfidence or risk-aversion. To exemplify, for potential gender heterogeneity, the interaction takes value one if a subject is in either of the competitive treatments (*Direct competition* (DC) or *Competitive threshold* (CT), respectively) and female. As such, its coefficient captures both direction and magnitude of any differential impacts of incentives and/or framing on women relative to men. Analogous examples apply to overconfidence as well as risk-aversion.

Separately examining each channel, we use full model regressions with robust standard errors for the dependent variable r_i

$$\begin{aligned} r_i =& \alpha + \beta_2 \times \mathrm{DC}_i + \beta_3 \times \mathrm{CT}_i + \delta_2 \times \mathrm{DC}_i \times \mathit{Channel}_i + \delta_3 \times \mathrm{CT}_i \times \mathit{Channel}_i \\ &+ \gamma_1 K_{ik} + ... + \gamma_7 K_{ik} + \epsilon_i, \end{aligned}$$

and the equivalent specification is used with effort, e_i, as dependent variable. Treatment dummies are included, and variable $\mathit{Channel}_i$ indicates one of three channel variables. All variable choices and regressions were pre-registered. For the channel of gender, we set $\mathit{Channel}_i = \mathit{Female}_i$ and include the full $k = 7$ control variables (see Section 6.1).

For the overconfidence channel, we increase the statistical power by focusing on one measure. Specifically, we use *Overestimation of winning*, which accounts for the difference between a subject's prediction of whether they will and whether they do win. As such, it more correctly captures overconfidence in competition, as it is the only measure which considers subjects' expectations of both their own effort and risk-taking, as well as how their output compares to others'. The other two measures, *Overestimation of performance* and *Overplacement*, only capture subjects' beliefs over their effort and how it compares to the average participant. Focusing on both effort and risk-taking allows us to capture for example if individuals who think they will win increase their risk-taking, leading them to lose despite high effort. As such, we define $\mathit{Channel}_i = \mathit{Overestimation\ of\ winning}_i$. To avoid clouding of the channel measure, we exclude *Overplacement* as a control variable and thus limit the model to $k = 6$ control variables.

Finally, for the risk preference channel we focus on risk-aversion specifically, rather than general risk preferences. We define people as risk-averse if they rate themselves below 5 on a scale from "Not willing to take risks" (0) to "Very willing to take risks" (10).[30] As such, we set $\mathit{Channel}_i = \mathit{Risk\text{-}aversion}_i$ and use *Risk-aversion*, rather than *General risk preferences*, as a control in the $k = 7$ control variable specification. Table C.4.1 provides definitions for all channel variables.

[30] We here follow previous literature, e.g. Ding et al. (2010) and Treibich (2015).

As discussed in the introduction to the section, our hypotheses do not differentiate between effects of competitive incentives and competitive framing, but rather we estimate the impacts of competition in general. For this reason, we use two-sided t-tests to test whether the interactions between each competitive treatment and gender (**H8-9**), overconfidence (**H10-11**), and risk-aversion (**H12-13**) individually have an impact on either effort (e_i) or risk-taking (r_i). We then use an F-test to test whether interaction coefficients for *Direct competition* and *Competitive threshold* differ from each other in each channel.[31] Table 9.1 outlines hypotheses and predictions for each channel. Note, predictions for directions of coefficients are the same for effort and for risk-taking.

Table 9.1: Statistical hypotheses for channels

	Direct competition	Competitive threshold	Differential effects
H_0	$\delta_2 = 0$	$\delta_3 = 0$	$\delta_2 = \delta_3$
H_1	$\delta_2 \neq 0$	$\delta_3 \neq 0$	$\delta_2 \neq \delta_3$
Predictions			
Gender (**H8,9**)	$\delta_2 < 0$	$\delta_3 < 0$	
Overconfidence (**H10,11**)	$\delta_2 > 0$	$\delta_3 > 0$	
Risk-aversion (**H12,13**)	$\delta_2 < 0$	$\delta_3 < 0$	

9.3 Results

An initial exploration of the channels indicate no definitive patterns in average effort or risk-taking in line with either hypotheses or predictions. As shown in Table C.4.2, the gap between genders is widest in *Neutral threshold* (NT) for risk-taking, but smallest in *Competitive threshold* (CT) for effort. Similarly, for the gap in risk-aversion, directions for competition are not clear for either effort or risk-taking. For overconfidence however, the difference in average effort increases with competitive treatments, in line with predictions. Turning to our channel analyses, Table 9.2 provides OLS regressions for risk-taking and effort, separated by channel, while F-statistics for hypotheses tests are presented in the subsequent Table 9.3.

Analysing the gender channel first, columns (1) and (2) of Table 9.2 show results similar to our main analyses: treatment coefficients are mostly in predicted directions yet insignificant. Importantly however, the regressions indicate that there is no evidence of any interaction effect between female gender and either of our competitive treatments. Only one of the interactions is negative as predicted, yet standard errors are too large for either to be significant – a result which holds true for both effort and risk-taking. Therefore, we cannot reject the null hypotheses for **H8** and **H9**. Similarly, as seen by the large p-values in columns (1) and (2) of Table 9.3, we cannot reject the last null hypotheses of identical interaction coefficients for the two treatments, either for effort or for risk-taking, at any relevant level of significance.

Similar conclusions can be drawn for the risk-aversion channel. Here, column (5) and (6) of Table 9.2 and Table 9.3 indicate neither interaction coefficient is significantly different from zero and they are also not significantly different from one another. In contrast to predictions, interactions are mostly positive for both effort and risk-taking, and also here are treatment coefficients unaffected. Similarly here, we cannot reject either **H12** or **H13**. As such, we cannot conclude either females and risk-averse people exert any lower effort or take any lower risks relative to men and risk-neutral/seeking individuals, when in competition.

[31] Note that our pre-analysis plan contained a typo that we were to use a t-test to test whether the coefficients were different from each other. Naturally, we meant to use an F-test.

Table 9.2: OLS regressions: Channels

	Gender differences		Overconfidence		Risk-aversion	
	(1)	(2)	(3)	(4)	(5)	(6)
	Risk-taking	Effort	Risk-taking	Effort	Risk-taking	Effort
DC	0.040	-0.161	0.512	-4.265**	0.140	0.581
	(0.483)	(0.994)	(0.497)	(1.377)	(0.416)	(0.930)
CT	0.707	1.222	0.946*	0.176	0.215	0.473
	(0.490)	(1.028)	(0.440)	(1.284)	(0.410)	(0.968)
Overestimation	0.406	0.670	1.057*	-2.950*	0.538	0.580
of winning	(0.294)	(0.597)	(0.502)	(1.361)	(0.296)	(0.607)
Overplacement	0.007	-0.521***			0.012	-0.517***
	(0.012)	(0.034)			(0.012)	(0.033)
General	0.338***	0.094	0.350***	-0.448*		
risk preferences	(0.061)	(0.125)	(0.059)	(0.175)		
Risk-aversion					-1.605**	0.030
					(0.556)	(1.052)
Female	-0.243	-1.548	-0.293	-2.078*	-0.405	-1.849**
	(0.509)	(1.039)	(0.294)	(0.910)	(0.299)	(0.619)
Female x	0.440	0.173				
DC	(0.704)	(1.480)				
Female x	-0.688	-1.208				
CT	(0.682)	(1.448)				
Overconfidence x			-0.591	7.826***		
DC			(0.695)	(2.100)		
Overconfidence x			-1.317	-0.865		
CT			(0.675)	(2.003)		
Risk-aversion x					0.184	-2.115
DC					(0.798)	(1.579)
Risk-aversion x					0.375	0.443
CT					(0.748)	(1.514)
Constant	2.705***	34.876***	2.213**	40.003***	5.388***	35.302***
	(0.815)	(1.826)	(0.785)	(2.466)	(0.719)	(1.674)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	417	417	417	417	417	417
R^2	0.148	0.559	0.149	0.114	0.113	0.561

Additional controls include Age, USA, and India. *Overconfidence* in interactions is defined as *Overstimation of winning*. Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For overconfidence, however, estimates in column (3) and (4) of Table 9.2 and 9.3 paint a different picture. On the one hand, for risk-taking we cannot reject the null hypotheses that the interaction between overconfidence and competition has an impact which is different from zero (**H10**). On the other hand, for effort, the interaction between *Direct competition* (DC) and overconfidence has a large and positive coefficient, in line with predictions. With a point estimate at 7.826 sliders, the interaction is significant at the 0.1%-level. Moreover, the F-test shows it is highly significantly different from the interaction between overconfidence and the *Competitive threshold* (CT) – an unsurprising result as CT's interaction coefficient is negative and insignificant. As such, we can reject null hypothesis **H11**, concluding that competitive incentives and overestimation of winning interact to increase effort. Furthermore, we find the control coefficient DC becomes negative and significant at the 1%-level when including the interaction. Together, the interpretation is overconfident people exert more effort when facing competitive incentives, while others exert effort more in line with predictions of the impact of competitive incentives.

Table 9.3: F-tests

	Gender		Overconfidence		Risk-aversion	
	Risk-taking (**H8**)	Effort (**H9**)	Risk-taking (**H10**)	Effort (**H11**)	Risk-taking (**H12**)	Effort (**H13**)
F-statistic	2.771	0.852	1.181	15.645	0.059	2.484
	(0.097)	(0.357)	(0.278)	(0.000)	(0.808)	(0.116)

p-values in parentheses.

Additionally, control coefficients for overestimation of winning become negative and significant when including the *Overconfidence* × DC interaction. Illustrating the interpretation, on average, an overconfident individual under competitive incentives completes approximately 3.6 sliders more than an individual who is not overconfident, and approximately 4.9 sliders more than another overconfident individual who is not facing competitive incentives. A result which is not only statistically significant, but also of some economic significance as the average number completed sliders is only 29.

9.4 Analysis

On the whole, we find mixed evidence on heterogeneity in responses to competition for risk-taking and effort. We do not find any evidence that either gender and risk-aversion yield differential impacts of competitive incentives or framing. We do however, find evidence overconfidence in combination with incentives leads to a relatively greater exerted effort. Instrumenting overconfidence by *Overestimation of winning*, we capture overconfidence over outcomes which depend on effort, risk, as well as a random component. Concretely, individuals who believe they will win, but do not, increase effort relative to those with correct beliefs and those who underestimate their winning. This hold across comparisons; when performance of both is evaluated against an unknown target with a rival prize; when compared with both under and overconfident individuals evaluated against a known target, whether it is personal or impersonal. While the result is highly statistically significant, the heterogeneity does not extend to the impact of competitive incentives on risk-taking.

Moreover, when including the interaction, we also find people who are not overconfident with regards to winning act more in accordance with theoretical predictions. While the common prediction is that in face of competitive incentives one should decrease effort (increasing risk-taking), only individuals who do not falsely predict their winning probability follow this. Moreover, predictions for competitive framing (higher effort and higher risk-taking) hold for risk-taking when the interaction for overconfidence is included, but as the interaction is not significant and the effect does not hold in other cases we cannot conclude that framing has any impact. Hence, overconfidence seem to matter more for effort than for risk-taking.

To validate our result, we must consider factors which may have caused it. First, our result may be a consequence of treatment imbalances. As we test joint impacts of individual attributes and treatments, it is possible that unbalanced treatment groups drive our results. For example, in the gender analysis, as women tend to be more risk-averse in general, an over-representation of women in the competitive treatments may yield a decreased risk-taking. If so, the impact we find is not driven by women's responses to competition, but rather by an imbalance in risk-averse preferences. However, as indicated in Table C.1.1, only one variable is significantly different between treatments: *Age*. Second, it may be a false positive (and negative) result. An interaction variable with a predicted effect which is smaller than, or equal to, the main effect quickly requires

a much greater sample size to retain statistical power.[32] However, as no previous research can offer us directly comparable effect sizes, and in particular so for the interaction effect, we cannot perform prior power analysis and estimate the true sample size needed to support our results.

Third, the heterogeneous impact of overconfidence is perhaps a true result, but dependent upon our specific measure. As we preregister and instrument overconfidence with overestimation of winning, the effect does not necessarily extend to other measures for overconfidence which regard only own choices or only effort. Preliminary robustness analysis suggests this may be true and thus we limit our conclusions to overestimation of winning. Despite this, it is reasonable that only overconfidence over effort, risk-taking and the rival's choices captures the impact, as it includes the effect of uncertainty on the trade-off between effort and risk-taking rather than simply incorrect beliefs of own effort. Finally, if overestimation of winning is indeed the channel through which competition impacts effort, then our finding is a true result. Clearly supporting this, our result is significant at the 0.1%-level, in line with predictions and robust to specifying the regression as a probit model, as seen in Table C.4.3 and C.4.4. Hence, while not all kinds of overconfidence may interact with competition, when it does it is likely to have a significant, positive impact on effort.

10 Discussion

Contributing to the evidence on whether choices of risk-taking and effort differ in and outside of competition, our analysis suggests that such research yields different outcomes than current literature. Our results differ from those in studies of risk-taking and effort in isolation, as well as those in studies of effort and risk-taking only within competition. Our overarching, robust finding is that neither competitive incentives nor framing significantly impact risk-taking or effort on average, in contrast to our hypotheses and their underlying theory. Beyond finding no separate impacts, we also do not find any evidence that risk-taking and effort act as either substitutes or complements between subjects. However, our conclusions do show competitive incentives and overestimation of winning interact to increase effort, but not risk-taking – a heterogeneity in responses which extends to neither gender nor risk-aversion.

As such, in an experimental task where subjects have an opportunity to gain a bonus payment if their output is sufficient, including competitive aspects does not seem to change subjects' choices or behaviours on average. In light of our result, we first discuss its theoretical context, followed by a consideration of the possibility that the result is a false null. Finally, we consider potential threats to internal and external validity.

10.1 Theoretical implications

Beyond contrasting previous experimental results, our results clashes particularly with theoretical predictions; both with our own models and with the theoretical literature in general. Assuming individuals understand the mechanism of the threshold evaluation, they should set effort to the threshold value ($\mathcal{T} = 273$, i.e. approximately 30 sliders) and take zero risk. Instead, we find average effort across treatments to be just below 30. As such, effort is also much above the direct competition predictions of minimised effort. For average risk-taking, we find excessive, i.e. non-zero, risk-taking in threshold treatments, but also not maximal risk-taking as predicted for the direct competition treatment. Thus, absent any identifiable differences between treatment averages, it is clear that individuals do not behave according to either model. Additionally, for both

[32] Gelman (2018) exemplifies this with a four times larger sample required for 80% statistical power when effect sizes are equal for interaction and main effect.

threshold treatments, excessive risk-taking is shown to be destructive. Around 18% of subjects end up loosing from excessive risk-taking, when their beliefs and their effort would have led them to meet the threshold and gain the bonus.

Additionally, our results provide no support for the theories of competitive framing. Performance in a competitive setting seems to be no different from performance in a neutral setting. However, our design does not allow us to separate impacts of "contingencies of reinforcement", availability heuristics, or social comparison from each other. Yet, the lack of impact of competitive framing in our results indicates that competitive framing overall leads individuals to behave no different to without competitive framing. Instead, our results suggest effort and risk choices are independent of both framing and incentives. Our results go beyond those of e.g. Dijk et al. (2014) and Kirchler et al. (forthcoming), as we also do not identify an additive impact of framing and incentives and as our main results are insignificant for both effort and risk-taking. In particular, inclusion of a known, competitive threshold should create a direct social reference point with which to compare one's own outcome, potentially inducing disappointment aversion (Gill and Prowse, 2012). However, our results show no such impact.

From examining the treatment outcomes we conclude that competition does not impact individuals on average. Through examining the link between personal characteristics and aspects of competition, we find evidence that responses to competition are persistent to heterogeneity in some, but not all, characteristics. In line with predictions, we find competitive incentives to interact with overconfidence, leading individuals to exert relatively higher effort. The equivalent is not found for gender, risk-aversion, or risk-taking, in contrast to predictions.

Interpreting the result in line with theoretical predictions by Gervais and Goldstein (2007) and Everett and Fairchild (2015) suggests subjects may overestimate the benefit of their marginal productivity of effort, overvaluing it relative to their cost of effort. On the one hand, while our slider task does require physical effort, its marginal effort cost is reasonably low. On the other hand, our experiment interlinks risk and effort choices. Hence, if an individual overvalues their own benefit from correctly placing sliders, they may be incentivised to exert more effort under the belief that doing so will secure them the prize. Such overvaluation can for example be explained by an individual not recognising the inherent variance in the final outcome created by risk-taking or if they falsely believe they can finish more sliders than others in general. Moreover, the significant result does not extend to the link between competition, overconfidence and risk-taking.

Returning to our research question, our concluding answer becomes that, in general, competition does not impact individuals' choice variables. However, there may be underlying differences in responses – a heterogeneity which deserve further research, particularly if the heterogeneity is not random across populations of interest.

10.2 Interpreting the null result

While our result is a null result in general, there are several possible explanations for it which must be considered before drawing a final conclusion. First, it may be that our result is a true null, and that competition does not affect risk-taking and effort on average. Recent research on competition, effort, and risk-taking also find null results, e.g. for effort (Gächter et al., 2017), for risk-taking (Filippin and Gioia, 2017), for substitution (Andersson et al., 2017) and partially for framing (Kirchler et al., forthcoming). Additionally, the failure of economics experiments to replicate (Open Science Collaboration, 2015; Camerer et al., 2016) implies the literature is likely riddled with false positive results. As seen in our power analysis in Section 2.4, when previous studies identify significant outcomes, effect sizes are quite small, requiring substantial sample sizes. In light of this, it is not unlikely that our result is a true null.

Second, while our null result may be true, the general intuition provided by the literature may simply be misguided. While predictions for direct competition, as derived by Hvide (2002), are generally robust to alterations of assumptions, our predictions for threshold evaluation depend on a core assumption. As outlined, we assume the marginal benefit of receiving the prize with certainty is larger than the cost of exerting enough effort to meet the threshold. If not, we would expect individuals in threshold settings to behave akin to those in Hvide (2002), as exerting effort is relatively costly.

Two considerations follow. First, we do not identify no effort and maximised risk-taking in either of the competitive treatments, unlike predictions by Hvide (2002). Second, as the slider task is timed, it requires subjects to wait throughout the time span to be eligible for any payment, even if they move no slider. The added cost of moving a slider is thus arguably low, whereas the bonus payment is comparatively larger than the guaranteed completion payment.[33] As such, our predictions for competitive incentives are unlikely to be misguided, and likewise applies to predictions for competitive framing. Here, the multiple facets of framing are indistinguishable in our experiment, hiding potentially opposing effects. While separation of factors is beyond the scope of this study, it also implies predictions for especially effort could be phrased in the opposite direction. Yet, also no such impact is found.

Thirdly, in contrast to many related studies (e.g. Gill and Prowse, 2012; Andersson et al., 2017; Gächter et al., 2017), we limit our experiment to one round in order to mimic a one-shot decision. As such, our result may be a true null, but learning in games may be an explanation for why competitive incentives are not found to impact risk-taking or effort in our study. With a one-shot decision there is no room for subjects to observe outcomes and adjust towards optimal strategies, whereas multiple decisions rounds may provide a chance for subjects to adapt to model predictions. However, learning in games cannot explain why we do not find evidence of competitive framing. Repeated interactions within a social comparison perspective would require continued matching with the same individual, and is thus rather a story of social history.

Finally, our result may also be a false null, i.e. competition does affect risk-taking and effort, but we have failed to identify the effect. On the one hand, a false null may be explained by too low statistical power to detect any difference. However, we designed our study to have standard 80% power, estimated using closely related studies (Andersson et al., 2017; Eriksen and Kvaløy, 2017; Filippin and Gioia, 2017), and thus argue it is unlikely that low power is a fundamental flaw. On the other hand, our elicitation of effort or risk may be too limited in scope to detect the impact, which we explore further by discussing validity of our results.

10.3 Validity

All methodological choices have benefits and drawbacks which affect reliability and generalisability of our results. This section discusses in particular how our subject pool and our experimental procedure may affect both internal and external validity, and in turn the interpretations of our results.

10.3.1 Internal validity

While experiments in general allow for large internal validity, our choice of an online experimental environment and an MTurk subject pool pose issues for internal validity in two ways. Firstly, as outlined, online experiments are vulnerable to bias from selective attrition. Further threatening this, MTurk workers need to perform many tasks to earn minimum wage, potentially causing selective attrition to increase if competitive settings

[33] The bonus payment is \$0.75 whereas the completion payment is only \$0.5.

are experienced as more exhausting. However, our results show no such bias: attrition is low and equal across groups.

Secondly, pressure to perform many tasks implies MTurker workers spend little time on each assignment which may imply they pay little attention to instructions (Horton et al., 2011; Straub, 2017). Limited attention can increase the level of noise in estimates, as subject simply do not understand experimental instructions. Yet, our large sample size limits the impact of individual noise. However, limited attention also implies subjects may not "receive" the treatment if instructions are not read carefully. Untreated subjects could bias any differential treatment impact downward, thus potentially explain our null result. To mitigate these last two issues, we exclude subjects who fail a majority of attention checks, a common way of minimising attention problems in MTurk experiments (Horton et al., 2011; Straub, 2017).

Another drawback with our sample population is the limited incentives MTurk workers may have to exert effort. As monetary payments are small, one may question the size of actual incentives to exert real effort in order to receive the bonus payment. If only small, effort estimates may be biased downwards across all three treatment groups, with no impact on our final result. Additionally, we find a higher average of completed sliders than previous studies, suggesting the bonus size is sufficient. If instead incentives appear relatively weaker in competitive settings, then this is simply a function of competitive framing and thus correctly captured in our estimates. Moreover, MTurk workers are shown to respond no different than other subjects to incentives (Horton et al., 2011; Amir et al., 2012). Thus, limited incentives concerns are no greater than in other experimental settings, and rather a concern for generalisability to the real world.

Pertinent to experiments is that subjects trust the experimental instructions. If subjects in competitively framed treatments do not believe they are in a competition, the treatment will be ineffective, biasing any treatment differences downward. Targeting this, we inform subjects all information is truthful and all choices have real consequences (see Appendix A.1). Moreover, supporting internal validity rather, our online setting allows us to abstract away from outside effects of social interaction that may cloud the direct impact of competition on effort and risk-taking. Situations with interaction history between competitors, e.g. workplaces, would yield more noisy estimates, as competitive framing here cannot be isolated from social history.

Another core experimental choice is our between-subjects design, rather than a within-subject design where subjects face all treatments in subsequent rounds. A downside of our choice is its reduction of any results' statistical power, but a benefit is the avoidance of contamination between treatments. Additionally, it allows for generalisations to cases where individuals cannot face the same choices in different contexts, e.g. if you are paid on relative performance you are unlikely to also be paid without competitive incentives. Crucial for a between-subject design, however, is a true randomisation process (Charness et al., 2012). As we have full control over the computerised randomisation and find no evidence of selective attrition, we are confident this condition holds.

10.3.2 External validity

Finally, the generalisability of our results must be considered. Similarly to laboratory experiments, the subject pool of MTurk is not representative of general populations. While experiments commonly test student samples (Henrich et al., 2010), MTurk samples consists of subjects residing primarily in the US and India (Difallah et al., 2018). Additionally, the MTurk population have selected into online work, in contrast to most workers.[34] However, experiments conducted on MTurk show comparable results to both online and face-to-face surveys (Clifford et al., 2015) and MTurk has been successfully

[34] However, not all workers have MTurk as their main source of income (Difallah et al., 2018), similar to how student subjects also have other incomes than from experiments.

used to replicate laboratory experiments (Horton et al., 2011; Amir et al., 2012; Arechar et al., 2017). As such, issues of external validity caused by our subject pool is not a core concern.

Moreover, external validity requires generalisability to other settings. Our real effort task allows extrinsic motivation to drive responses as sources of intrinsic motivation are not relevant (Gill and Prowse, 2012). As such, our effort results may not extend to situations in which intrinsic motivation is fundamental, for example to effort of artists or researchers. Hence, while focusing on extrinsic motivation provides a clearer measure of effort, it impedes generalisability to some degree. Nonetheless, there is no shortage of situations in which the work is dull and tedious, regardless of its setting. Moreover, similar criticism applies to "theoretical" effort tasks, e.g. investment in the mean of the output variable, instead of exerting real effort (e.g. Andersson et al., 2017). In contrast, our task has the advantage of capturing a more realistic effort type.

Furthermore, the domain of choices in our task may limit generalisability to other populations. Hanoch et al. (2006) and Charness et al. (2013) show risk-taking behaviours are not stable across domains and that subjects who are risk-taking in one domain may not be risk-taking in others. As our subjects are familiar with computer-based tasks, their domain-specific preferences may bias our risk-taking estimates upwards relative to a general population. However, as computers are ubiquitous in today's society, we believe similar results hold true for most populations. Furthermore, the Gneezy and Potters (1997) task is more akin to gambling than computer-specific tasks, a domain which is not specific to MTurk or to online behaviours. While implying that those who gamble otherwise may also choose higher risks in our experiment, our results are robust to controlling for risk-aversion in gambling. Hence, domain-specific risk-measures are a lesser concern.

In conclusion, while our design allows for control over experimental components and internal validity, caveats of external validity must be kept in mind when interpreting the results. However, our design allows us to focus on the mechanism of competition, and to abstract from potentially confounding factors. Although our experiment provides a simple way to identify causal effects, it is possible that confounding factors, such as being observed by others, matter for real-world impacts of competition on risk-taking and effort. Competition is rarely as simple as the slider task and the lottery investment in our study, and as a result, our findings must be interpreted with caution and in light of evidence from other studies.

11 Conclusion

This thesis employs an online, experimental design with 417 subjects to investigate whether competition between individuals affects risk-taking and effort. Our results suggest the decisions we make under competition are no different from the choices we make when performing in non-competitive situations. Particularly, our study contributes new, contrasting, evidence on the impact of two aspects of competition: competitive incentives and framing. To our knowledge, we are first to isolate the separate, as well as additive, impact of these two aspects on risk-taking and effort. As our core results are consistently robust, not driven by specific control variables and resistant to concerns of validity, they contribute new considerations for the literature. Additionally, we go further by analysing potential heterogeneity in responses to competition, finding competitive incentives to increase exerted effort among individuals who exhibit overestimation of winning. However, no further evidence of heterogeneous responses to competition is identified, either across genders or risk preferences, or for risk-taking. Nevertheless, our results suggest the implications of competitive incentives are not unrelated to personal characteristics.

In turn, our study provides answers to concerns about the impact of competition. Previous literature suggests competition gives rise to negative externalities through increasing risk-taking as well as create discrepancies in principal-agent contracts. However, our results indicate that neither of these possibilities should be feared if competition is introduced. Individuals take excessive risks also when performing against a fixed, impersonal target, and risk-taking is not further increased in competition. Moreover, individuals also exert effort when evaluated against a threshold, but the magnitude does not increase with competition, as many hope. Our evidence thus suggests negative externalities are persistent and must be mitigated in other ways than by simply removing competition. Additionally, it also indicates alternative metrics should be used by principals who seek to increase agents' effort.

As such, our conclusion yields implications for the literature on relative performance evaluation in two ways. First, as competition is not found to increase excessive risk-taking, calls for regulating competitive incentives to reduce risk-taking in the financial industry may be misguided. Second, RPE may simply be lacking from observed compensation contracts as our results suggest competition is no better than exogenous thresholds for increasing effort on average. A caveat, however, is that if overestimation of winning is more common in for example financial industries, RPE may potentially live up to its hopes of incentivising greater effort, without higher risk-taking here.

Beyond implications for policy, our results open up venues for future research. Firstly, it emphasises the importance of replication of previous studies and experimental designs. Casting doubt on the differential impact of competition, our result becomes one in a growing line of findings where previous literature fail to replicate in other settings (Open Science Collaboration, 2015; Camerer et al., 2016). While not a direct replication, as our design combines tasks from Gill and Prowse (2012) with that of Gneezy and Potters (1997), our null result for competition joins that of Gächter et al. (2017).

Secondly, our evidence provides suggestions for future experimental research. While the link between overconfidence and effort is analysed extensively in theoretical studies, empirical evidence is more limited. The indication in our results of an interplay between effort, overestimation of winning, and competitive incentives warrants further exploration, in particular through experiments explicitly designed to test this link. Additionally, beyond the three specific channels of heterogeneity we focus on, future research may consider e.g. culture, ability asymmetries, or other-regarding preferences. Moreover, our lack of significant differential impacts of both incentives and framing on effort and risk-taking suggest that studies of the average impact of competition can employ either aspect without loss of generalisability.

Finally, our conclusions highlight the need for further understanding of competition's different aspects. As our results contrast predictions for differences between competitive and non-competitive situations on average, future research may consider to further test components of these predictions. For example, factors of competitive framing, such as availability heuristics and social comparison, may not impact risk-taking and effort in the same direction. If so, experimental designs which invoke one or the other could potentially separate the impacts. Alternatively, components of competition beyond effort and risk-taking warrant exploring to understand if the lack of impact remains. In particular, the degree of competition could be explored by testing differences between our two-person case and an increased number of agents.

As such, while competition is ubiquitous in our lives, our conclusions indicate that individuals in competition do not risk it all. Rather individuals take some risk and exert some effort, but the levels are no different outside of competition.

Bibliography

Agranov, M., Bisin, A. and Schotter, A. (2013), 'An experimental study of the impact of competition for other people's money: The portfolio manager market', *Experimental Economics* **17**(4), 564–585.

Alain, C., Ernst, F., Benedikt, H. and Frédéric, S. (2014), 'Social comparison and effort provision: Evidence from a field experiment', *Journal of the European Economic Association* **12**(4), 877–898.

Amir, O., Rand, D. G. and Gal, Y. K. (2012), 'Economic games on the internet: The effect of $1 stakes', *PLOS ONE* **7**(2), 1–4.

Andersson, O., Holm, H. J., Tyran, J.-R. and Wengström, E. (2013), Risking other people's money: Experimental evidence on bonus schemes, competition, and altruism, Working Paper Series 989, Research Institute of Industrial Economics.

Andersson, O., Holm, H. J. and Wengström, E. (2017), Grind or gamble? An experimental analysis of effort and spread seeking in contests, Working Paper Series 1149, Research Institute of Industrial Economics.

Apicella, C. L., Dreber, A., Gray, P. B., Hoffman, M., Little, A. C. and Campbell, B. (2011), 'Androgens and competitiveness in men', *Journal of Neuroscience, Psychology, and Economics* **4**(1), 54–62.

Arechar, A. A., Gächter, S. and Molleman, L. (2017), Conducting interactive experiments online, IZA Discussion Papers 10517, Institute for the Study of Labor (IZA).

Barber, B. M. and Odean, T. (2001), 'Boys will be boys: Gender, overconfidence, and common stock investment', *Quarterly Journal of Economics* **116**(1), 261–292.

Barber, B. M. and Odean, T. (2002), 'Online investors: Do the slow die first?', *The Review of Financial Studies* **15**(2), 455–488.

Barsbai, T., Schmidt, U. and Zirpel, U. (forthcoming), Overconfidence and risk-taking in the field: Evidence from Ethiopian Farmers, Working paper, Kiel Institute for the World Economy (IfW).

Becker, B. E. and Huselid, M. A. (1992), 'The incentive effects of tournament compensation systems', *Administrative Science Quarterly* **37**(2), 336–350.

Brown, K. C., Harlow, W. V. and Starks, L. T. (1996), 'Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry', *Journal of Finance* **51**(1), 85–110.

Buser, T. and Dreber, A. (2016), 'The flipside of comparative payment schemes', *Management Science* **62**(9), 2626–2638.

Buser, T. and Yuan, H. (2016), Do Women give up competing more easily? Evidence from the lab and the Dutch math olympiad, Tinbergen Institute Discussion Papers 16-096/I, Tinbergen Institute.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M. and Wu, H. (2016), 'Evaluating replicability of laboratory experiments in economics', *Science* **315**(6280), 1433–1436.

Charness, G. and Gneezy, U. (2012), 'Strong Evidence for Gender Differences in Risk Taking', *Journal of Economic Behavior & Organization* **83**(1), 50–58.

Charness, G., Gneezy, U. and Imas, A. (2013), 'Experimental methods: Eliciting risk preferences', *Journal of Economic Behavior & Organization* **87**(C), 43–51.

Charness, G., Gneezy, U. and Kuhn, M. A. (2012), 'Experimental methods: Between-subject and within-subject design', *Journal of Economic Behavior & Organization* **81**(1), 1–8.

Charness, G., Masclet, D. and Villeval, M. C. (2014), 'The dark side of competition for status', *Management Science* **60**(1), 38–55.

Clifford, S., Jewell, R. M. and Waggoner, P. D. (2015), 'Are samples drawn from Mechanical Turk valid for research on political ideology?', *Research and Politics* **2**(4), 1–9.

Croson, R. and Gneezy, U. (2009), 'Gender differences in preferences', *Journal of Economic Literature* **47**(2), 448–474.

Dechenaux, E., Kovenock, D. and Sheremeta, R. (2015), 'A survey of experimental research on contests, all-pay auctions and tournaments', *Experimental Economics* **18**(4), 609–669.

Dekel, E. and Scotchmer, S. (1999), 'On the evolution of attitudes towards risk in winner-take-all games', *Journal of Economic Theory* **87**(1), 125–143.

Difallah, D., Filatova, E. and Ipeirotis, P. (2018), Demographics and dynamics of mechanical turk workers, *in* 'Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining', WSDM '18, ACM, New York, NY, USA, pp. 135–143.

Dijk, O., Holmen, M. and Kirchler, M. (2014), 'Rank matters – The impact of social competition on portfolio choice', *European Economic Review* **66**(C), 97–110.

Ding, X., Hartog, J. and Sun, Y. (2010), Can we measure individual risk attitudes in a survey?, IZA Discussion Papers 4807, Institute for the Study of Labor (IZA).

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.

Downs, G. W. and Rocke, D. M. (1994), 'Conflict, agency, and gambling for resurrection: The principal-agent problem goes to war', *American Journal of Political Science* **38**(2), 362–380.

Dreber, A., Rand, D., Wernerfelt, N., Garcia, J., Vilar, M., Lum, J. and Zeckhauser, R. (2011), 'Dopamine and risk choices in different domains: Findings among serious tournament bridge players', *Journal of Risk and Uncertainty* **43**(1), 19–38.

Eriksen, K. W. and Kvaløy, O. (2014), 'Myopic risk-taking in tournaments', *Journal of Economic Behavior & Organization* **97**, 37–46.

Eriksen, K. W. and Kvaløy, O. (2017), 'No guts, no glory: An experiment on excessive risk-taking', *Review of Finance* **21**(3), 1327–1351.

Everett, C. R. and Fairchild, R. J. (2015), 'A theory of entrepreneurial overconfidence, effort, and firm outcomes', *Journal of Entrepreneurial Finance* **417**(1), 1–26.

Fafchamps, M., Kebede, B. and Zizzo, D. J. (2015), 'Keep up with the winners: Experimental evidence on risk taking, asset integration, and peer effects', *European Economic Review* **79**(C), 59–79.

Falk, A., Huffman, D. B. and Sunde, U. (2006), Self-confidence and search, IZA Discussion Papers 2525, Institute for the Study of Labor (IZA).

Falk, A. and Ichino, A. (2006), 'Clean evidence on peer effects', *Journal of Labor Economics* **24**(1), 39–58.

Fehr, E. and Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *Quarterly Journal of Economics* **114**(3), 817–868.

Filippin, A. and Gioia, F. (2017), Competition and subsequent risk-taking behaviour: Heterogeneity across gender and outcomes, IZA Discussion Papers 10792, Institute for the Study of Labor (IZA).

Frydman, C. and Jenter, D. (2010), 'CEO compensation', *Annual Review of Financial Economics* **2**(1), 75–102.

Föllmi, R., Legge, S. and Schmid, L. (2016), 'Do professionals get it right? Limited attention and risk-taking behaviour', *Economic Journal* **126**(592), 724–755.

Gaba, A., Tsetlin, I. and Winkler, R. L. (2004), 'Modifying variability and correlations in winner-take-all contests', *Operations Research* **52**(3), 384–395.

Gamba, A., Manzoni, E. and Stanca, L. (2017), 'Social comparison and risk taking behavior', *Theory and Decision* **82**(2), 221–248.

Garicano, L. and Palacios-Huerta, I. (2014), Sabotage in tournaments: Making the beautiful game a bit less beautiful, *in* I. Palacios-Huerta, ed., 'Beautiful game theory', Princeton University Press, Princeton, NJ, USA, chapter 8, pp. 124–151.

Gelman, A. (2018), 'You need 16 times the sample size to estimate an interaction than to estimate a main effect'. *http://andrewgelman.com/2018/03/15/need-16-times-sample-size-estimat -interaction-estimate-main-effect/* [Accessed 2018-05-04].

Gervais, S. and Goldstein, I. (2007), 'The positive effects of biased self-perceptions in firms', *Review of Finance* **11**(3), 453–496.

Gill, D. and Prowse, V. (2012), 'A structural analysis of disappointment aversion in a real effort competition', *American Economic Review* **102**(1), 469–503.

Gill, D., Prowse, V. and Vlassopoulos, M. (2013), 'Cheating in the workplace: An experimental study of the impact of bonuses and productivity', *Journal of Economic Behavior & Organization* **96**(C), 120–134.

Gilpatric, S. M. (2009), 'Risk taking in contests and the role of carrots and sticks', *Economic Inquiry* **47**(2), 266–277.

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

Gächter, S., Huang, L. and Sefton, M. (2017), 'Disappointment aversion and social comparisons in a real-effort competition', *Economic Inquiry* . doi: 10.1111/ecin.12498 [Accessed 2018-05-12].

Gächter, S., Nosenzo, D. and Sefton, M. (2013), 'Peer effects in pro-social behavior: Social norms or social preferences?', *Journal of the European Economic Association* **11**(3), 548–573.

Gächter, S. and Thöni, C. (2010), 'Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis', *Journal of Economic Behavior & Organization* **76**(3), 531–543.

Hanoch, Y., Johnson, J. G. and Wilke, A. (2006), 'Domain specificity in experimental measures and participant recruitment', *Psychological Science* **17**(4), 300–304.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C. and Bigham, J. P. (2018), A data-driven analysis of workers' earnings on amazon mechanical turk, paper no. 449, *in* 'Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems', CHI '18, ACM, New York, NY, USA, pp. 1–14.

Henrich, J., Heine, S. J. and Norenzayan, A. (2010), 'The weirdest people in the world?', *Behavioral and Brain Sciences* **33**(2-3), 61–135.

Hill, S. E. and Buss, D. M. (2010), 'Risk and relative social rank: Positional concerns and risky shifts in probabilistic decision-making', *Evolution and Human Behavior* **31**, 219–226.

Holmström, B. (1979), 'Moral hazard and observability', *Bell Journal of Economics* **10**(1), 74–91.

Holmström, B. (1982), 'Moral hazard in teams', *Bell Journal of Economics* **13**(2), 324–340.

Holt, C. A. and Laury, S. K. (2002), 'Risk aversion and incentive effects', *American Economic Review* **92**(5), 1644–1655.

Horton, J., Rand, D. and Zeckhauser, R. (2011), 'The online laboratory: Conducting experiments in a real labor market', *Experimental Economics* **14**(3), 399–425.

Hvide, H. K. (2002), 'Tournament rewards and risk taking', *Journal of Labor Economics* **20**(4), 877–898.

HyLown Consulting (2018), 'Compare 2 means: 2-sample, 2-sided equality'. *"http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equality* [Accessed 2018-04-03].

Ipeirotis, P. G. (2010), Demographics of Mechanical Turk, NYU Working Paper CEDER-10-01, New York University.

Jensen, M. C. and Murphy, K. J. (1990), 'Performance pay and top-management incentives', *Journal of Political Economy* **98**(2), 225–264.

Jullien, B., Salanie, B. and Salanie, F. (1999), 'Should more risk-averse agents exert more effort?', *Geneva Risk and Insurance Review* **24**(1), 19–28.

Kahneman, D. and Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**(2), 263–291.

Kirchler, M., Lindner, F. and Weitzel, U. (forthcoming), 'Rankings and risk-taking in the finance industry', *Journal of Finance* . doi: 10.2139/ssrn.2760637 [Accessed 2018-05-10].

Kleinlercher, D., Huber, J. and Kirchler, M. (2014), 'The impact of different incentive schemes on asset prices', *European Economic Review* **68**(C), 137–150.

Konrad, K. A. and Schlesinger, H. (1997), 'Risk aversion in rent-seeking and rent-augmenting games', *Economic Journal* **107**(445), 1671–1683.

Koszegi, B. and Rabin, M. (2007), 'Reference-dependent risk attitudes', *American Economic Review* **97**(4), 1047–1073.

Kräkel, M. (2008), 'Optimal risk taking in an uneven tournament game with risk averse players', *Journal of Mathematical Economics* **44**(11), 1219–1231.

Kräkel, M. and Sliwka, D. (2004), 'Risk taking in asymmetric tournaments', *German Economic Review* **5**(1), 103–116.

Lazear, E. P. and Rosen, S. (1981), 'Rank-order tournaments as optimum labor contracts', *Journal of Political Economy* **89**(5), 841–864.

Lezzi, E., Fleming, P. and Zizzo, D. J. (2015), Does it matter which effort task you use? A comparison of four effort tasks when agents compete for a prize, CBESS Discussion Paper 15-05, School of Economics, University of East Anglia.

Linde, J. and Sonnemans, J. (2012), 'Social comparison and risky choices', *Journal of Risk and Uncertainty* **44**(1), 45–72.

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

Millner, E. L. and Pratt, M. D. (1991), 'Risk aversion and rent-seeking: An extension and some experimental evidence', *Public Choice* **69**(1), 81–92.

Moore, D. and Healy, P. (2008), 'The trouble with overconfidence', *Psychological review* **115**(2), 502–517.

Mueller, A. (2007), Impact of overoptimism and overconfidence on economic behavior: Literature review, measurement methods and empirical evidence, Diploma thesis, Otto Besheim School of Management.

Mutz, D. C., Pemantle, R. and Pham, P. (2018), 'The perils of balance testing in experimental design: Messy analyses of clean data', *American Statistician* pp. 1–11. doi: 10.1080/00031305.2017.1322143 [Accessed 2018-05-10].

Nalebuff, B. J. and Stiglitz, J. E. (1983), 'Information, competition, and markets', *American Economic Review* **73**(2), 278–283.

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

Niederle, M. and Vesterlund, L. (2011), 'Gender and competition', *Annual Review of Economics* **3**(1), 601–630.

Nieken, P. (2010), 'On the choice of risk and effort in tournaments-experimental evidence', *Journal of Economics & Management Strategy* **19**(3), 811–840.

Nieken, P. and Sliwka, D. (2010), 'Risk-taking tournaments – theory and experimental evidence', *Journal of Economic Psychology* **31**(3), 254–268.

Odean, T. (1998), 'Volume, volatility, price, and profit when all traders are above average', *Journal of Finance* **53**(6), 1887–1934.

Open Science Collaboration (2015), 'Estimating the reproducibility of psychological science', *Science* **349**(6251), 943.

Prokudina, E., Renneboog, L. and Tobler, P. (2015), Does confidence predict out-of-domain effort?, Discussion Paper 2015-055, Tilburg University, Center for Economic Research.

Schmidt, U., Friedl, A. and Lima de Miranda, K. (2015), Social comparison and gender differences in risk taking, Kiel Working Papers 2011, Kiel Institute for the World Economy (IfW).

Skinner, B. F. (1969), *Contingiencies of reinforcement*, B. F. Skinner Foundation, Cambridge, MA, USA.

Stewart, N., Ungemach, C., Harris, A. J. L., Bartels, D. M., Newell, B. R., Paolacci, G. and Chandler, J. (2015), 'The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers', *Judgment and Decision Making* **10**(5), 479–491.

Stigler, G. J. and Becker, G. S. (1977), 'De gustibus non est disputandum', *American Economic Review* **67**(2), 76–90.

Straub, T. J. (2017), Incentive engineering for collaborative online work: The case of crowdsourcing, Doctoral thesis, Department of Economics and Management, Karlsruhe Instute of Technology.

Thöni, C. and Gächter, S. (2015), 'Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis', *Journal of Economic Psychology* **48**(C), 72 – 88.

Treibich, C. (2015), Are survey risk aversion measurements adequate in a low income context?, AMSE Working Papers 1517, Aix-Marseille School of Economics, Marseille, France.

Tsetlin, I., Gaba, A. and Winkler, R. L. (2004), 'Strategic choice of variability in multiround contests and contests with handicaps', *Journal of Risk and Uncertainty* **29**(2), 143–158.

Tversky, A. and Kahneman, D. (1973), 'Availability: A heuristic for judging frequency and probability', *Cognitive Psychology* **5**, 207–232.

A Experiment instructions

A.1 Online experiment instructions

*Each of the following sections represents the full text included on each slide/page in the MTurk and Qualtrics survey. Instructions in grey are not included in the actually survey but used here for explanation. For sections where alternative instructions are used in the different treatment groups these are surrounded by three stars, e.g. ***text***, and the text for the neutral threshold treatment is presented. Alternative instructions for the competitive target and the direct competition groups then follow after the full slide.*

[Amazon Mechanical Turk HIT information]
Instructions:

We are conducting an academic survey on economic choices and behaviors. The survey takes about 8 minutes to complete and is conducted on another web page.

If you complete the task, you will earn $0.50, and have a high probability of receiving a bonus payment of $0.75. At the end of the survey, you will receive a code to paste into the box below to complete the HIT.

You will **only be able to take the survey once**, so please don't attempt it through different HITs.

When ready, click the link below to begin the task.

Survey link:

Make sure to leave this window open as you complete the survey. When you are finished, you return to this page to paste the code into the box to finish the HIT.

Area to fill in code

[Page 1]
This experiment is conducted as a part of a research project at the Stockholm School of Economics. By participating you make a contribution to research, but you also have the opportunity to earn money!

Participation will take around 8 minutes. The HIT consists of three parts; a preparation phase with instructions and a trial task, an experimental phase, and a final questionnaire.

All information and choices in this study are real. Your choices will remain confidential and be used for research purposes only. Remember to answer all questions truthfully.

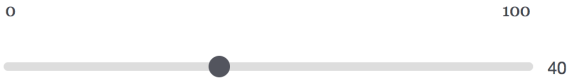
You will be asked a number of Attention Questions throughout the survey. You need to answer the majority of these correctly to be paid for the HIT.

Thank you for participating!

Please click the button below to continue.

[Page 2]

Your main task is to position sliders correctly on a line. The picture below shows a slider placed at the 40-mark:



On the next page you can practice the task. Please click the button to continue.

[Page 3]
Below are two sliders. The current position of the slider is shown to the right of the line, and the instructions for where to position it is shown on the left. You can adjust a slider as many times as you want.

You need to **place both sliders correctly** before you continue.

Place the first slider at 40 and the second slider at 50.

Once correct, please click the button to continue.

[Page 4]
Good job! For completing this step you will be paid $0.50 at the end of the HIT.

Please click to continue.

[Page 5]
In this step, your goal is to gain points by correctly placing sliders. You will be given **2 minutes** to place as many sliders as possible. Each correctly placed slider adds **9 points** towards your total.

***You will perform against a **threshold**.

If you collect **more points** than the threshold which is 273 points, you are paid a **bonus payment** of $0.75, in addition to a completion payment of $0.50.***

Attention question: What is the highest possible **total (completion and bonus) payment** from this HIT? (Checking question for understanding)
Answer in total $. Free-answer question.
If you do not get most of these attention questions correct, you will **not be paid** for the HIT.

Check your answer and then click for the next instructions.

ALTERNATIVE: COMPETITIVE THRESHOLD GROUP
You will compete against **another participant**.

If you collect **more points** than your opponent who got 273 points, you are paid a **bonus payment** of $0.75, in addition to a completion payment of $0.50.

ALTERNATIVE: DIRECT COMPETITION GROUP
You will compete against **another participant**.

If you collect **more points** than your opponent, you are paid a **bonus payment** of $0.75, in addition to a completion payment of $0.50.

[Page 6]
Before starting the task, you can choose to **bet a part of the points** you earn from each slider in the following lottery:

You have a 50% chance of winning 2.5 *times the amount you bet and a 50% chance of loosing the amount you bet.*

The points you do not bet in the lottery go directly to your total points.

Remember, to get the bonus payment $0.75 you need **more points than the threshold total points, 273.

The next page gives some examples. After that you can choose how much to bet.

Please click the button to continue.

COMPETITIVE THRESHOLD GROUP INSTRUCTIONS
Remember, to get the bonus payment $0.75 you need **more points than your competitor's points, 273.**

DIRECT COMPETITION GROUP INSTRUCTIONS
Remember, to get the bonus payment $0.75 you need **more points than your competitor's points**.

[Page 7]
***Example 1**
You finish 10 sliders and bet all 9 points from each slider. The outcome of the lottery is positive, so you get total points 225. The threshold is 112 points, so you get final payment $1.25.

- If the outcome of the lottery instead is negative, you get 0 points. This is less than the threshold 112 points, so your final payment is $0.50.

Attention question: How much did you bet in the example?
Drop-down-menu

Example 2
You finish 40 sliders and bet 0 points from each slider. The outcome of the lottery is positive, so you get 360 points in total. The threshold is 400 points, so you get final payment $0.50.

- If you instead bet 5 of your points from each slider, then your total points would be 660 which is above the threshold 400, so you get final payment $1.25.

Attention question: Was the outcome of the lottery positive in Example 2?
Yes-no-answer***

Please click the button to continue.

ALTERNATIVE: COMPETITIVE THRESHOLD GROUP
Example 1
You finish 10 sliders and bet all 9 points from each slider. The outcome of the lottery is positive, so you get total points 225. Your opponent got 112, so you win and get final payment $1.25.

- If the outcome of the lottery instead is negative, you get 0 points. This is less than your opponent's 112, so you lose and your final payment is $0.50.

Attention question: How much did you bet in the example?
Drop-down-menu with 0 to 9 points

Example 2
You finish 40 sliders and bet 0 points from each slider. The outcome of the lottery is positive, so you get 360 points in total. Your opponent got 400 points, so you lose and get final payment $0.50.

- If you instead bet 5 of your points from each slider, then your total points would be 660 which is above your competitor's 400, so you win and get final payment $1.25.

Attention question: Was the outcome of the lottery positive in Example 2?
Yes-no-answer

ALTERNATIVE: DIRECT COMPETITION GROUP
Example 1
You finish 10 sliders and bet all 9 points from each slider. The outcome of the lottery is positive, so you get total points 225. Your opponent got fewer points, so you win and get final payment $1.25.

- If the outcome of the lottery instead is negative, you get 0 points. This is less than your opponent's points, so you lose and your final payment is $0.50.

Attention question: How much did you bet in the example?
Drop-down-menu

Example 2
You finish 40 sliders and bet 0 points from each slider. The outcome of the lottery is positive, so you get 360 points in total. Your opponent got more points, so you lose and get final payment $0.50.

- If you instead bet 5 of your points from each slider, then your total points would have been 660 which is above your competitor's points, so you win and your final payment is $1.25.

Attention question: Was the outcome of the lottery positive?
Yes-no-answer

[Page 8]
Now **you can decide** how much to bet from each correct slider.

Each slider gives 9 **points**.
Choose how many points from each **you wish to bet** below: drop-down menu.
Choose your bet in the list, between 0 and 9*.*

Once decided, please click the button to continue.

[Page 9]
You will soon perform the slider task.

Before you start, answer the following three questions.

1. How many sliders do you think you will place correctly?
   Choose any number between 0 and 60.
   Drop-down menu, between 0 and 60.
2. How many sliders do you think people on average place correctly?
   Choose any number between 0 and 60
   Drop-down menu, between 0 and 60.
3. ***Do you think you will get as many or more points than the threshold?
   Yes-no-answer***

You will have 2 minutes to complete as many sliders as possible starting from when you click to the next screen. Please click to start.

ALTERNATIVE: COMPETITIVE THRESHOLD & DIRECT COMPETITION GROUPS

3. Do you think you will get as many or more points than your opponent? Yes-no-answer

---

[Page 10]
Complete as many sliders as possible within 2 minutes.

Sliders on screen until time is up and clock showing time left.

[Page 11]
You have now completed the experimental phase, but to be paid you will need to answer a few questions.

Your answers will not affect whether you receive the bonus payment or not, but you **need to complete** the questionnaire in order to be payed for the HIT.

Remember to answer truthfully.

See Appendix A.2 for the questions.

Please click the button to continue.

[Page 12]
For the last step, we will supply you with an MTurk Completion Code. Copy and paste this number into the field on MTurk to validate the HIT.

Thank you for participating!

Click to get your completion code.

A.2 Questionnaire

General questions

1. What is your gender?
   - Female
   - Male
   - Other
2. How old are you?
   Answer in whole years, for example 25 or 56.
3. In which country do you live?
   - Drop-down box with all countries.

Risk preference questions

1. How do you see yourself: are you generally a person who is fully prepagray to take risks or do you try to avoid taking risks?
   Answer by dragging your slider to the value of your choice
   - Scaled answer from 0 to 10, with 0 being "Not willing to take risks" and 10 "Very willing to take risks"
2. Have you participated in any gambling activities, such as betting on sports or visiting a casino, in the last month? Yes/no answer
3. Have you ever started your own company? Yes/no answer

B Proof for models of competitive incentives

The following section provides additional information and proof for the model of direct competition (Section 3.1) and the model of threshold evaluaiton (Section 3.2).

B.1 Proof of uniqueness in Hvide (2002)

Providing a full proof for Proposition 1 we here expand the four original cases considered by Hvide (2002) which were not included in the main text:

(3) $e_{i,DC} < e_{j,DC}$: As j exerts higher effort than i, agent i chooses risk-level $\sigma^2_{i,DC} = \infty$ to maximise the probability of winning – which will then be 50% as outlined above. As $\sigma^2_{i,DC} = \infty$, the level of effort becomes irrelevant for the winning probability and hence i chooses $e_{i,DC} = 0$ to minimise cost. However, as shown above, the best response for agent j to $e_{i,DC} = 0, \sigma^2_{i,DC} = \infty$ is $e_{j,DC} = 0, \sigma^2_{j,DC} = \infty$, which contradicts $e_{i,DC} < e_{j,C}$. Hence, $e_{i,DC} < e_{j,DC}$ cannot be a Nash equilibrium.

(4) $e_{j,DC} < e_{i,DC}$: Due to symmetry and (1), $e_{j,DC} < e_{i,DC}$ is not a Nash equilibrium.

(5) $e_{i,DC} = e_{j,DC} > 0$: If $e_{i,DC} = e_{j,DC}$, each agent's probability of winning is 50%. However, as there is a positive cost of effort, either agent can improve her utility by choosing $e_{i,DC} = 0, \sigma^2_{i,DC} = \infty$, and retain the 50 % probability of winning, but with a lower effort cost. Hence, $e_{i,DC} = e_{j,DC} > 0$ cannot be a Nash equilibrium.

(6) $e_{i,DC} = e_{j,DC} = 0$ and $\sigma^2_{i,DC}, \sigma^2_{j,DC} < \infty$: As the risk-levels chosen by both agents are finite, an agent can increase her probability of winning by exerting any positive effort as the probability of winning is strictly increasing in effort as long as risks are finite $\sigma_{i,DC}, \sigma_{j,DC} < \infty$. We obtain the optimal choice of effort by rearranging Equation 1

$$V'(e_{i,DC}) = f(e_{i,DC} - e_{j,DC}) \times W > 0 \tag{3}$$

for any $e_{i,DC} > 0$. Hence, as $V'(0) = 0$ and $V'(e_{i,DC}) > 0$, agent i will exert positive effort. This contradicts $e_{i,DC} = e_{j,DC} = 0$, and as such $\{e_{i,DC} = e_{j,DC} = 0$, $\sigma^2_{i,DC}$ and $\sigma^2_{j,DC} < \infty\}$ cannot be a Nash equilibrium.

B.2 Exclusion of cases in model of threshold evaluation

Table B.2.1 below shows the possible combinations of effort and risk-taking in the model for threshold evaluation and acts as an illustrative supplement to the exclusion of cases in search for the optimal solution.

Table B.2.1: Visualization of Choice Combinations

	$e_{i,T} = 0$	$e_{i,T} \in (0, \mathcal{T})$	$e_{i,T} = \mathcal{T}$	$e_{i,T} > \mathcal{T}$
$\sigma^2_{i,T} = 0$	Excluded (3)	Excluded (3)	Optimal if $\frac{W}{2} \geq V(\mathcal{T})$	Excluded (1)
$\sigma^2_{i,T} \in (0, \infty)$	Excluded (3)	Excluded (3)	Excluded (2)	Excluded (2)
$\sigma^2_{i,T} = \infty$	Optimal if $\frac{W}{2} < V(\mathcal{T})$	Excluded (4)	Excluded (2)	Excluded (2)

Below we provide a formal exclusion of ten of the cases when optimising the agent's choices. The ten cases can be divided into four subgroups:

(1) $\left\{e_{i,T} \geq \mathcal{T}, \sigma^2_{i,T} = 0\right\}$: the probability of winning $\mathbb{P}(y_{i,T} = \mathcal{T}) = \mathbb{P}(y_{i,T} > \mathcal{T}) = 1$, i.e. the agent will be at or above the target with probability one, regardless of whether $e_{i,T} = \mathcal{T}$ or $e_{i,T} > \mathcal{T}$. As effort is costly, choosing $e_{i,T} = \mathcal{T}$ strictly dominates $e_{i,T} > \mathcal{T}$. Hence, we can rule out the case $\left\{e_{i,T} > \mathcal{T}, \sigma^2_{i,T} = 0\right\}$.

(2) $\{e_{i,T} \geq \mathcal{T}, \sigma^2_{i,T} = 0\}$: the probability of winning $\mathbb{P}(y_{i,T} \geq \mathcal{T}) = 1$. This strictly dominates the alternative choice $\{e_{i,T} \geq \mathcal{T}, \sigma^2_{i,T} \in (0, \infty) \cup \infty\}$, as it only gives $\mathbb{P}\left(y_{i,T}(e_{i,T} \geq \mathcal{T}, \sigma^2_{i,T} > 0) \geq \mathcal{T}\right) < 1$. Essentially, if the agent will surely win, introducing variance to this will only decrease the probability of winning as the outcome may also fall below threshold if we have a negative realisation of $\varepsilon_{i,T}$. Hence, we can exclude the four cases where $\{e_{i,T} \geq \mathcal{T}, \sigma^2_{i,T} > 0\}$.

(3) $\{e_{i,T} \in [0, \mathcal{T}), \sigma^2_{i,T} \in [0, \infty)\}$: this gives $\mathbb{P}(y_{i,T} \geq \mathcal{T}) < 0.5$, as there is a strictly positive probability that $\varepsilon_{i,T} < \mathcal{T} - e_{i,T}$ and in turn that the prize is not won, given a positive realisation of $\varepsilon_{i,T}$. This choice is strictly dominated by the alternative choice $\{e_{i,T} = 0, \sigma^2_{i,T} = \infty\}$, which gives the preferable $\mathbb{P}(y_{i,T} \geq \mathcal{T}) = 0.5$ at zero cost of effort. This rules out the four cases when $\{e_{i,T} \in [0, \mathcal{T}), \sigma^2_{i,T} \in [0, \infty)\}$.

(4) $\{e_{i,T} \in (0, \mathcal{T}), \sigma^2_{i,T} = \infty\}$: this gives $\mathbb{P}(y_{i,T} \geq \mathcal{T}) = 0.5$. The same probability can also be achieved by choosing $\{e_{i,T} = 0, \sigma^2_{i,T} = \infty\}$ at no cost of effort. Hence, the latter option dominates the former, so we can exclude it.

Hence, the two remaining cases ($\{e_{i,T} = \mathcal{T}, \sigma^2_{i,T} = 0\}$ and $\{e_{i,T} = 0, \sigma^2_{i,T} = \infty\}$) are considered in Section 3.2 to identify the optimal combination of effort and risk-taking.

C Additional results and robustness checks

The following sections provide additional empirical strategies, results and robustness checks for the main and the channel analyses. Throughout the sections we use the following abbreviations for our treatments: NT for *Neutral threshold*, DC for *Direct competition*, and CT for *Competitive threshold*.

C.1 Descriptive statistics

This section provides additional descriptive statistics for control variables across treatments, for distributions of effort and risk-taking across treatments, and for attrition.

Table C.1.1 provides summary statistics for control variables, divided by treatment groups, as well as ANOVA statistics for differences between treatment groups.

Table C.1.1: Summary statistics: Control variables

	NT	DC	CT	ANOVA
Age	38.557	35.565	39.086	3.568*
	(11.922)	(10.710)	(12.750)	(0.029)
Female	0.507	0.493	0.460	0.317
	(0.502)	(0.502)	(0.500)	(0.728)
USA	0.771	0.681	0.705	1.510
	(0.421)	(0.468)	(0.458)	(0.222)
India	0.157	0.268	0.216	2.566
	(0.365)	(0.445)	(0.413)	(0.078)
Overplacement	0.272	1.594	1.612	0.507
	(11.329)	(13.466)	(13.352)	(0.603)
Overestimation	0.464	0.399	0.482	1.080
of winning	(0.501)	(0.491)	(0.501)	(0.341)
General	5.571	5.884	5.950	0.712
risk preferences	(2.926)	(2.777)	(2.783)	(0.491)
Risk-aversion	0.393	0.341	0.309	1.093
	(0.490)	(0.476)	(0.464)	(0.336)
Observations	140	138	139	417

In column (1)-(3) mean coefficients and standard deviations in parentheses.

In column (4) F-statistic and p-values from ANOVA analysis in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.1.1 Distributions

Figure C.1 displays the cumulative frequency of effort and risk-taking choices for our three treatment groups.

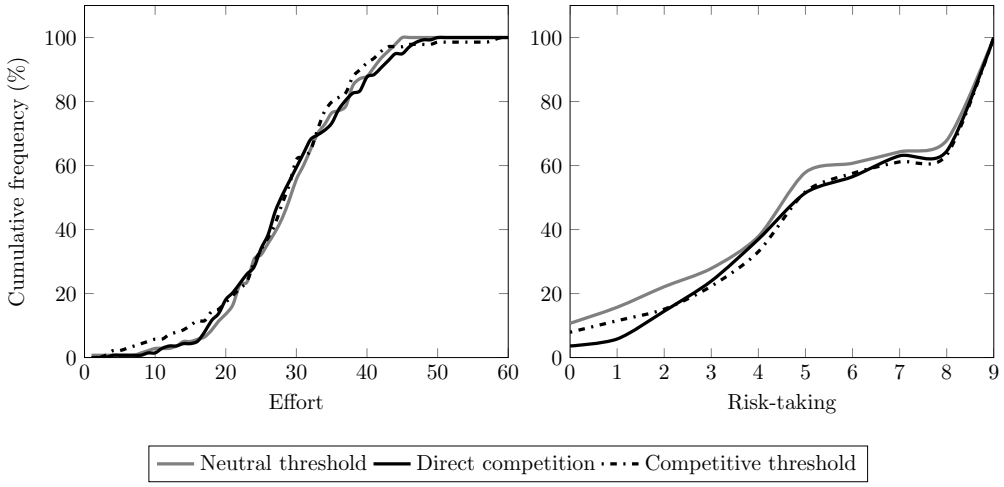


Figure C.1: Cumulative frequencies

The equality of distributions is tested between each pair of treatment using a Kolmogrov-Smirnov test. Table C.1.2 below gives the relevant p-values for the null hypotheses of equal distributions. As can be observed, all values are large, suggesting there is no relevant difference in distributions between any of our three treatments.

Table C.1.2: Two-sample Kolmogorov-Smirnov p-values

	NT vs. DC	NT vs. CT	DC vs. CT
Effort	0.939	0.837	0.895
Risk-taking	0.446	0.850	0.969

NT is *Neutral threshold*, DC is *Direct competition*, CT is *Competitive threshold*

C.1.2 Attrition

Table C.1.3 below gives the contigency table from a Fisher's exact test for selective attrition for the core sample. The p-value is sufficiently large to not reject the null hypotheses of no selective attrition.

Table C.1.3: Contingency table for attrition

	NT	DC	CT	Total
Dropped out	3	5	2	10
Final sample	140	138	139	417
Total	143	143	141	427

Fisher's exact test: $p = 0.620$

Table C.1.4 below gives the contigency table from a Fisher's exact test for selective attrition for the sample also including individuals who dropped out but also wouldn't

have passed the attention checks and thus would not have been in the main subject pool. The p-value is sufficiently large to not reject the null hypotheses of no selective attrition at relevant significance levels (5%).

Table C.1.4: Contingency table for attrition: Including rejected responses

	NT	DC	CT	Total
Dropped out	4	12	5	21
Final sample	140	138	139	417
Total	144	150	144	438

Fisher's exact test: $p = 0.092$

C.2 Probit model

As outlined in Section 6.2 for the second robustness test, we use a probit model to test our estimation strategy's robustness for our research question. In particular, one could imagine that while the OLS regressions give a small, significant effect of competition on risk-taking, yet that this effect is aggregated to lower levels of risk-taking. Interpreting this, competition would have an effect on risk-taking, yet perhaps with lesser economic importance as the propensity to take *high* risk may not have increased. Hence, we first discuss our specific empirical strategy followed by the results from our probit regressions.

C.2.1 Empirical strategy

As such, we estimate the probability of a subject exerting high effort or taking high risk when in competition. This has the added benefit of increasing power, by splitting our dependent variables in to high and low effort and risk-taking, respectively. To achieve this, risk-taking is split down the middle ($r_i^L \in [0, 4]$ and $r_i^H \in [5, 9]$), indicating more and less risk-averse bets, and effort is split in accordance with the mean in previous studies (e.g. Gill and Prowse (2012); Buser and Dreber (2016) , where $\mu_e \approx 26$, thus $e_i^L \in [0, 25]$ and $e_i^H \in [26, 60]$).

The probit model is preferable to a Linear Probabilities Model (LPM), as it does not risk predicting probabilities outside the closed interval $[0, 1]$, even though the results from the LPM can be easier to interpret. As we are more interested in the direction than the magnitude of the coefficient, especially in this robustness check, we do not see this as a large drawback. As such, we also follow related literature (see e.g. Eriksen and Kvaløy, 2017). Moreover, we note that that probit and linear probablity models often generate identical results.

As such, we specify the following:

$$\mathrm{P}(r_i^H = 1|\cdot) = \alpha + \beta_2 \times \mathrm{DC}_i + \beta_3 \times \mathrm{CT}_i + \gamma_1 K_{i1} + ... + \gamma_k K_{ik} + \epsilon_i$$

Once again, we run separate analyses for risk-taking (r_i^H) and effort (e_i^H), as in the OLS regressions, include the same three control specifications and perform the same collection of test for our statistical hypotheses (**H1-H6**), yet with different interpretations of the coefficients compared to before.[35]

[35] Throughout the results we provide comparisons with Table 7.3 in Section 7.2.

C.2.2 Results

Examining the probit regressions, Table C.2.1 below shows that while treatment coefficients in the OLS regressions (Table 7.3) remained positive throughout, the *Competitive threshold* coefficient becomes negative in fuller probit models, i.e. columns (2) and (3), contrary to predictions. Similarly, unlike in the OLS regressions, also the *Direct competition* coefficient turns positive in fuller models, (5) and (6). However, across all regressions both treatment coefficients are insignificantly different from zero.

Table C.2.1: Probit regressions

	(1) r_i^H	(2) r_i^H	(3) r_i^H	(4) e_i^H	(5) e_i^H	(6) e_i^H
DC	0.128	0.075	0.051	-0.046	0.137	0.157
	(0.154)	(0.161)	(0.161)	(0.156)	(0.180)	(0.181)
CT	0.024	-0.013	-0.010	-0.053	0.122	0.155
	(0.153)	(0.160)	(0.162)	(0.156)	(0.189)	(0.188)
Overestimation		0.221	0.198		0.056	0.082
of winning		(0.137)	(0.136)		(0.152)	(0.156)
Overplacement		0.002	0.001		-0.094***	-0.095***
		(0.006)	(0.006)		(0.012)	(0.012)
General		0.147***	0.136***		0.025	0.025
risk preferences		(0.025)	(0.027)		(0.029)	(0.031)
Age			0.010			-0.009
			(0.006)			(0.007)
Female			-0.151			-0.229
			(0.136)			(0.159)
USA			0.079			-0.317
			(0.265)			(0.350)
India			0.352			-0.679
			(0.296)			(0.389)
Constant	0.309**	-0.582***	-0.945*	0.464***	0.449*	1.276**
	(0.108)	(0.176)	(0.384)	(0.110)	(0.201)	(0.471)
Observations	417	417	417	417	417	417
Pseudo R^2	0.001	0.093	0.104	0.000	0.326	0.340

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As such, we can draw a first conclusion: our results for the impact of competitive framing are robust to our choice of estimation strategy. Neither here can we reject the null hypotheses of any significant effect for the coefficient of the *Competitive threshold* treatment on both higher risk-taking (**H3**) and higher effort (**H4**).

For the remaining hypotheses, Table C.2 shows the χ^2-statistics and p-values. Similarly as for the OLS regressions, all p-values are much too large to reject either hypotheses, for any regression specifications. As such, our results for the impact of competitive incentives or the join effect is robust to our estimation strategy.

As a final detail, in Table C.2.1 fewer of the control variables are significant than in the OLS regressions. Only higher general risk-preferences show indicate a greater propensity to take higher risks, and more overplacing individuals have a lower propensity to exert

high effort. This difference may be explained by that individual characteristics may perhaps influence the on average, but not the propensity to exert higher effort.

Table C.2.2: χ^2-tests

	r_i^H		e_i^H	
	H1	**H5**	**H2**	**H6**
Model 1, 4	0.453	0.775	0.002	0.236
	(0.501)	(0.679)	(0.966)	(0.934)
Model 2, 5	0.291	0.340	0.006	0.688
	(0.590)	(0.844)	(0.940)	(0.709)
Model 3, 6	0.137	0.160	0.000	0.976
	(0.712)	(0.923)	(0.995)	(0.614)

Model numbers refers to regressions in Table C.2.1

p-values in parentheses.

C.3 Robustness checks

As outlined in the Section 6.2, in addition to inclusion of control variables and the probit model, we perform another two kinds of robustness checks: regressions with normal standard errors and variations of control variable specification. This section outlines the results and process of these checks.

To check the robustness of the results to the choice of standard errors, Table C.3.1 below gives main OLS regressions for risk-taking and effort, but with normal instead of robust standard errors.

Table C.3.1: OLS regressions: Normal standard errors

	(1)	(2)	(3)	(4)	(5)	(6)
	Risk-taking	Risk-taking	Risk-taking	Effort	Effort	Effort
DC	0.413	0.262	0.235	-0.925	-0.264	-0.102
	(0.360)	(0.337)	(0.338)	(1.095)	(0.765)	(0.738)
CT	0.447	0.356	0.368	-0.125	0.569	0.620
	(0.361)	(0.338)	(0.342)	(1.097)	(0.768)	(0.746)
Overestimation		0.439	0.422		0.435	0.674
of winning		(0.287)	(0.288)		(0.651)	(0.629)
Overplacement		0.009	0.008		-0.522***	-0.520***
		(0.011)	(0.011)		(0.025)	(0.025)
General		0.348***	0.333***		0.081	0.089
risk preferences		(0.051)	(0.058)		(0.117)	(0.126)
Age			0.014			-0.111***
			(0.012)			(0.027)
Female			-0.325			-1.891**
			(0.290)			(0.633)
USA			0.247			-0.507
			(0.569)			(1.242)
India			0.465			-4.014**
			(0.623)			(1.361)
Constant	5.350***	3.205***	2.661**	29.364***	28.852***	34.944***
	(0.254)	(0.374)	(0.815)	(0.773)	(0.849)	(1.778)
Observations	417	417	417	417	417	417
R^2	0.005	0.136	0.142	0.002	0.518	0.557

Normal standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As can be seen, as treatment coefficients remain insignificant and with very large standard errors, our results are robust to the use of robust standard errors.

For the final level of robustness checks, Table C.3.2 provides variable definitions for control variables used in alternative specifications to test robustness of our estimates.

Table C.3.2: Robustness checks: control variables

	Range	Description
Overestimation of performance	-60 to 60	$E(e_i) - e_i$, i.e. the difference between predicted and realized effort
Risk-aversion	0, 1	1 if *General risk preferences* < 1
Risk-aversion in entrepreneurship	0, 1	1 if subject has not started own company
Risk-aversion in gambling	0, 1	1 if subject has not partaken in gambling activities in past month

Examining risk-taking first and effort subsequently, Table C.3.3 shows the robustness checks for risk-taking. In column (1) to (3) the overconfidence measures are adjusted. In column (4) to (6) the risk preference measures are adjusted.

Table C.3.3: OLS regressions: Robustness checks for risk-taking

	Overconfidence			Risk preferences		
	(1)	(2)	(3)	(4)	(5)	(6)
DC	0.237	0.238	0.236	0.212	0.277	0.303
	(0.350)	(0.347)	(0.349)	(0.354)	(0.360)	(0.356)
CT	0.335	0.374	0.337	0.353	0.344	0.405
	(0.340)	(0.339)	(0.340)	(0.346)	(0.355)	(0.351)
Overestimation		0.439		0.529	0.737*	0.707*
of winning		(0.291)		(0.293)	(0.295)	(0.291)
Overplacement	0.012			0.012	0.018	0.014
	(0.015)			(0.012)	(0.012)	(0.012)
Overestimation	-0.003	0.002	0.002			
of performance	(0.009)	(0.008)	(0.008)			
General	0.350***	0.339***	0.355***			
risk preferences	(0.060)	(0.061)	(0.059)			
Risk-aversion				-1.426***		
				(0.353)		
Company					0.429	
					(0.311)	
Gambling						1.076**
						(0.335)
Age	0.015	0.014	0.014	0.010	0.002	0.011
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Female	-0.310	-0.333	-0.329	-0.403	-0.645*	-0.559
	(0.291)	(0.293)	(0.294)	(0.298)	(0.298)	(0.296)
USA	0.233	0.250	0.257	0.054	0.172	0.184
	(0.475)	(0.479)	(0.472)	(0.487)	(0.530)	(0.523)
India	0.492	0.470	0.515	0.601	1.000	0.974
	(0.520)	(0.524)	(0.517)	(0.535)	(0.566)	(0.550)
Constant	2.713***	2.628***	2.703***	5.349***	4.837***	4.320***
	(0.779)	(0.778)	(0.777)	(0.705)	(0.729)	(0.751)
Observations	417	417	417	417	417	417
R^2	0.138	0.141	0.136	0.112	0.076	0.093

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.3.4 shows the robustness checks for effort. In column (1) to (3) the overconfidence measures are adjusted. In column (4) to (6) the risk-preference measures are adjusted.

Table C.3.4: OLS regressions: Robustness checks for effort

	Overconfidence			Risk preferences		
	(1)	(2)	(3)	(4)	(5)	(6)
DC	-0.080	-0.009	-0.007	-0.114	-0.084	-0.101
	(0.747)	(1.007)	(1.005)	(0.747)	(0.742)	(0.742)
CT	0.581	0.463	0.486	0.615	0.643	0.598
	(0.740)	(0.939)	(0.936)	(0.742)	(0.745)	(0.738)
Overestimation		-0.278		0.689	0.719	0.750
of winning		(0.782)		(0.603)	(0.613)	(0.609)
Overplacement	-0.505***			-0.519***	-0.514***	-0.513***
	(0.041)			(0.033)	(0.033)	(0.033)
Overestimation	-0.015	-0.224***	-0.224***			
of performance	(0.019)	(0.022)	(0.022)			
General	0.122	-0.093	-0.103			
risk preferences	(0.127)	(0.175)	(0.177)			
Risk-aversion				-0.493		
				(0.687)		
Company					-0.431	
					(0.647)	
Gambling						-0.630
						(0.737)
Age	-0.106***	-0.075*	-0.075*	-0.111***	-0.109***	-0.116***
	(0.027)	(0.036)	(0.036)	(0.027)	(0.027)	(0.027)
Female	-1.831**	-1.009	-1.012	-1.891**	-2.012**	-2.051***
	(0.634)	(0.859)	(0.858)	(0.622)	(0.614)	(0.610)
USA	-0.598	-1.606	-1.610	-0.563	-0.606	-0.587
	(1.280)	(1.809)	(1.806)	(1.272)	(1.269)	(1.246)
India	-4.029**	-4.980*	-5.009**	-4.014**	-3.784**	-3.797**
	(1.426)	(1.935)	(1.925)	(1.404)	(1.349)	(1.351)
Constant	34.934***	35.400***	35.352***	35.702***	35.566***	35.856***
	(1.721)	(2.485)	(2.469)	(1.613)	(1.575)	(1.580)
Observations	417	417	417	417	417	417
R^2	0.557	0.251	0.251	0.557	0.557	0.558

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.4 Channel analysis

This section includes additional variable definitions, descriptive statistics, as well as robustness checks for our channel analyses.

Table C.4.1 below includes key variables for the interactions in the channel analyses; the treatment dummies as well as the three channel dummies.

Table C.4.1: Channel variable definitions

	Range	Description
DC	0 to 1	1 if subject was in the *Direct competition* treatment
CT	0 to 1	1 if subject was in the *Competitive threshold* treatment
Female	0, 1	1 if *Gender* = female
Overestimation	0, 1	1 if subject believed she would win and did not,
of winning		0 otherwise
Risk-aversion	0, 1	1 if *General risk-preferences* < 5

C.4.1 Descriptive statistics

Table C.4.2 gives summary statistics for the dependent variables across groups used for channel analysis: gender, overestimation of winning, and risk-aversion.

Table C.4.2: Summary statistics: Channel variables

	NT		DC		CT		Total	
Female	0	1	0	1	0	1	0	1
Effort	29.594	29.141	29.371	29.103	28.933	27.859	29.290	28.724
	(8.992)	(8.168)	(9.143)	(9.064)	(10.346)	(9.101)	(9.495)	(8.749)
Risk-taking	5.768	4.944	6.543	5.029	5.907	5.594	6.070	5.177
	(3.308)	(2.971)	(2.558)	(2.937)	(2.796)	(3.245)	(2.905)	(3.047)
Overconfidence	0	1	0	1	0	1	0	1
Effort	30.747	27.769	31.229	26.236	26.806	30.194	29.687	28.187
	(8.726)	(8.133)	(8.816)	(8.690)	(8.686)	(10.605)	(8.926)	(9.340)
Risk-taking	4.720	6.077	5.651	6.018	5.250	6.313	5.222	6.144
	(3.003)	(3.198)	(2.903)	(2.765)	(3.339)	(2.506)	(3.088)	(2.826)
Risk-aversion	0	1	0	1	0	1	0	1
Effort	27.659	32.000	28.033	31.574	28.729	27.791	28.162	30.614
	(8.758)	(7.582)	(9.061)	(8.720)	(10.239)	(8.719)	(9.380)	(8.452)
Risk-taking	6.118	4.164	6.374	4.681	6.344	4.465	6.283	4.421
	(2.851)	(3.265)	(2.533)	(3.101)	(2.679)	(3.305)	(2.679)	(3.210)

Value 0 indicates that the group to which the characteristics does not apply, 1 indicates the groupp to which it does.

Mean coefficients, standard deviations in parenthesis.

C.4.2 Robustness checks

Table C.4.3 below gives the core robustness check for the channel analysis, i.e. probit regressions for risk-taking and effort for the three channels. The same empirical strategy as outlined in the main analysis is used and the same control variables are included in each regressions as in the OLS regressions.

Table C.4.3: Robustness check: Probit regressions for channels

	Gender differences		Overconfidence		Risk-aversion	
	(1)	(2)	(3)	(4)	(5)	(6)
	r_i^H	e_i^H	r_i^H	e_i^H	r_i^H	e_i^H
DC	0.028	0.296	0.167	-0.483*	0.117	0.399
	(0.227)	(0.254)	(0.218)	(0.216)	(0.204)	(0.227)
CT	0.164	0.437	0.172	-0.083	0.035	0.260
	(0.238)	(0.272)	(0.210)	(0.214)	(0.207)	(0.237)
Overestimation	0.200	0.094	0.427	-0.416	0.222	0.051
of winning	(0.137)	(0.157)	(0.231)	(0.226)	(0.136)	(0.159)
Overplacement	0.001	-0.096***			0.003	-0.095***
	(0.006)	(0.012)			(0.006)	(0.012)
General	0.138***	0.024	0.142***	-0.054*		
risk preferences	(0.027)	(0.031)	(0.027)	(0.027)		
Risk-aversion					-0.507*	0.227
					(0.233)	(0.272)
Female	-0.059	0.034	-0.135	-0.215	-0.178	-0.217
	(0.228)	(0.254)	(0.138)	(0.138)	(0.135)	(0.161)
Female x DC	0.061	-0.275				
	(0.322)	(0.366)				
Female x	-0.331	-0.542				
CT	(0.326)	(0.373)				
Overconfidence x			-0.260	0.969**		
DC			(0.331)	(0.321)		
Overconfidence x			-0.439	0.085		
CT			(0.328)	(0.321)		
Risk-aversion x					-0.194	-0.743
DC					(0.334)	(0.386)
Risk-aversion x					-0.111	-0.328
CT					(0.327)	(0.409)
Constant	-0.972*	1.184*	-1.105**	1.615***	0.112	1.364**
	(0.400)	(0.492)	(0.393)	(0.402)	(0.340)	(0.451)
Observations	417	417	417	417	417	417
Pseudo R^2	0.107	0.344	0.107	0.046	0.088	0.346

Additional controls include Age, USA, and India. *Overconfidence* in interactions is defined as *Overstimation of winning*. Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.4.4 shows the χ^2-statistics and p-values for the probit regressions in our channel analysis. Similarly as for the OLS regressions, all but one p-value are much too large to reject the hypotheses. The one p-value that is significant at the 1%-level is, as expected, for hypotheses **H11**.

Table C.4.4: χ^2-tests

	Gender differences		Overconfidence		Risk-aversion	
	r_i^H	e_i^H	r_i^H	e_i^H	r_i^H	e_i^H
	(**H8**)	(**H9**)	(**H10**)	(**H11**)	(**H12**)	(**H13**)
χ^2-statistic	1.427	0.469	0.295	7.59	0.059	1.032
	(0.232)	(0.494)	(0.587)	(0.006)	(0.808)	(0.310)

Column numbers refers to regressions in Table C.4.3

p-values in parentheses.

D Pre-analysis plan

The following section contains the full pre-analysis plan as compiled on April 7th at 9.45 am. It is uploaded to the Open Science Framework along with the data set and codes for replication of the results.[36] *As a pre-analysis plan, it is necessarily repetitive of what has been presented throughout the paper but was completed prior to data collection.*

Nota bene: notation and names for e.g. treatments differ in some places between the main study and the pre-analysis plan, but no changes have been made to the substance of these.

---

1 Introduction

The purpose of this project is to answer the research question:

Does competition between actors affect risk-taking and effort when the outcome of the competition is determined by both?

To answer this, we will employ an experimental methodology where we expand upon previous research by developing a design in which subjects choose preferred risk-level (through an investment choice based on Gneezy and Potters (1997)) and perform a real effort task (as outlined by Gill and Prowse (2012)). Subjects will be randomised into three treatments in order to distinguish between the effects of competition due to tournament incentives and due to social comparison. The experiment will be performed on Amazon Mechanical Turk, with experimental details are provided below.

Without expending too much time on background information the literature on competition, effort and risk-taking broadly identifies two channels. Firstly, competition can increase risk-taking and decrease effort through the inherent economic incentives in competitions. Hvide (2002) and Gilpatric (2009) both show that strategic incentives exist to substitute effort for risk, and hence minimise effort and maximise risk-taking in order to win the prize in a competition. Secondly, there is some evidence of a behavioural channel, where competition and social comparison, *per se*, trigger changes in effort and risk-taking. Ranking, social comparison, and a competitive frame have been shown to affect risk-taking and effort, even when there is no incentive for this change in behaviour (e.g. Kirchler et al., forthcoming; Eriksen and Kvaløy, 2017; Falk and Ichino, 2006). With our experiment, we seek to to test these two channels.

2 Hypotheses

2.1 Primary hypotheses

To generate our hypotheses, we turn to the results of a model proposed by Hvide (2002) and an extension by us which includes competition against a set threshold, as well as results from behavioural research. From the model's observations, we derive our first set of hypotheses:

1. *Direct competition leads to a higher level of risk-taking than threshold incentives:* $\sigma^2_{i,T} < \sigma^2_{i,C}$
2. *Direct competition leads to a lower level of effort than threshold incentives:* $e_{i,T} > e_{i,C}$

[36] Link: `https://osf.io/8wy5b/`

A model such as Hvide (2002) makes no difference between if the threshold is framed in a neutral way (threshold setting $T = N$) or in a way that is associated with social comparison and competition (threshold setting $T = S$). However, behavioral, cognitive and experimental research has shown individuals to react to competition per se, i.e. the absence of a threshold but inclusion of a social comparison framing. In other words, by introducing a competitive frame we derive two further hypotheses:

3. *A competitively framed threshold leads to a higher level of risk-taking than a neutrally framed threshold:* $\sigma^2_{i,N} < \sigma^2_{i,S}$
4. *A competitively framed threshold leads to a higher level of effort than a neutrally framed threshold:* $e_{i,N} < e_{i,S}$

Key here is that the competitive setting is included in both the competitive threshold and the direct competition, but not in the neutral threshold. We hypothesise that the effects of social comparison and competitive incentives are additive, and as such we formulate the following hypothesis for risk-taking:

5. *Direct competition leads to higher risk-taking than a competitively framed threshold:* $\sigma^2_{i,C} > \sigma^2_{i,S}$

However, for the hypothesis for effort the question becomes less straightforward and we hypothesise that for effort:

6. *Direct competition leads to lower effort than competitively framed threshold but higher effort than a neutral threshold*: $e_{i,N} < e_{i,C} < e_{i,S}$

2.2 Secondary hypotheses

While our primary hypotheses regard our core research question and are rooted in economic theory as well as previous research, the secondary hypotheses are more exploratory in nature, and are thus disclosed here for completeness. Our secondary hypotheses come in two forms:

- **Substitution between effort and risk-taking**: motivated by theoretical predictions Hvide (2002) and findings by Nieken (2010) and Andersson et al. (2017) we will analyze the possibility for risk-taking and effort to function as complements or substitutes.
- **Channels** for competition's effect on risk-taking and effort: evidence from a vast literature suggests that competition may have different impacts on risk-taking and effort for females, overconfident or risk-averse individuals.

Substitution between effort and risk-taking
Building on Andersson et al. (2017) subjects in our study can affect their performance measure through both exerting effort and taking risks. This is in line with the findings in Nieken (2010) and Hvide (2002). Hence, we hypothesise:

7. *Effort and risk-taking are substitutes to one another, i.e. subjects who exert high effort exert lower risks and vice versa.*

Gender differences
There exists a vast literature that explores gender differences in competitiveness, risk-taking, and factors that influence economic decisions. Many studies have found that men are more likely to enter competition than women (e.g. Niederle and Vesterlund, 2007; Buser and Yuan, 2016), but men's performance have been found to increase more as a result of competition than women's (see review by Niederle and Vesterlund, 2011). Furthermore, men have been found to be more risk-taking than women (Charness et al., 2013), and as a result, we hypothesise:

8. *Men increase risk-taking more than women as a result of competition*

9. *Men increase effort more than women as a result of competition*

Overconfidence

The literature on gender differences in risk-taking and competitiveness has found overconfidence to play a large role in the decision to enter competition and take risks (Croson and Gneezy, 2009; Niederle and Vesterlund, 2007, 2011), where overconfident people take more risks and are more likely to enter competition, even if they perform badly. With regards to effort, however, things are a bit more complicated. Either, overconfident agents overestimate their marginal product and exert higher effort (Gervais and Goldstein, 2007), or they exert low effort as they "believe they will achieve their goals anyways" (Mueller, 2007, p. 16). Given the theoretical groundings in Gervais and Goldstein (2007); Everett and Fairchild (2015), we believe that the former effect will dominate, but are open to the possibility of being wrong. As such, we hypothesise:

10. *When in competition, overconfident individuals increase risk-taking more than others*
11. *When in competition, overconfident individuals increase effort more than others*

Risk-aversion

The literature on gender and competition has found risk-attitudes play a role in the decision to enter competition (Niederle and Vesterlund, 2007, 2011), although not as large as for example overconfidence. Risk-averse people are also more likely to exert effort more effort in order to avoid bad outcomes (Jullien et al., 1999). As such, we test the following hypotheses:

12. *Risk-averse participants increase risk-taking less than others as a result of competition*
13. *Risk-averse participants increase effort less than others as a result of competition*

3 Experimental design

The experiment will be carried out using Amazon Mechanical Turk. There will be three treatments, one with a neutral threshold, one with a competitive threshold, and one with direct competition, as described below. Participants are randomised into the treatments, using the individual as the unit of randomisation. The experimental task itself is performed using Qualtrics.

The data collection will be carried out in April 2018, with the experiment being uploaded to Amazon Mechanical Turk on April 7th at approximately 3 pm CEST.[37] Collection of data will end once we have the predetermined number of subjects, i.e. 420. Criteria for exclusion of observations will be explained below, together with an analysis of power in the experiment.

3.1 Power analysis and sample size

Generally, statistical power is not discussed enough in the experimental literature, which has led to a number of problems concerning replication (Camerer et al., 2016; Open Science Collaboration, 2015). We want to be able to have 80% power in our experiment, and use previous research as a guide with regards to effect sizes. As three of the closest studies to ours are Andersson et al. (2017), Eriksen and Kvaløy (2017), and Filippin and Gioia (2017), we use these papers to estimate the sample size needed in the experiment.

[37] The careful reader may notice that the final edit of this pre-analysis plan also was posted on April 7th, however prior to the initiation of data collection.

Doing so, estimate a needed sample size of around 130 subjects in each treatment based on the effect size in Eriksen and Kvaløy (2017),[38] 70 based on the effect size in Filippin and Gioia (2017),[39] and 65 based on the effect size in Andersson et al. (2017).[40] As a result, we settle on using 140 subjects in each treatment, in order to leave some margin. Power calculations were made using the power calculator provided by HyLown Consulting (2018).

3.2 Attrition and exclusion of observations

In any experiment, especially online, selective attrition can be a problem. To avoid attrition, we make it clear that no payment will be made to participants unless they finish the whole experiment and submit their unqiue identifier to MTurk. To enforce this, we only provide the completion code necessary to submit the HIT at the end of the survey.

Following Buser and Dreber (2016), we will also use Fisher's exact test to check for selective attrition in the cases where subjects themselves have opted to end the experiment. Such individuals are noted as "responses in progress" in our survey. However, a potentially confounding factor here is that the total number of submitted HITs may have reached its limit while the subject is taking the survey – prompting her to stop, even though she would have wanted to finish. As such, subjects which have completed all questions in the survey, yet do not appear in our list of submitted HITs will be excluded from both the main data set and the attrition as they did not attrite, but also cannot be paid.

Regarding screening of subjects and exclusion of observation, we have three main exclusion criteria. Firstly, we use so called "Attention Questions" to exclude inattentive subjects. When using an online workforce, we make sure the participants are paying attention (for reference, see e.g. Straub, 2017) by asking them three simple questions to which the correct answer is found in the adjacent text. If a participant fails two or three of these attention checks, they are excluded from the sample and not paid for their participation – something they are also informed of when accepting the HIT. Moreover, we also see these Attention Checks as a screen for the participant's ability to understand English.

Secondly, we exclude participants based on multiple participation. As our survey records IP adresses, we identify and exclude all but the first submission from subjects with multiple entries. Also here participants are notified that multiple submissions will not be accepted and paid.

Finally, in our analyses of the gender channel we drop observations who do not identify as gender binary. Hence, we drop subjects who enter gender "Other", however these observations will be used in the main analyses (as we there simply control for gender "Female" to avoid problems with multicollinearity as the number of subjects who identify neither as "Female" or "Male" is likely to be small). As such, participants who identify as "Other" are naturally paid.

3.3 Primary outcome variables

In the experiment, participants make choices over risk and real effort levels in a task where the outcome is dependent on both. By performing a real effort slider task (designed by Gill and Prowse, 2012), participants collect points from each correctly placed

[38] Based on treatments WInfo4 and NoInfo4.

[39] Based on correct coins in Table 1. We do not look at their risk-measure as they look at spill-overs to risk, we do not.

[40] Based on treatments WTA and SFAS.

slider. From these points they can choose to invest as share in a lottery with a 50% chance of a positive outcome (as designed by Gneezy and Potters, 1997). In order to win a bonus payment from the task, subjects must reach a "high enough" total points. The number of points a subjects must reach, and how it is presented, will change between treatments.

We have two dependent variables: *effort* and *risk-level*. For the former, we interpret a subject's number of correctly placed sliders, within the Slider Task designed by Gill and Prowse (2012), as a subject's exerted effort. For the latter, we interpret the number of points from each slider as subject invests in a lottery, designed based on Gneezy and Potters (1997)'s risk preference elicitation, as a subject's choice of risk-level.

3.4 Treatments

Our main treatment variable is the language we use to describe how many points the subjects need to win the bonus payment as well as the condition for paying the bonus. More concretely, we randomise subjects into one of three treatments in which we vary the nature of the required total points to win the bonus payment: a "neutral" threshold (1), a "competitive" threshold (2) and direct competition (3) as follows:

1. The target number of total points to beat is the median number of points gained by subjects in our pilot study. The choice of the median is to give a high probability for participants to win the bonus. However, to create the "neutral" setting, Group 1 is simply informed that they need to beat a threshold value (T). Hence, a participant in Group 1 wins the bonus if $y_i > T$.
2. The target number of total points to beat is the median number of points from the pilot study (i.e. same as in Group 1). However, to create the "competitive" threshold, a subject i in Group 2 is informed that the threshold value (y_j) is the score of a previous participant j, and they need to beat this previous participant. As in Group 1, a participant in Group 2 wins the bonus if $y_i > y_j$.
3. Group 3, in contrast to the previous two, is not provided with an explicit threshold, but is informed that they are in direct competition with another, unnamed participant. Concretely, in difference to Group 2, they are not informed of y_j, but for a participant in Group 3 to win $y_i \geq y_j$ is required. Hence, there is complete uncertainty over the competitor's effort and risk-choices. The subjects are also rivals in competing for the good, i.e. a "true" competition.

Throughout the treatments identical instructions are used, apart from the few lines of key treatment information. Retaining all other factors, e.g. task, order of steps, timing, questions, equal across treatments – the nature of the target is the sole treatment variable.

3.5 Pilot study

A pilot study was carried out 31 March - 2 April 2018 in order to get a threshold value for the main experiment, as well as to test the design. This was done with 30 participants who all received the direct competition treatment. These participants will not be included with the data to test our hypotheses.

Some small changes were made to the survey design, such as forcing subjects to stay two minutes on the slider task, increase the number of sliders made, and well some minor changes to clarify the text. We also made it more difficult to take the survey more than once, and stated that multiple answers would be rejected.

4 Empirical strategy

In answering our research question, we use a three-step empirical strategy: descriptive analysis, regression analysis for primary and for secondary hypotheses. Here, we will specify the latter two as the descriptive analysis will mainly consist of description of data with the use of averages, standard deviations and histograms.

4.1 Regression analysis

4.1.1 Primary hypotheses

Firstly, the separate impact of competition on effort and risk-taking will be analyzed using Ordinary Least Squared (OLS) regressions of the form:

$$y_i = \alpha + \beta_2 G2_i + \beta_3 G3_i + \gamma_1 K_{i1} + ... + \gamma_k K_{ik} + \epsilon_i$$

where y_i is the dependent variable (effort e_i or risk-taking x_i), $G2_i$ and $G3_i$ are treatment dummies (with treatment Group 1, G1, used as baseline), and $K_{i1}, ..., K_{ik}$ a list of k potential control variables. We will use three versions of the specification:

- $k = 0$, i.e. a pure model with only treatment dummies and a constant.
- $k = 3 = \{$overestimation of winning, overplacement, general risk-preference$\}$, i.e. an expanded model with key personality controls
- $k = 6 = \{$overestimation of winning, overplacement, general risk-preferences, gender, age, country of residence$\}$, i.e. a full model including personality as well as personal characteristic controls

The coefficient of interest to us is the treatment effects across the groups, where we seek to test if $\beta_2 = 0$, $\beta_3 = 0$ and $\beta_2 = \beta_3$. These hypotheses will be tested using standard, two-sided t-tests.

A final analysis to increase power we will run a Probit model where we split the sample into two parts and seek the probability of an individual taking higher risks or higher effort as competitive frames are added. To achieve this, risk-taking is split down the middle ($x_i^L \in [0, 4]$ and $x_i^H \in [5, 9]$), and effort is split in accordance with the mean in previous studies (e.g. Gill and Prowse (2012); Buser and Dreber (2016), where $\mu_e \approx 26$, thus $e_i^L \in [0, 25]$ and $e_i^H \in [26, 49]$). We run two separate analyses, similar to the OLS regressions above, and similarly include the same extensions using controls.

4.1.2 Secondary hypotheses

Substitution between effort and risk-taking
Secondly, while competition may have an effect on effort and risk in general, it is possible that risk-taking and effort functions as complements or substitutes. Therefore, we analyze the joint effect using OLS regressions of the form:

$$e_i = \alpha + \beta_1 x_i + \beta_2 G_i + \beta_3 G3_i + \gamma_1 K_{i1} + ... + \gamma_k K_{ik} + \epsilon_i$$

where e_i is the dependent variable effort, x_i is the risk-taking level, $G2_i$ and $G3_i$ are treatment dummies, and $K_{i1}, ..., K_{ik}$ a list of k potential control variables, as described above. Similarly to the primary analyses, we run the regression with three control specifications.

Here, the direction and significance of the coefficient on risk in the regression for effort is of interest. As we seek to test the whether the dependent variable are substitutes we test whether $\beta_1 = 0$. If effort and risk are substitutes, we would expect $\beta_1 < 0$, if they are compliments $\beta_1 > 0$. Additionally, analysis of the coefficients on the treatments

may be interesting from an exploratory perspective but are not contained within our hypotheses.

Channels
Beyond our main hypotheses, we also seek to explore potential channels through which competition may affect effort and risk-taking. We will explore the effect of overconfidence, risk-aversion, and gender och the effect of competition on risk-taking and effort. This will be done using an interaction term with the OLS regressions with the following specification:

$$y_i = \alpha + \beta_2 G2_i + \beta_3 G3_i + \delta_2 G2_i I_i + \delta_3 G3_i I_i + \gamma_1 K_{i1} + ... + \gamma_1 k_{ik} + \epsilon_i$$

where $G2_i$ and $G3_i$ are treatment dummies, as in the primary hypotheses analyses. In contrast to before, the model now includes interactions between the treatment variables and the variable I which depends on the channel of interest. I is defined as follows:

1. **Gender**: I is a dummy defined as one if the subject defines as "Female", and zero otherwise.
2. **Risk-aversion**: I is a dummy defined as one if the subject is sees herself as risk-averse (a rating <5 on the general risk-preference scale), and zero otherwise
3. **Overconfidence**: In the primary OLS regressions above we control for both overestimation of winning and overplacement. In order to increase power for our channel analysis, we simply interact one of these. Here, we use a measure of overconfidence that takes both risk and effort into account (overestimation of winning). Hence, we use a dummy that is equal to one if the subject thought she would win and did not, and zero otherwise.

Moreover, the specification includes control variables $K_{i1}, ..., K_{ik}$, including overestimation of winning, gender, age, and country of residence (as in the primary analyses). Additionally, for the channels of gender and overconfidence K includes general risk-preferences (as in the primary analyses), but for the channel of risk-aversion it includes a risk-aversion control. For the channels of gender and risk-aversion K includes overplacement (as in the primary analyses). We do not include the control for overplacement for the channel of overconfidence to limit the channel analysis to one overconfidence measure.

Now, the coefficient of interest to us is the joint effect of treatment and the channel, hence we seek to test if $\delta_2 = 0$, $\delta_3 =$ and $\delta_2 = \delta_3$. Similarly here, we will perform a two-sided t test.

4.2 Robustness

By including control variables in our three regression specifications we test the robustness of our results to different model specifications. Additionally, to analyze the robustness of our results to our specific choices of variables we collect additional measures of overconfidence and risk-preferences. Switching our chosen variables with the alternatives and performing the same analysis we compare the outcomes to see if the hypotheses hold. We perform these robustness checks for our primary hypotheses, as they cannot be done for our secondary hypotheses.

Risk-taking: Alternative specifications using *Risk-aversion in entrepreneurship*, *Risk-aversion in gambling* as well as *Risk-aversion*.

Overconfidence: Alternative specifications using *Overestimation of performance*.

4.3 Variable definitions

Through our experiment we are able to collect directly collect data on the chosen risk level ($x_i \in [0, 9]$), on general risk-preferences from the Likert-scale (*risk preference* $\in [0, 10]$) and on age (*age* $\in [0, \infty)$). The remainder of the variables require (minor) transformations:

- **Effort**: Qualtrics collects data on each slider, we measure total number of correct slider by comparing the value allocated with the correct, tabulating the total number of correctly placed sliders
- **Overestimation of performance (robustness)**: As seen in other literature (for reference, see e.g. Filippin and Gioia, 2017), we measure over(under)estimation of performance as difference between predicted and executed effort: $\mathrm{E}(e_i) - e_i$
- **Overplacement**: By adapting Moore and Healy (2008), we calculate $(\mathrm{E}(e_i) - \mathrm{E}(\bar{e})) - (e_i - \bar{e})$, i.e. difference between predicted distance to mean effort and the actual distance to mean effort. The mean here is taken from the subject's treatment group and not the mean of the entire subject pool. In contrast to Moore and Healy (2008), we use the mean instead of the effort of another participant for measuring overplacement in order to keep the measurement constant between treatments.
- **Overestimation of winning**: A dummy variable which is the 1 if the individual thinks it will surpass the target, but does not, and 0 otherwise.
- **General risk-averse preferences**: Include a dummy variable to indicate general risk-preferences below 5 on the Likert-scale as is common (Ding et al., 2010; Treibich, 2015).
- **Risk-aversion robustness measures**: We collect information on individuals experience with entrepreneurship (if they have started an own company) and on gambling habits (in the past month, at e.g. a casino or online) which we transform from yes-no answers to dummy-variables with 0 for yes, 1 for no as not partaking in either activity indicates relative risk-aversion.
- **Winning**: After having performed the draw of lottery outcomes and tallied total points, we note winning as a dummy variable.
- **Country of residence**: Subjects provide information on actual country, but group these according to the two largest MTurk origins (Difallah et al., 2018): USA, India, other using dummy variables for USA and India.
- **Gender**: Dummy variable which is equal to one if the subject is female, and zero otherwise.
- **High and low risk-taking**: Where we choose to use a probit model and/or a grouped risk-taking variable we split the sample down the middle ($x_i^L \in [0, 4]$ and $x_i^H \in [5, 9]$).
- **High and low effort**: Where we choose to use a probit model and/or a grouped effort variable we split the sample in accordance with the mean in previous studies (e.g. Gill and Prowse (2012); Buser and Dreber (2016), where $\mu_e \approx 26$, thus $e_i^L \in [0, 25]$ and $e_i^H \in [26, 60]$)