Gender Robustness of Overconfidence and Excess Entry

Danková, Katarína and Servátka, Maroš

MGSM Experimental Economics Laboratory, Macquarie Graduate School of Management, University of Economics in Bratislava

15 January 2019

Gender Robustness of Overconfidence and Excess Entry

Katarína Danková[a] and Maroš Servátka[b, c]

[a] MGSM Experimental Economics Laboratory, [b] MGSM Experimental Economics Laboratory, Macquarie Graduate School of Management in Sydney, Australia, [c] University of Economics in Bratislava, Slovakia

January 15, 2019

Abstract: Camerer and Lovallo (1999; henceforth CL) present thought-provoking experimental evidence that overconfidence might lead to excess entry into markets. As their findings are based on the majority of sessions exclusively consisting of male participants, we conduct two experiments in an attempt to replicate their study while including both men and women in all of our sessions. Our Experiment 1 closely follows CL design whereas Experiment 2 employs a gender-neutral addition task and provides more control in assessing gender differences in overconfidence and excess entry. Using participants of both genders we are unable to replicate CL's main finding that market entry decisions are driven by overconfidence. Contrary to CL, where self-selection increases the entry rate, in our Experiment 1 self-selection leads to less entry. This result is driven by self-selected females who rationally enter the market less often than self-selected males as their rank-determining performance on a sports and current events trivia quiz is worse than the performance of self-selected males. In Experiment 2 we find no effect of self-selection on entry and no gender differences in entry rates or performance in the addition task. Our results point out that (i) the finding that overconfidence leads to excess entry is not robust to a population consisting of both genders; and (ii) the self-selection effect is sensitive to both gender and task that is used to determine the rank upon entry.

JEL classifications: C9, C72, D2, D47
Keywords: Experiment, Gender, Market Entry, Overconfidence, Real Effort, Replication, Robustness, Self-selection

Acknowledgements: This paper is based on Katarína Danková's dissertation chapter written at the University of Canterbury (Danková, 2015). We wish to thank the guest editor Juergen Huber, two anonymous referees, Jeremy Clark, Miloš Fišar, Duncan James, Daniel Woods, and Lyla Zhang provided helpful comments and suggestions. We are particularly thankful to Miloš Fišar for programming mazes and helping us run Experiment 2 at the Masaryk University Experimental Laboratory. Financial support was provided by the University of Canterbury, College of Business and Economics and Macquarie Graduate School of Management.

1 Introduction

Do optimistic biases predictably influence economic behavior of firms when entering into markets? A large body of psychology and social psychology literature documents that people are overconfident about their relative abilities or unreasonably optimistic about their future (Alicke, 1985; Dunning et al., 1989; Messick et al., 1985; Svenson, 1981; Taylor and Brown, 1988). Overconfidence of entrepreneurs and managers could therefore be crucial for understanding failures of new businesses. Camerer and Lovallo (1999; henceforth CL) present thought-provoking experimental evidence that overconfidence might lead to excess entry into markets. As women are usually less overconfident than men (Lundeberg et al., 1994), CL findings are based on the majority of their sessions exclusively consisting of male participants.[1] With the increasing numbers of women managers, entrepreneurs and startup owners over the past decades, it is paramount to ascertain whether the excess entry finding is robust to a population consisting of both genders. We therefore replicate CL study while including both men and women in all of our sessions. We conduct two separate experiments; the first one closely replicates CL's design whereas the second one provides more control in assessing gender differences in overconfidence and excess entry.

CL offer three possible explanations for business failures: (1) Quick exits that appear to be failures are actually hit-and-run entries that are profitable but brief. Profits are made if entering the market during the high peak, i.e. when profitability is high, and then leaving (or "failing") when profitability dies down. Because of the fleeting nature of many business opportunities, a failure within a year of startup is probable and expected (Forbes, 2009). (2) Business entries are similar to lottery tickets, i.e. most firms expect to lose money and fail, but if they become successful, the payoff is large and worth the risk. Entrepreneurs understand the nature of risky entries and often report that the key to success is making profits on average rather than with every single investment. (3) Many entry decisions are simply mistakes due to underestimating the competitors or overconfidence about own abilities. Such mistakes are often hard to correct if the performance feedback is imperfect.

Distinguishing which one of the three explanations influences business failure and to what extent can be challenging with happenstance data. To explore the third explanation, CL design an incentivized laboratory experiment testing for the effect of overconfidence in one's skills on market entry decisions, i.e. whether overconfidence amplifies the market entry rate. In our study we use a mixed-gender subject pool to address the same two main questions as CL: 1. Is there more entry when people are betting on their own skills? 2. Are participants neglecting the reference group when they volunteer to participate in the experiment, knowing that their payoffs will depend on their skills? The answers to these questions deepen our understanding of the origins of business failures and can help design better performance feedback mechanisms.

Overconfidence occurs when an individual's certainty that his predictions are correct exceeds the accuracy of those predictions (Klayman et al., 1999). Hoelzl and Rustichini (2005) identify three sources of overconfidence: people may overestimate their own abilities, perceive themselves more favorably than others perceive them, or perceive themselves more favorably than they perceive others. Indeed, a large body of psychology and social psychology literature provides evidence that people are overconfident about their relative abilities or unreasonably

[1] CL's experiment consists of 8 sessions with both genders participating only in sessions 1 and 2. Sessions 3-8 were composed solely of male participants. CL use data from sessions 1-8 to analyze the link between overconfidence and excess entry as well as the impact of self-selection on entry decisions and data from sessions 3-8 to analyze whether the excess entry was caused by overconfidence or underestimating how many participants will enter the market in total.

optimistic about their future (Alicke, 1985; Dunning et al., 1989; Messick et al., 1985; Taylor and Brown, 1988). The effect has been labelled "better than average".[2]

Related theoretical and empirical literature in economics and finance focuses on explaining economic phenomena and particular aspects of behavior associated with overconfidence (Bénabou and Tirole, 2002; Bénabou and Tirole, 2003; Daniel et al., 1998; Gervais and Odean, 2001; Weinberg, 2009). Overconfidence in one's skills or relative ability can, in financial markets, lead to excessive trading and lower returns (Barber and Odean, 2001), distortions in corporate investment decisions (Malmendier and Tate, 2005), value-reducing mergers (Roll, 1986) and to security market anomalies (Daniel et al., 1998). It has been shown to influence the estimation of one's own ability, performance (Clayson, 2005), level of control (Presson and Benassi, 1996), speed with which one can get work done (Buehler et al., 1994), accuracy of one's beliefs (Alpert and Raiffa, 1982; Klayman et al., 1999; Soll and Klayman, 2004; Healy and Moore, 2007) or even expert judgements such as the accuracy of diagnoses (Christensen-Szalanski and Bushyhead, 1981). Overconfidence thus appears to be a robust phenomenon present in a wide array of professional and business-related decisions, with market entry being one of them.

The experimental literature, to a great degree triggered by CL, identifies numerous factors, such as task difficulty (Hoelzl and Rustichini, 2005; Moore and Cain, 2007) leading to overconfidence. An example of such task where overconfidence plays an important role is the introduction of risky products to the market (Simon and Houghton, 2003). Similarly, greater overconfidence tends to lead to aggressive behavior in the pursuit of higher wealth (Deaves et al., 2009). Interestingly, experience and specialization can, in some scenarios, contribute to overconfidence; for example studies by Kirchler and Maciejovsky (2002); Glaser et al. (2005); Glaser et al. (2007) show that experts are more likely to be overconfident than relatively inexperienced subjects. However, some other studies find that overconfidence tends to decrease with experience (Christoffersen and Sarkissian, 2002; Gervais and Odean, 2001; Locke and Mann, 2001). The evidence on gender effects in overconfidence is also mixed as the findings appear to vary with the task, activity, and/or environment. For example, Beyer (1990) observes that the tendency to ascribe success to personal effort and failure to external forces is less pronounced in women, while Deaves et al. (2009) find little evidence that gender influences trading activity, hinting that more research is necessary to understand the prevalence of overconfidence and its driving factors. Testing market entry decisions in a population composed of both genders is a step in this direction.

Gender differences in overconfidence are closely linked to competitive attitudes of men and women. A large literature in this area demonstrates that competitive environments tend to hinder the performance of women, especially if the task favors men (Gneezy et al., 2003; Günther et al., 2010) and that women are less likely to self-select into competitive environments (e.g. Gneezy and Rustichini, 2006), even if there are no significant differences in performance and when overconfidence is controlled for (Niederle and Vesterlund, 2007).[3] Other recent studies examine factors that mediate the relationship between gender and preference for competition (e.g. Balafoutas and Sutter, 2012; Cadsby et al., 2013; Gneezy et al., 2009; Vandegrift and Yavas, 2009; Wozniak et al., 2014) and tasks and environments that mitigate this relationship (e.g. Andersen et al., 2013; Balafoutas and Sutter, 2012; Ertac and Szentes, 2011; Flory et al., 2014; Healy and Pate, 2011; Müller and Schwieren, 2012; Niederle et al., 2013; Vandegrift and Yavas, 2009). Our contribution to this literature stems from extending the analysis of gender differences to the market entry game and exploring whether the

[2] A popular example of overconfidence is asking a group of average people about their driving ability. Most of them will say they are above average even though only about half can be better than average (Svenson, 1981).

[3] A notable exception is Price (2011) who does not replicate Niederle and Vesterlund (2007) competitiveness finding using a seemingly identical experimental design.

willingness to compete based on one's skills depends on self-selecting into the experiment. Furthermore, by varying the nature of the task, performance in which is used to determine the rank, we are able to explore the sensitivity of gender differences in market entry decisions to whether the task favors one of the genders or not. CL's design allows one to identify whether market entry, which is also a competitive task, is driven by overconfidence. In the market entry game, whether an entry is successful or not depends on the market capacity and the entrant's rank. In one scenario, henceforth referred to as skill-rank, the rank is determined by performance in a task and thus the decision whether to enter depends on one's confidence in his skills relative to others. In the control scenario, henceforth referred to as random-rank, the rank is determined randomly. CL find that in rounds when payoffs from entry depend on skills, excess entry is higher than in rounds when payoffs are determined randomly, providing evidence of overconfident behavior. Furthermore, excess entry is highest when the participants are told in advance that their payoffs will depend on their skills, suggesting that participants who self-select into such sessions neglect consideration of the reference group with which they will be competing.[4] CL use expected average profits to distinguish whether the excessive entry is caused by overconfidence or underforecasting and find that overconfidence is the main driver.

To the best of our knowledge, there is no study testing the robustness of the CL's finding with respect to both genders. Including female participants, who have been shown to be less overconfident than males in various other contexts, constitutes a more conservative test of the effect of overconfidence on market entry decisions and increases the external validity of CL's study. Our Experiments 1 and 2, designs of which are presented in detail in sections 2 and 4 respectively, test the original CL's hypotheses while also allowing us to compare the behavior of men and women.

Hypothesis 1: There is more entry (and thus lower industry profit) in skill-rank rounds than in random-rank rounds.

If participants are overconfident, they will enter the market more often in skill-rank rounds, which will result in lower industry profits, i.e. if the number of entrants is higher than the market capacity, the industry profit will be negative.

Hypothesis 2: The profit differential between skill-rank and random-rank rounds in sessions with self-selection is larger than in sessions with no self-selection.

The larger the skill-rank and random-rank profit differential, the more entry will be observed in the skill-rank rounds. If the entrants neglect the reference group, i.e. enter more often because of overconfidence in their skills but ignore that all other entrants are doing the same, the differential between the skill-rank and random-rank rounds will be larger in sessions with self-selection than in session with no self-selection.

Hypothesis 3: The expected average profit is smaller in skill-rank rounds than in random-rank rounds.

The expected average profit is calculated based on the forecasts of participants. If participants decide to enter because they think fewer people will enter, then the expected average profit will be higher in skill-rank rounds than in random-rank rounds. If, however,

[4] Reference group neglect is also known in the literature as egocentrism (Kruger, 1999; Windschitl and Chambers, 2004).

participants enter more often because they are overconfident about their relative skills, the expected average profit will be lower in skill-rank rounds than in random-rank rounds.

As CL's design varies market capacities across sessions, we construct a normalized entry rate, a novel measure that allows us to compare data across Experiment 1 sessions that do not have the same capacities in each round. Despite that, we do not obtain the same results as CL. In Experiment 1 we find only very weak evidence that the industry profit is lower in skill-rank rounds than in random-rank rounds due to more entry in the skill-rank rounds. We also find that self-selection leads to less entry, in stark contrast to CL. This result is driven by self-selected females who rationally enter less often than self-selected males, likely due to the nature of the employed tasks. We also find no difference in expected average profits between random-rank and skill-rank rounds, suggesting that overconfidence might not be as strong of a driving factor of entry decisions when both genders are represented amongst the market participants. Regarding performance in tasks that determine the rank in skill-rank rounds, we observe that males perform significantly better on a trivia quiz about sports and current events (a task that could be perceived as favoring males) and are faster to complete the mazes task (which also favors males according to previous literature), however the latter result is not statistically significant. In Experiment 2 employing a gender-neutral addition task, we find no evidence that the industry profit is lower in skill-rank rounds than in random-rank rounds, signifying no difference in entry rates due to overconfidence. We also find no effect of self-selection on entry and no gender differences in entry rates or performance in the addition task.

In summary, using participants of both genders we are unable to replicate CL findings that there is more entry in the skill-rank compared to the random-rank rounds or that self-selection increases entry. Our results also point out that the self-selection effect is sensitive to both gender and task that is used to determine the rank upon entry.

2 Experiment 1 design and procedures

CL employ the market entry game introduced by Selten and Güth (1982) to study the link between overconfidence and decisions to enter the market.[5] In what follows we present the CL modification of the game with rank-based payoffs. The design of Experiment 1 follows CL in terms of the implemented parameters, session ordering, and procedures to the extent known to us. Any differences are discussed below.

In the repeated market entry game, each participant is endowed with \$10 and is informed about the market capacity "c", where $c \in \{2, 4, 6, 8\}$ (for the capacity in each round used in the same sequence in both CL and our Experiment 1, see Table C1 in Appendix C). The participants in each round simultaneously choose whether to enter the market or not. The payoff to the entrants depends on the overall number of entrants, market capacity c announced at the beginning of the round and the entrant's rank. Entrants ranked below c lose their initial endowment, while entrants ranked c or above earn a positive sum of money (see Table 1). The top c entrants share \$50 proportionally, with higher-ranked entrants earning more relative to other entrants.[6] Non-entrants do not earn or lose any money; they keep their initial endowment.

The rank is assigned randomly or based on the participant's skills as determined by performance in a real-effort task. In CL sessions 1 and 2, participants were asked to solve "puzzles," details of which were not reported as part of the experimental protocol. We

[5] For a theoretical analysis of the standard version of the market entry game and a recent review of the empirical literature see Collins et al. (2017).

[6] If the number of entrants is lower than c, the entrants share \$50 proportionally, i.e. the entrant with the lowest rank receives the smallest \$ amount, the entrant with the second lowest rank receives twice as much as the previous one etc.

operationalize solving “puzzles” as completing “mazes,” a task that is comparable in terms of skills.[7] In sessions 1 and 2 of our Experiment 1, skill-ranks are thus determined by the speed of completing five mazes. In sessions 3-8 of both CL and our Experiment 1, skill-ranks are determined by the number of correct answers on a trivia quiz about sports and current events.

Table 1. Experiment 1 rank-based payoffs*

Rank	Market Capacity			
	c=2	**c=4**	**c=6**	**c=8**
1	33	20	14	11
2	17	15	12	10
3		10	10	8
4		5	7	7
5			5	6
6			2	4
7				3
8				2

* Payoff in $ for successful entrants as a function of “c”

The game is played in two blocks, each consisting of 12 rounds (24 rounds in total). In one of the two blocks the rank is determined randomly (R), in the other block the rank depends on skills (S). This feature is implemented in a within-subject design, i.e. the same participants participate in both blocks of rounds. The participants are told in advance in which block of rounds the rank is assigned randomly and in which it depends on their skills. To control for order effects, in half of the sessions the block of rounds with random-rank is conducted first, followed by the block of rounds with skill-dependent rank. In the other half of the sessions the order is reversed, i.e. the block of rounds with skill-dependent rank is conducted first and the random-rank second. With the exception of sessions 1 and 2 as in the original CL experiment, the random-rank rounds have the exact same order of c’s as the skill-rank rounds and thus the two blocks are directly comparable.

The design ensures all participants make decisions in both random-rank and skill-rank rounds and thus their decisions in the random-rank rounds act as a within-subject control for risk preferences, a potentially important consideration when deciding whether to enter the market or not. Given the considerable variation of risk-taking behavior observed in previous research (nicely summarized in e.g. Deck et al. (2013), such approach is simpler and likely to generate less noise than eliciting risk preferences using an unrelated risky task. Along with their individual entry decisions, the participants forecast how many entrants they expect in that round. For each correct forecast, the participants earn $1. CL use these forecasts to distinguish between participants who enter because they underestimate the number of competitors and participants who are overconfident about their skills and who therefore enter because they think their performance on the quiz or maze is better than average. The participants’ ranks are not revealed until the end of the experiment, i.e. after their market entry decisions for all 24 rounds.

A total of 118 participants, 59 males and 59 females, took part in our Experiment 1. The experimental sessions were conducted in the New Zealand Experimental Economics Laboratory (NZEEL) at the University of Canterbury. Participants were recruited using the online database system ORSEE (Greiner 2015). Each participant only participated in a single

[7] In an email conversation, CL provided further details that the puzzles in the first two sessions were “brain teasers” taken from a book they bought. Note that CL’s design involved solving ten puzzles whereas our Experiment 1 design involved solving five mazes. We calibrated the number of mazes based on the expected time (10 minutes) it would take the participants to solve them. Our objective was to implement a task that would require display of skills but that would not unnecessarily prolong the experimental session.

session of the study, and had not participated in any similar market entry experiment run at NZEEL.

The invitation, similarly as in CL, differed in information provided to the participants before signing up for the experiment. In sessions 1-4 the participants were invited to participate in the experiment with an opportunity to make money. In addition to that, in sessions 5-8, the participants were told in the invitation email that their payoff in the experiment would depend on their skills, especially their knowledge about current events and sports.[8] In these latter sessions it was possible for participants confident of their abilities to self-select into the experiment (see Table 2 for the session overview).

Unlike CL experiment, our Experiment 1 was fully computerized (i.e. including the mazes) using z-Tree (Fischbacher, 2007). As reported in Table 2, the number of participants in a session varied from 12 to 16, following CL. All sessions were run under a single-blind social distance protocol in which there was a complete anonymity between the participants but not with respect to the experimenter. On average, a session lasted 50 minutes including the payment. The participants earned 13.80 NZD on average.[9]

Upon entering the laboratory, the participants were asked to sit in a cubicle of their choice. At the beginning of the experiment instructions (provided in Appendix A) were handed out, as well as projected onto a screen and read aloud by the experimenter. The participants then had a few minutes to go through the instructions again, this time in their own pace. Any questions arising were answered in private. All participants had to answer the control questions (provided in Appendix B) correctly before they could proceed to the decision-making part of the experiment. This procedure allowed us to assess the understanding of instructions and clarify any confusion. After the control questions, the participants first entered their decisions in each of 24 rounds and only then engaged in a task that determined their rank for the skill-dependent rounds. Upon the completion of the experiment, they were also asked to fill out a questionnaire. Participants were then called one by one to collect their payment in private in the control room at the back of the laboratory.

Table 2. Experiment 1 sessions overview

Session #	n	Invitation	Block Order	Task
1	12	No self-selection	R/S	Maze
2	14	No self-selection	S/R	Maze
3	16	No self-selection	R/S	Quiz
4	16	No self-selection	S/R	Quiz
5	16	Self-selection	R/S	Quiz
6	16	Self-selection	S/R	Quiz
7	14	Self-selection	R/S	Quiz
8	14	Self-selection	S/R	Quiz

R= random-rank, S=skill-rank

3 Experiment 1 results

The section is organized as follows: We attempt to replicate CL's results by applying the tests they use to our data. These results are always reported first. It is important to note that

[8] CL do not report the details how their participants were invited to the experiment or what the communication channel was. The information included in the invitation email to participants in the NZEEL database was as follows: "Earn money in an experiment in which performance on sports and current events trivia will determine your payoff. If you are very good you might earn a considerable sum of money." The latter part of the sentence is reproduced from the CL paper.

[9] For reference, at the time of the experiment 1 NZD = 0.7883 USD and the adult minimum wage in New Zealand was 14.25 NZD per hour.

in CL's design, which Experiment 1 closely replicates, sessions 1 and 2 have a different order of c's across rounds. That is, the participants in session 1 face a different order of c's than the participants in session 2, making these two sessions not directly comparable. To rectify the issue, in addition to CL's analysis, we (i) analyze data only from sessions 3-8 and (ii) we calculate the normalized entry rate that addresses the different order of c's in sessions 1 and 2 and thus allows us to perform tests on data from all sessions (i.e. 1-8).

The industry profit, calculated by adding profits of successful entrants and losses of unsuccessful entrants in a given round, is strictly positive in 81 (=84%), negative in five, and zero in the remaining ten out of 96 random-rank rounds (12 rounds/session x 8 sessions = 96 rounds in each block). The average industry profit across random-rank rounds is $29.28. Out of 96 skill-rank rounds, the industry profit is strictly positive in 76 (=79%), negative in nine, and zero in eleven rounds. The average profit across skill-rank rounds is $26.14.

Industry profit and market entry

Hypothesis 1 states there will be a lower industry profit resulting from more entry in the skill-rank rounds than in the random-rank rounds. The hypothesis is based on a conjecture that when participants are betting on their own skill they will enter more often, which will in turn lower the total industry profit.

Following CL, we first test for differences in the industry profit between the skill-rank and random-rank rounds using a matched pairs t-test.[10] Recall that each experimental session consists of two blocks composed of twelve random-rank and twelve skill-rank rounds. The test is conducted as follows. The industry profit from the first twelve random-rank rounds in session 1 is matched with industry profit from the first twelve skill-rank rounds in session 2. Similarly, the industry profit from the skill-rank rounds in session 3 is matched with the industry profit from the random-rank rounds in session 4. In the same way session 5 with 6, and 7 with 8 are matched. In sessions 3-8, each pair of rounds being compared has the same value of c's, the same history (or path) of previous values of c's, and differs only in how the rank was determined. We followed this procedure in order to replicate the original design by CL and so preserved the order of c's in sessions 1 and 2. The matched pairs t-test does not detect a difference between profits in the random-rank rounds and skill-rank rounds (p-value=0.193). Our Experiment 1 result differs from the one obtained by CL who find that the industry profit is significantly lower in the skill-rank rounds than in the random-rank rounds when using individual data from sessions 1-8 (CL's p-value < 0.001).

The fact that participants in session 1 face a different order of c's than participants in session 2, makes these sessions not directly comparable. We thus exclude these two sessions from the matched pairs t-test. The t-test for sessions 3-8 supports CL's finding that there is more entry in the skill-rank rounds (i.e. more overconfidence in one's skill) as the industry profits are lower in the skill-rank than in the random-rank rounds (p-value=0.084) albeit this effect is weaker than in CL.

In order to be able to use the data from sessions 1 and 2 (which do not have the same order of c's) we calculate a normalized entry rate for each round. The normalized entry rate is the ratio of the number of entrants and the actual capacity c in the respective round, where 100% means that the number of entrants was exactly the same as c in the given round. If the normalized entry rate is higher than 100%, there are more entrants than c. If it is less than 100%, the market is not saturated and it is possible for more participants to enter the market and make profit. By calculating the normalized entry rate we are able to control for different c's in the given round between sessions 1 and 2. Using the normalized entry rate we then test whether

[10] We report the t-test in order to make our results easily comparable with CL.

there is more entry (a higher normalized entry rate) in the skill-rank rounds than in the random-rank rounds. The t-test does not detect a statistically significant difference in normalized entry rates between the skill-rank and random-rank rounds (p-value=0.482).

In addition to tests reported by CL, one can also test for within-subject comparisons as each participant took part both in the random-rank and skill-rank rounds. To test whether there is a difference in industry profit as well as in normalized entry rates between random-rank rounds and skill-rank rounds within a session we use the Wilcoxon matched-pairs signed-rank test; p-values for each session are reported in Table 3 below. Except for the industry profit in session 6, where the profit in the random-rank rounds is significantly higher than in the skill-rank rounds, none of the other tests show that industry profits are significantly different in the random-rank rounds than in the skill-rank rounds of the same session.

In summary, using participants of both genders we find weak evidence of excess entry due to overconfidence when comparing behavior in the skill-rank rounds compared to the random-rank rounds.

Table 3. The Wilcoxon matched-pairs signed-rank test for the difference in industry profits and normalized entry rate between the random-rank and skill-rank rounds in Experiment 1 (within subjects)

	Industry profit differences			Normalized entry rate		
Session #	**Random-rank**	**Skill-rank**	**Wilcoxon matched-pairs signed-rank test (p-value)**	**Random-rank [%]**	**Skill-rank [%]**	**Wilcoxon matched-pairs signed-rank test (p-value)**
1	370	450	0.310	140.0	131.8	0.843
2	330	260	0.478	176.4	189.7	0.3762
3	180	120	0.174	241.7	240.3	0.237
4	330	310	0.657	191.0	195.1	0.693
5	400	370	0.250	158.3	173.0	0.172
6	330	160	0.012	184.0	231.3	0.009
7	340	330	0.809	165.7	171.9	0.691
8	530	510	0.657	94.9	108.8	0.265

Reference group neglect

Hypothesis 2 states that the profit differential between the skill-rank and random-rank rounds in sessions with self-selection is larger than sessions without self-selection due to the reference group neglect.

Following CL, we first conduct a matched pairs t-test comparing the skill-random profit differential between sessions 1-4 (no self-selection) and 5-8 (self-selection), the result of which does not support Hypothesis 2 (p-value=0.432). Unlike us, CL observe that the reference group neglect produces a significantly larger skill-random rank entry differential in sessions with self-selected participants than in sessions without self-selection (p-value < 0.001); see Table 4 for a comparison of our results with CL.

In addition to the test performed using data from sessions 1-8 as in CL, we run a matched pairs t-test comparing the skill-random profit differential between sessions 3-4 and 5-8 (i.e. excluding sessions 1 and 2 that have different order of c's). This test also does not support the hypothesis that the differential is larger in sessions with self-selection than without (p-value=0.659).

In summary, applying the t-test (used by CL) to our data, we find that the profit differential between skill-rank and random-rank rounds does not differ in sessions with self-selection than without self-selection, pointing out that the overconfidence effect does not increase with self-selection in a population composed of both genders.

Table 4. Comparison of CL's results and Experiment 1 results

	CL (sessions 1-8)	**Experiment 1 (sessions 1-8)**	**Experiment 1 (sessions 3-8)**
Avg. profit random-rank	\$16.87	29.27	\$29.31
Avg. profit skill-rank	\$-1.56	\$26.15	\$25.00
Matched pairs t-test	t=-7.43 p<0.001	t=1.311 p=0.193	t=1.755 p=0.084
Avg. profit without self-selection, random-rank	\$19.79	\$25.21	\$21.25
Avg. profit without self-selection, skill-rank	\$10.83	\$23.75	\$17.92
Avg. profit self-selection, random-rank	\$13.96	\$33.33	\$33.33
Avg. profit self-selection, skill-rank	\$-13.13	\$28.54	\$28.54
Matched pairs t-test	t=-4.08 p<0.001	t=0.793 p=0.432	t=-0.447 p=0.659

Expected profit differential in skill-rank and random-rank rounds

The results in the previous subsections provide weak evidence for overconfidence resulting in excess entry and demonstrate that self-selection does not increase the strength of the overconfidence effect. These tests, however, do not control for all possible explanations. Excessive entry in the skill-rank rounds may not necessarily be due to overconfidence about one's skills, but due to the underestimating how many participants will enter in total. CL call this the "blind spot" hypothesis. If the number of expected entrants is underestimated, it decreases the participants' payoffs because they enter even though they should not. In order to test whether the expectations are correct, we ask participants to saliently forecast the number of entrants in each round.[11]

On average, the number of forecasted entrants in all sessions is 6.07 and 6.23 in the random-rank and skill-rank rounds, respectively. The actual number of entrants in all sessions is on average 5.75 and 6.26 in the random-rank rounds and skill-rank rounds, respectively. The difference between the forecasted and actual number of entrants in the random-rank rounds is not statistically significantly different (Mann-Whitney, p-value=0.599).[12] In the skill-rank rounds this difference is not statistically significant either (Mann-Whitney, p-value=0.916). In the random-rank rounds participants forecast about 0.32 entrants too high and in the skill-rank rounds their forecast converges to the actual number of entrants.

To separate overconfidence from incorrect estimates of others' entry CL use the obtained forecasts to compute the profit that a participant expects the average entrant to earn, calculated in a following way:

$$E_j(\pi_{ijt}) = (50-10*(F_{ijt} - c_{it}))/ F_{ijt}, \tag{1}$$

where $E_j(\pi_{ijt})$ is the expected average profit, F_{ijt} is the forecast of participant j used to calculate the profit that participant j expects the average entrant to earn, and c_{it} is the capacity in the particular round.

Separating overconfidence from incorrect estimates of others' entry requires testing the hypothesis that the expected average profit is larger in the random-rank rounds than in the skill-rank rounds. If participants decide to enter in the skill-rank rounds because they think that fewer other participants will enter, the expected average profit in the skill-rank rounds will be larger. Including $E_j(\pi_{ijt})$ in the entry regression, reported in the next subsection, will separate out the effect falsely attributed to skills. If, on the other hand, the participants enter because they are more overconfident in the skill-rank rounds compared to the random-rank rounds, not taking into account the number of entrants they expect to enter, the expected average profits will be smaller in the skill-rank rounds than in the random-rank rounds. The overconfident participants will expect to earn more than the average entrant and enter even when the expected average profit is low.

Following CL, we therefore calculate the differential between expected average profits in the random-rank rounds (denoted π_r) and in the skill-rank rounds (denoted π_s), using only the rounds in which participants entered. A negative differential, i.e. larger profits in the skill-rank rounds than in the random-rank rounds, represents the incorrect estimation of entrants, whereas a positive differential represents overconfidence. In Table 5 we report the mean differential π_r - π_s, averaged across entering participants, the number and percentage of participants who have a negative mean (i.e. who expect less average profit in the skill-rank rounds), and the number and percentage of participants whose expected average profit is negative, on average, across the random-rank rounds and skill-rank rounds. For completeness, in rows 3 and 4 we also report a percentage of entrants whose profit is lower than 0 in the random-rank rounds and skill-rank rounds.

The mean differential π_r - π_s is negative in sessions 1 and 2, suggesting an incorrect estimation of number of entrants by participants, and positive in sessions 3 and 4, suggesting the presence of overconfidence. In session 1, 44% of the participants expect to earn less in the skill-rank rounds than in the random-rank rounds. In sessions 2, 3 and 4, it is respectively 57%, 46% and 67% of participants. In the self-selection sessions, the mean differential π_r - π_s is

[11] The specific question we asked before each round is: "How many people (including yourself) do you expect to enter the market in this round?" If a participant forecasted the number of entrants correctly, $1 was added to his/her payoff in the respective round.
[12] CL only report a regression in which they use data from sessions 3-8.

negative in sessions 5-7, while in session 8 the differential is positive and suggests the presence of overconfidence. In session 5 only 18% of the participants expect to earn less in the skill-rank rounds than in the random-rank rounds. In session 6, 7, and 8 it is 57%, 38% and 75%, respectively. The t-test does not detect a significant difference in the average differential of expected profits per person between the self-selection sessions and sessions without self-selection (p-value=0.913). In other words, there is no difference in the expected average profit between the skill-rank rounds and random-rank rounds.

Table 5. The average differential in expected profits per entrant between the random-rank and skill-rank rounds in Experiment 1

Measure	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Total
Πr - Πs	-1.126	-0.665	0.023	1.832	-1.094	-0.886	-1.718	4.483	1.036
# of entrants with Πr - Πs<0 (percent)	5/9 (56%)	6/14 (43%)	7/13 (54%)	4/12 (33%)	9/11 (82%)	6/14 (43%)	8/13 (62%)	2/8 (25%)	47/94 (50%)
# of entrants with Πr <0 (percent)	0/9 (0%)	0/13 (0%)	3/12 (25%)	1/11 (9%)	0/10 (0%)	0/11 (0%)	0/12 (0%)	0/8 (0%)	4/86 (5%)

Logistic regression

Table 6 reports a fixed-effects logistic regression of the entry decision. Model 1 is the basic model. Model 2 controls for the demographic characteristics. In line with the t-test results using the industry profits as well as the normalized entry rates, we find that the entry rate in the skill-rank rounds is not higher than in the random-rank rounds.

Unlike the results of the t-test, our regression results show that participants who received an invitation email saying that their payoffs in the experiment (variable Self-selection in the regression) will depend on their skills enter less often than those who received a generic invitation email without such information, which is in stark contrast to CL where self-selection increases the entry rate. The coefficient associated with the interaction term of being male and self-selection is statistically significant in our data, meaning that in our self-selection sessions males enter more often than females. The difference in entry rates of males and females is highly statistically significant (p-value < 0.001; chi-square test). This fact also likely drives the result that self-selection leads to overall less entry. Table 7 presents a comparison of male and female entries separately for self-selection and no self-selection sessions. While we observe a slightly higher entry rate by males in self-selection sessions than in no self-selection sessions (the direction observed also by CL in their male-only population), this difference is not statistically significant (p-value = 0.296; chi-square test). On the other hand, females enter significantly less often in self-selection sessions than in no self-selection sessions (p-value < 0.001, chi-square test).

Table 6. Experiment 1 fixed-effects logistic regression of the entry decision

Experiment 1	Model 1			Model 2		
Variable	**Coef (Robust Std. Err)**	**Marginal**	**z-statistic (p-value)**	**Coef (Robust Std. Err)**	**Marginal**	**z-statistic (p-value)**
Intercept	-0.096 (0.143)		-0.67 (0.502)	-0.367 (0.403)		-0.91 (0.363)
C	-0.037* (0.021)	-0.009	-1.74 (0.081)	-0.038* (0.022)	-0.008	-1.70 (0.089)
Maze	0.179 (0.113)	0.041	1.57 (0.115)	0.208* (0.123)	0.044	1.69 (0.091)
$E(\pi_{ijt})$	-0.017*** (0.004)	-0.004	-3.76 (0.001)	-0.016*** (0.005)	-0.003	-3.50 (0.001)
Self-selection	-0.682*** (0.152)	-0.042	-4.48 (0.001)	-1.051*** (0.175)	-0.100	-6.00 (0.001)
Skill rank	0.045 (0.132)	0.032	0.34 (0.735)	0.052 (0.137)	0.033	0.38 (0.706)
Male	0.038 (0.136)	0.114	0.28 (0.778)	-0.051 (0.154)	0.101	-0.33 (0.741)
Age				-0.059*** (0.010)	-0.013	-5.72 (0.001)
Non NZ Nationality				-0.172 (0.124)	-0.037	-1.39 (0.165)
Siblings				0.093*** (0.033)	.020	2.85 (0.004)
Relative Income				-0.101* (0.059)	-0.021	-1.71 (0.087)
City size				0.466*** (0.053)	0.100	8.75 (0.001)
Living with others				0.111*** (0.027)	0.024	4.17 (0.001)
Money				0.001 (0.001)	0.001	0.06 (0.952)
Finance study				0.001 (0.001)	0.001	0.62 (0.534)
Rely				-0.043** (0.020)	-0.009	-2.22 (0.027)
Law				0.703*** (0.157)	0.149	4.49 (0.001)
Humanities				-0.070 (0.179)	-0.015	-0.39 (0.696)
Natural Sciences				0.286** (0.132)	0.061	2.16 (0.031)
Other Social Sciences				0.846*** (0.162)	0.180	5.24 (0.001)
Engineering				0.609 (0.143)	0.129	4.24 (0.001)
Medical Science				-0.408 (0.401)	-0.087	-1.02 (0.308)
Self-selection *Skill rank	0.119 (0.157)		0.76 (0.447)	0.126 (0.164)		0.77 (0.442)
Self-selection *Male	0.844*** (0.159)		5.32 (0.001)	0.979*** (0.187)		5.25 (0.001)
Skill rank*Male	0.068 (0.157)		0.43 (0.666)	0.077 (0.164)		0.47 (0.641)

Log-likelihood	-1855.82	Pseudo R2 = 0.0303	Log-likelihood	-1732.88	Pseudo R2 = 0.095

Sessions 1-8 (CL report sessions 3-8; we provide such regression in Table C2 in appendix C), n=2832, standard errors are not clustered at the session level because of the small number of sessions.
Description of demographic variables: non New Zealander represents participants who are not from New Zealand; Siblings represents the number of siblings; Relative income represents whether the income is far below average, below avg., avg., above avg., or far above avg. (from 1 to 5). City size, Living with others, Money, and Finance study represent respectively the size of the city from 2000 to 100 000+; number of people in a household, the size of the monthly budget, and the share of monthly expenses one finances alone. Rely is a self-reported variable that indicates on a scale 1 to 9 the reliability of the information provided in a questionnaire with 9 being the most reliable. Law, Humanities, Natural Sciences, Other Social Sciences, Engineering, Medical Science are dummy variables representing the fields of study; the omitted variable is Economics.
Run on StataSE 13.0. Robust standard errors used.
*, **, *** refer to statistical significance at the 10%, 5% and 1% levels, respectively.

The remaining interaction terms in the logistic regression, Skill rank*Male and Skill rank*Self-selection, are not significant, suggesting that the effect of skill-rank on the entry decision is not mediated by gender or self-selection. We also find no gender differences in risk preferences as measured by the entry rate in the random-rank rounds.

Similarly to CL, we also find the effect of the expected profit to be negative and significant. CL hypothesize that this is due to participants planning to enter and forecasting a lot of entry, so the expected average profit is lower when they enter and relate the explanation to the false consensus effect in which people use their own decision as a clue about what others will do. Finally, law, natural sciences and other social sciences students enter more often than economics students and participants in sessions with mazes enter more often than participants in sessions with the quiz when controlling for all demographics (Model 2 in Table 6).

CL in their paper report regression results only for sessions 3-8, consisting solely of male participants. Recall that the rank in skill-rank rounds in sessions 3-8 is determined by performance on a trivia quiz about sports and current events. When restricting the dataset to sessions 3-8 of our Experiment 1, we find our two main results of self-selection decreasing the entry rate and self-selected males entering the market more often than self-selected females to be robust as both Self-selection and the interaction term Self-selection*Male remain significant. For the regression results from sessions 3-8 only, see Table C2 in Appendix C.

Table 7. Number of decisions in which participants entered the market

	Female	Male	Total
Self-selection	189 (out of 696)	355 (out of 744)	544
No Self-selection	307 (out of 720)	302 (out of 672)	609
Total	496	657	

Note: In sessions 1-8 we have 1416 female decisions and 1416 male decisions. In no self-selection sessions we have 1392 decisions and self-selection sessions1440 decisions.

Gender differences in performance on the quiz and in the mazes task

To analyze whether the nature of the task affects performance, we compare the scores achieved by males and females. The results and Mann-Whitney tests comparing the performance for each session in Experiment 1 are summarized in Table 8. At the session level, we find that males perform better on the quiz only in session 5. Using pooled data from sessions 3-8, however, we find that males perform significantly better on the quiz than females (Mann-Whitney test, p-value =0.020), confirming that the quiz is not a gender- neutral task. We find that effect to be mostly driven by self-selected males who perform significantly better than self-selected women (p-value = 0.016), whereas there is no statistical difference between the performance of males and females in the sessions without self-selection (p-value = 0.606). The observed gender difference in performance in self-selection sessions is in line with participants' decisions in the market entry game, as self-selected males enter the market more often than self-selected females.

At the overall level, we observe no difference in performance of participants in the no self-selection and self-selection sessions (sessions 3-4 and 5-8, respectively; p-value = 0.590, Mann-Whitney test), signifying that self-selection into the experiment when using participants of both genders does not have the effect conjectured by CL. (Note that CL's paper does not report the performance results of their participants.) Recall that reference group neglect, which is the reason for including self-selection sessions in the design, builds on the idea that excess entry rates could be the result of participants failing to acknowledge the strength of their competitors who also self-selected into these sessions.

For the mazes task, we find that on average females are slower than males, but the difference is not statistical significant. However, this result is likely driven by the small number of observations (10 males and 16 females in total) as mazes were used only in the no self-selection (1 and 2) as in CL's design.

Table 8. Average performance across sessions in Experiment 1

	Experiment 1		M-W test
	Male	**Female**	
Session 1*	602.9 sec	623.5 sec	z=0.092 p=0.926
Session 2*	625.0 sec	641.1 sec	z=-0.192 p=0.848
Session 3	11.55	12.60	z = 0.57 p = 0.567
Session 4	13.29	10.56	z = -1.12 p = 0.264
Session 5	16.11	11.43	z = -3.11 p = 0.002
Session 6	12.71	11.33	z = -0.85 p =0.393
Session 7	12.00	11.38	z = -0.33 p =0.744
Session 8	12.00	9.20	z = -1.08 p = 0.282

*In Sessions 1 and 2, the lower the time of completing the mazes, the better the performance. The score in sessions 3-8 refers to the average number of correctly answered questions on the quiz. There is no difference in performance between no self-selection (Session 3 and 4) and self-selection sessions (Sessions 5-8), (Mann-Whitney, p-value=0.590).

4 Experiment 2 design and procedures

While Experiment 1 closely replicates CL's design, it also contains certain design features that might influence the behavior of males and females in different ways and thus affect the obtained results. First and foremost, mazes (or puzzles) together with a trivia quiz about sports and current events might be viewed as male tasks (Günther et al., 2010), potentially affecting our results in two ways. Namely, including a male task in the design might create a direct effect of gender over rank and an indirect effect stemming from the participant's confidence regarding the performance on the implemented task. To control for the task favoring a particular gender, in Experiment 2 we implement the so-called addition task, demonstrated by Niederle and Vesterlund (2007) to be gender-neutral. We implement the addition task in all 8 sessions.

Additionally, the gender composition within each session might also play a role. Rather than ensuring that the overall number of males and females is constant, Experiment 2 ensures that an equal number of males and females participate in all sessions. We also keep the total number of participants constant at 14 across all sessions and announce this number in the instructions. Finally, the order of market capacities is also kept constant to make the skill-rank round directly comparable with the random-rank rounds.

Experiment 2 was run in the Masaryk University Experimental Economics Laboratory (MUEEL) in Brno, Czech Republic since in the meantime both of the authors moved and NZEEL, where Experiment 1 was run, no longer existed. Experiment 2 was conducted in English by the same experimenter (Katarína Danková) and using the same instructions as Experiment 1 with payoffs denoted in Czech Crowns (Česká koruna or CZK).[13] The MUEL lab policy was to pay on average 150-160 CZK for incentivized decisions, which was approximately ten times as much in nominal terms as the payment policy in NZEEL, where the average payments ranged between $14-16. The rank-based payoffs were thus calculated to reflect the ratios used in CL and Experiment 1 and set to be equal to the nominal values from Experiment 1 multiplied by ten. This approach ensures that the maximum feasible payment ($33 in NZEEL and 330 CZK in MUEEL) is approximately twice as much as the average payment in the given lab.

Each participant in Experiment 2 is thus endowed with 100 CZK each round and the top c entrants share 500 CZK proportionally (see Table 9 for details). The rank is assigned randomly or based on the participant's skills as determined by number of correct answers in the addition task. All other procedures and parameterizations were analogous to Experiment 1.

Table 9. Experiment 2 rank-based payoffs*

Rank	Market Capacity			
	c=2	c=4	c=6	c=8
1	330	200	140	110
2	170	150	120	100
3		100	100	80
4		50	70	70
5			50	60
6			20	40
7				30
8				20

* Payoff in CZK for successful entrants as a function of "c"

A total of 112 participants, 56 males and 56 females, took part in the experiment. Participants were recruited using the online database system HROOT (Bock et al., 2012). Each participant only participated in a single session of the study, and had not participated in any similar market entry experiment run at MUEEL. On average, a session lasted 45 minutes

[13] The exchange rate at the time of the experiment was 1 CZK = 0.045 USD. The show up fee in MUEL was 100 CZK.

including the payment. The participants earned on average 245 CZK.

As in Experiment 1, the invitation differed in information provided to the participants before signing up for the experiment. In sessions 1, 3, 5, and 7 the participants were invited to participate in the experiment with an opportunity to make money. In addition to that, in sessions 2, 4, 6, and 8 the participants were told in the invitation email that their payoff in the experiment would depend on their skills, especially on their performance in an addition task. In these latter sessions it was possible for participants confident in their abilities to self-select into the experiment (see Table 10 for Experiment 2 session overview).

Table 10. Experiment 2 sessions overview

Session #	n	Invitation	Block Order
1	14	No self-selection	S/R
2	14	Self-selection	S/R
3	14	No self-selection	R/S
4	14	Self-selection	R/S
5	14	No self-selection	S/R
6	14	Self-selection	S/R
7	14	No self-selection	R/S
8	14	Self-selection	R/S

R= random-rank, S=skill-rank

5 Experiment 2 results

The industry profit is strictly positive in 76 (=79%), negative in twelve, and zero in the remaining eight out of 96 random-rank rounds. The average industry profit across random-rank rounds is 235.42 CZK. Out of 96 skill-rank rounds, the industry profit is strictly positive in 69 (=72%), negative in twelve, and zero in the remaining fifteen rounds. The average profit across skill-rank rounds is 207.29 CZK.

Industry profit and market entry

As in CL and Experiment 1, we first test for differences in the industry profit between the skill-rank and random-rank rounds (Hypothesis 1). In line with our Experiment 1, the matched pairs t-test does not detect a difference between profits in the random-rank rounds and skill-rank rounds (p-value = 0.192). Results from both of our experiments therefore differ from the result obtained by CL that the industry profit is significantly lower in the skill-rank rounds than in the random-rank rounds when using individual data from sessions 1-8.

Next we test whether there is a difference in industry profit between the random-rank rounds and skill-rank rounds within a session. The Wilcoxon matched-pairs signed-rank test p-values for each session are reported in Table 11 below. Except for the industry profit in session 3, where the profit in the random-rank rounds is statistically higher than in the skill-rank rounds, none of the other tests show that industry profits are statistically different in the random-rank rounds from the skill-rank rounds of the same session.

In summary, using participants of both genders we find almost no evidence of excess entry due to overconfidence when comparing behavior in the skill-rank rounds with the random-rank rounds.

Table 11. The Wilcoxon matched-pairs signed-rank test for the difference in industry profits and normalized entry rate between the random-rank and skill-rank rounds in Experiment 2 (within subjects)

Session #	Random-rank	Skill-rank	Wilcoxon matched-pairs signed-rank test (p-value)
1	3400	3200	0.874
2	2300	2300	1.000
3	2600	1000	0.021
4	1400	1600	0.937
5	2400	2200	0.873
6	3400	3000	0.411
7	4100	3700	0.297
8	3000	2900	0.966

Reference group neglect

Recall that Hypothesis 2 states that due to the reference group neglect the profit differential between the skill-rank and random-rank rounds in sessions with self-selection is larger than in sessions without self-selection. To test the hypothesis, we thus compare the profit differential in the selection sessions (2, 4, 6, and 8) with the profit differential in the sessions without self-selection (1, 3, 5, and 7). As in Experiment 1, the matched-pairs t-test does not support Hypothesis 2 (p-value=0.298) or CL's finding that the reference group neglect produces a significantly larger skill-random rank entry differential in sessions with self-selected participants than in sessions without self-selection (see Table 12 for a comparison of Experiment 2 results with CL.)

In summary, we do not find that the profit differential between skill-rank and random-rank rounds is larger in sessions with self-selection than without self-selection, pointing out that overconfidence does not increase with self-selection in a population composed of both genders.

Table 12. Comparison of CL's results and Experiment 2 results

	CL (sessions 1-8)	Experiment 2 (sessions 1-8)
Avg. profit random-rank	\$16.87	235.42 CZK
Avg. profit skill-rank	\$ -1.56	207.29 CZK
Matched pairs t-test	t=-7.43 p<0.001	t= 1.315 p= 0.192
Avg. profit without self-selection, random-rank	\$19.79	260.42 CZK
Avg. profit without self-selection, skill-rank	\$10.83	210.42 CZK
Avg. profit self-selection, random-rank	\$13.96	210.42 CZK
Avg. profit self-selection, skill-rank	\$ -13.13	204.17 CZK
Matched pairs t-test	t=-4.08 p<0.001	t= -1.053 p= 0.298

Expected profit differential in the skill-rank and random-rank rounds

Recall that excessive entry in the skill-rank rounds may not necessarily be due to overconfidence about one's skills, but due to underestimating how many participants will enter in total. In Experiment 2 we find that on average the number of forecasted entrants is 7.28 and 7.31 in the random-rank and skill-rank rounds, respectively. The actual number of entrants is on average 7.20 and 7.53 for the random-rank rounds and skill-rank rounds, respectively. The difference between the forecasted and actual number of entrants in the random-rank rounds is not statistically significantly different (Mann-Whitney test, p-value=0.833). In the skill-rank rounds this difference is not statistically significant either (p-value=0.674). In both random-rank and skill-rank rounds the participants' forecast converges to the actual number of entrants.

To separate overconfidence from incorrect estimates of others' entry, we calculate the differential between expected average profits in the random-rank rounds and in the skill-rank rounds using only the rounds in which participants entered (see Table 13). The mean differential $\pi_r - \pi_s$ is negative in sessions 1, 2, 4, 6, and 8. In sessions 3 and 5, the positive mean differential suggests the presence of overconfidence. In session 1, 41.7% of the participants expect to earn less in the skill-rank rounds than in the random-rank rounds. In sessions 3, 5, and 7, it is respectively 33.3%, 38.5%, and 36.4% of participants. In the self-selection sessions (2, 4, 6, and 8), the mean differential $\pi_r - \pi_s$ is always negative. In session 2 only 35.7% of participants expect to earn less in the skill-rank rounds than in the random-rank rounds. In session 4, 6, and 8 it is 57%, 53.8%, and 66.7%, respectively. The t-test does not detect a significant difference in the average differential of expected profits per person between the self-selection sessions and sessions without self-selection (p-value=0.223).

Table 13. The average differential in expected profits per entrant between the random-rank and skill-rank rounds in Experiment 2

Measure	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Total
Πr - Πs	-1.313	-1.679	0.069	-14.840	7.139	-9.071	-1.266	-6.172	-3.39
# of entrants with Πr - Πs<0 (percent)	5/12 (41.7%)	5/14 (35.7%)	4/12 (33.3%)	8/14 (57%)	5/13 (38.5%)	7/13 (53.8%)	4/11 (36.4%)	8/12 (66.7%)	46/101 (45.5%)
# of entrants with Πr <0 (percent)	0/12 (0%)	0/14 (0%)	0/14 (0%)	1/14 (7%)	1/14 (7%)	0/13 (0%)	0/12 (0%)	2/13 (15%)	4/106 (3.8%)

Logistic regression

Table 14 reports a fixed-effects logistic regression of the entry decision for Experiment 2. In line with the t-test results using the industry profits, we find that the entry rate in the skill-rank rounds is not higher than in the random-rank rounds. Our Experiment 2 regression results further show that participants in self-selection sessions do not enter more often than in sessions without self-selection. The coefficient associated with the interaction term of being male and self-selection is not statistically significant, meaning that in self-selection sessions males do not enter more often than females when a gender-neutral task is used to determine the ranking. (See Table 15 for a comparison of male and female entries in self-selection and no self-selection sessions. While it appears that males enter more often in self-selection sessions than in no self-selection sessions, the effect is not statistically significant at the 5% level with p-value = 0.071, chi-square test. There is no difference in the female entry rate; p-value = 0.702, chi-square test.)

Table 14. Experiment 2 fixed-effects logistic regression of the entry decision

Experiment 2	Model 1			Model 2		
Variable	**Coef (Robust Std. Err)**	**Marginal**	**z-statistic (p-value)**	**Coef (Robust Std. Err)**	**Marginal**	**z-statistic (p-value)**
Intercept	0.106 (0.135)		0.78 (0.434)	1.682*** (0.621)		2.71 (0.007)
C	0.043** (0.019)	0.010	2.20 (0.028)	0.038** (0.020)	0.009	1.93 (0.054)
$E(\pi_{ijt})$	-0.007*** (0.001)	-0.002	-6.87 (0.001)	-0.006*** (0.001)	-0.001	-6.64 (0.001)
Self-selection	0.014 (0.135)	0.018	0.11 (0.915)	0.016 (0.145)	0.018	0.11 (0.914)
Skill rank	0.165 (0.135)	0.030	1.22 (0.223)	0.170 (0.137)	0.029	1.23 (0.217)
Male	0.014 (0.135)	0.034	0.10 (0.919)	0.218 (0.147)	0.079	1.49 (0.137)
Age				-0.058*** (0.019)	-0.013	-3.03 (0.002)
Slovak/Czech				-0.388*** (0.123)	-0.088	-3.16 (0.002)
Siblings				0.044 (0.057)	0.010	0.77 (0.441)
Relative Income				0.025 (0.063)	0.006	0.40 (0.689)
City size				-0.034 (0.053)	-0.008	-0.64 (0.524)
Living with others				-0.175*** (0.052)	-0.040	-3.36 (0.001)
Money				0.001*** (0.001)	0.001	3.86 (0.001)
Finance study				0.004*** (0.001)	0.001	2.68 (0.007)
Rely				0.019 (0.019)	0.004	1.00 (0.317)
Law				0.011 (0.199)	0.003	0.06 (0.954)
Humanities				0.013 (0.165)	0.003	0.08 (0.935)
Natural Sciences				0.174 (0.164)	0.040	1.06 (0.288)
Other Social Sciences				0.621*** (0.231)	0.142	2.68 (0.007)
Engineering				-0.442** (0.208)	-0.101	-2.13 (0.033)
Medical Science				0.721*** (0.208)	0.164	3.46 (0.001)
Political				-0.816*** (0.244)	-0.186	-3.34 (0.001)
Self-selection *Skill rank	-0.108 (0.157)		-0.69 (0.492)	-0.110 (0.161)		-0.69 (0.493)
Self-selection *Male	0.232 (0.157)		1.47 (0.140)	0.231 (0.170)		1.35 (0.176)
Skill rank*Male	.029 (0.157)		0.18 (0.855)	0.027 (0.161)		0.17 (0.865)
Log-likelihood	-1799.63	Pseudo R2	= 0.03	Log-likelihood	-1737.71	Pseudo R2 = 0.07

Experiment 2 sessions 1-8, n=2688, standard errors are not clustered at the session level because of a small number of sessions.
Description of demographic variables: Slovak/Czech is 1 if the participant was Slovak or Czech, 0 otherwise; Siblings represents the number of siblings; Relative income represents whether the income is far below average, below avg., avg., above avg., or far above avg. (from 1 to 5). City size, Living with others, Money, and Finance study represent respectively the size of the city from 2000 to 100 000+; number of people in a household, the size of the monthly budget, and the share of monthly expenses one finances alone. Rely is a self-reported variable that indicates on a scale 1 to 9 how reliable is the information provided in a questionnaire, 9 being most reliable. Law, Humanities, Natural Sciences, Other Social Sciences, Engineering, Medical Science and Political Economy are dummy variables representing the fields of study, the omitted variable is Economics.
Run on StataSE 13.0. Robust standard errors used.
*, **, *** refer to statistical significance at the 10%, 5% and 1% levels, respectively.

The interaction terms Skill rank*Male and Skill rank*Self-selection in the logistic regression are not significant either, suggesting that the effect of skill-rank on the entry decision is not mediated by gender or self-selection. Finally, Experiment 2 regression results reveal the effect of the expected profit $E(\pi_{ijt})$ to be negative and significant, along the lines observed in CL and Experiment 1. We also find that with higher capacity the participants enter more often.

Table 15. Number of decisions in which participants entered the market

	Female	Male	Total
Self-selection	342 (out of 672)	378 (out of 672)	720
No Self-selection	349 (out of 672)	345 (out of 672)	694
Total	691	723	

Note: Due to the design of Experiment 2, in all sessions we have 1344 female decisions and 1344 male decisions. In no self-selection sessions we also have 1344 decisions and in self-selection sessions 1344 decisions as well.

Gender differences in performance in the addition task

The performance of our participants in the addition task, implemented in Experiment 2, provides support to Niederle and Vesterlund's (2007) finding that the task is gender-neutral. At the session level, there are no differences between the performance of males and females in sessions 1-7 (see Table 16). In session 8 females perform better than males. Pooling data from sessions 1-8 we find no statistical difference between the performance of males and females (p-value = 0.696). This result holds true for both the selection (p-value = 0.532) and no self-selection sessions (p-value = 0.304).

Unsurprisingly, we also find no overall difference in performance of participants in the self-selection and no self-selection sessions (even- and odd-numbered sessions, respectively; p-value = 0.979, Mann-Whitney test.) While this implies that also in Experiment 2 self-selection did not result in better performance in the addition task, the performance of participants is in line with their behavior in the market entry game where also observe no gender differences in entry decisions and no effect of self-selection on the entry rate.

Table 16. Average performance across sessions in Experiment 2

	Experiment 2		
	Male	**Female**	**M-W test**
Session 1	6.29	7.86	z = 0.71 p = 0.480
Session 2	8.29	8.43	z = -0.06 p = 0.949
Session 3	11.29	7.29	z = -1.28 p = 0.199
Session 4	10.29	8.86	z = -0.45 p = 0.653
Session 5	8.57	8.86	z = -0.39 p = 0.697
Session 6	9.14	7.71	z = -0.58 p = 0.563
Session 7	13.43	8.71	z = -1.35 p = 0.178
Session 8	6.14	12.57	z = 2.26 p = 0.024

The score in sessions 1-8 of Experiment 2 refers to the average number of correctly solved addition problems. There is no difference in performance between no self-selection (odd sessions) and self-selection sessions (even sessions), (Mann-Whitney, p-value= 0.979).

6. Discussion

CL propose a novel idea that business failures might be caused by overconfidence of those who decide to enter the market. In testing their conjecture, they find that males overconfident about their skills are more likely to enter the market and that overconfidence increases with self-selection. In our two experiments, we seek to replicate CL using a sample composed of both genders, making it a more conservative, and given the increased number of female managers and entrepreneurs observed in recent years also more timely test of their conjecture. Apart from including both male and female participants, our Experiment 1 closely follows CL design and employs mazes and a trivia quiz about sports and current events as tasks based on which the rank is determined. Experiment 2 employs a gender-neutral task, controls for the number of participants in a session ensuring that the number of males and females participants is equal, and makes the number of possible entrants common knowledge. The order of market capacities is also kept constant to make the skill-rank rounds directly comparable with the random-rank rounds.

Using both male and female participants, we are unable to replicate CL finding that the industry profit is lower and thus that there is more entry in the skill-rank compared to the random-rank rounds. While we find only very weak evidence of excess entry due to overconfidence in Experiment 1, we find no such evidence in Experiment 2. Furthermore, we are unable to replicate the finding that self-selection increases the entry rate in either of our two experiments. In Experiment 1 self-selection actually decreases the entry rate as self-selected females enter less often than self-selected males. In Experiment 2 we find no effect of self-selection on the entry rate. In both experiments we observe that when participants are expecting higher profits, they enter less often, in line with CL.

Regarding the implemented tasks, our data show that males perform significantly better on the quiz than females and that this effect is driven by self-selected males who perform significantly better than self-selected females. Females in self-selection sessions correctly anticipate their worse performance and rationally enter less often than self-selected males. We also observe that males are faster to complete the mazes task, however, this result is not statistically significant. Finally, in the addition task we observe similar performance of males and females; the addition task thus appears to be gender-neutral. Consistently with performance, we observe no gender differences in entry rates in Experiment 2, whether in self-selection or no self-selection sessions.

Apart from the above-described differences between our experiments and CL, there might be additional reasons for the diverging results. First and foremost, our Experiment 1 was conducted in New Zealand and Experiment 2 in Czech Republic as opposed to the U.S., some two decades later than the original CL study. Furthermore, we ran both of our experiments with undergraduate students whereas CL experiments also included two sessions with MBA students. Second, as seen in other areas of experimental research, for example dictator games, the results are often sensitive to a variety of seemingly innocuous variations (Cooper and Kagel, 2009). It is therefore possible that our procedures deviated from the original ones to some extent (e.g., the recruitment protocol for the experimental sessions) and these minor procedural differences have in turn affected the observed behavior. The lesson in all this is that while we are unable to replicate CL findings that there is more entry in the skill-rank compared to the random-rank round due to overconfidence, we find the self-selection effect to be sensitive to the gender and experimental conditions, especially the task used to determine the rank upon entry.

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology,* 49**,** 1621.

Alpert, M. & Raiffa, H. (1982). A progress report on the training of probability assessors.

Andersen, S., Ertac, S., Gneezy, U., List, J. A. & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics,* 95**,** 1438-1443.

Balafoutas, L. & Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science,* 335**,** 579-582.

Barber, B. M. & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly journal of Economics***,** 261-292.

Bénabou, R. & Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly journal of economics***,** 871-915.

Bénabou, R. & Tirole, J. (2003). Self-knowledge and self-regulation: An economic approach. *The psychology of economic decisions,* 1**,** 137-167.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of personality and social psychology,* 59**,** 960.

Buehler, R., Griffin, D. & Ross, M. (1994). Exploring the" planning fallacy": Why people underestimate their task completion times. *Journal of personality and social psychology,* 67**,** 366.

Cadsby, C. B., Servátka, M. & Song, F. (2013). How competitive are female professionals? A tale of identity conflict. *Journal of Economic Behavior & Organization,* 92**,** 284-303.

Camerer, C. & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American economic review***,** 306-318.

Christensen-Szalanski, J. J. & Bushyhead, J. B. (1981). Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human perception and performance,* 7**,** 928.

Christoffersen, S. & Sarkissian, S. (2002). Location overconfidence. *McGill University*.

Clayson, D. E. (2005). Performance overconfidence: Metacognitive effects or misplaced student expectations? *Journal of Marketing Education,* 27**,** 122-129.

Collins, S. M., James, D., Servátka, M. & Woods, D. (2017). Price-setting and attainment of equilibrium: Posted offers versus an administered price. *Games and Economic Behavior,* 106**,** 277-293.

Cooper, D. J. & Kagel, J. H. (2009). The role of context and team play in cross-game learning. *Journal of the European Economic Association,* 7**,** 1101-1139.

Daniel, K., Hirshleifer, D. & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance,* 53**,** 1839-1885.

Danková, K. (2015). Experiments on reciprocity, social comparison and overconfidence. *Mimeo: University of Canterbury.*

Deaves, R., Luders, E. & Luo, G. Y. (2009). An experimental test of the impact of overconfidence and gender on trading activity. *Review of Finance,* 13**,** 555-575.

Deck, C., Lee, J., Reyes, J. A. & Rosen, C. C. (2013). A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. *Journal of Economic Behavior & Organization,* 87**,** 1-24.

Dunning, D., Meyerowitz, J. A. & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of personality and social psychology,* 57**,** 1082.

Ertac, S. & Szentes, B. The effect of information on gender differences in competitiveness: Experimental evidence. 2011. TÜSİAD-Koç University Economic Research Forum working paper series.

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

Flory, J. A., Leibbrandt, A. & List, J. A. (2014). Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions. *The Review of Economic Studies,* 82**,** 122-155.

Forbes, W. (2009). *Behavioural finance*, John Wiley & Sons.

Gervais, S. & Odean, T. (2001). Learning to be overconfident. *Review of Financial studies,* 14**,** 1-27.

Glaser, M., Langer, T. & Weber, M. (2005). Overconfidence of professionals and lay men: Individual differences within and between tasks?

Glaser, M., Langer, T. & Weber, M. (2007). On the trend recognition and forecasting ability of professional traders. *Decision Analysis,* 4**,** 176-193.

Gneezy, U., Leonard, K. L. & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica,* 77**,** 1637-1664.

Gneezy, U., Niederle, M. & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics,* 118**,** 1049-1074.

Gneezy, U. & Rustichini, A. (2006). Executive versus teachers: Gender, competition and self-selection. *Working Paper. University of Chicago*.

Günther, C., Ekinci, N. A., Schwieren, C. & Strobel, M. (2010). Women can't jump?—an experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization,* 75**,** 395-401.

Healy, A. & Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal,* 121**,** 1192-1204.

Healy, P. & Moore, D. (2007). Bayesian overconfidence.

Hoelzl, E. & Rustichini, A. (2005). Overconfident: Do you put your money on it?*. *The Economic Journal,* 115**,** 305-318.

Kirchler, E. & Maciejovsky, B. (2002). Simultaneous over-and underconfidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty,* 25**,** 65-85.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C. & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational behavior and human decision processes,* 79**,** 216-247.

Kruger, J. (1999). Lake wobegon be gone! The" below-average effect" and the egocentric nature of comparative ability judgments. *Journal of personality and social psychology,* 77**,** 221.

Locke, P. R. & Mann, S. C. House money and overconfidence on the trading floor. AFA 2002 Atlanta Meetings, 2001.

Lundeberg, M. A., Fox, P. W. & Puncochar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of educational psychology,* 86**,** 114.

Malmendier, U. & Tate, G. (2005). Ceo overconfidence and corporate investment. *The journal of finance,* 60**,** 2661-2700.

Messick, D. M., Bloom, S., Boldizar, J. P. & Samuelson, C. D. (1985). Why we are fairer than others. *Journal of Experimental Social Psychology,* 21**,** 480-500.

Moore, D. A. & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes,* 103**,** 197-213.

Müller, J. & Schwieren, C. (2012). Can personality explain what is underlying women's unwillingness to compete? *Journal of Economic Psychology,* 33**,** 448-460.
Niederle, M., Segal, C. & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science,* 59**,** 1-16.
Niederle, M. & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The quarterly journal of economics,* 122**,** 1067-1101.
Presson, P. K. & Benassi, V. A. (1996). Illusion of control: A meta-analytic review. *Journal of Social Behavior and Personality,* 11**,** 493.
Price, C. R. (2011). Do women shy away from competition? Do men compete too much?: A (failed) replication. *Working Paper, University South Indiana*.
Roll, R. (1986). The hubris hypothesis of corporate takeovers. *Journal of business***,** 197-216.
Selten, R. & Güth, W. Equilibrium point selection in a class of market entry games. Games, Economic Dynamics, and Time Series Analysis, 1982. Springer, 101-116.
Simon, M. & Houghton, S. M. (2003). The relationship between overconfidence and the introduction of risky products: Evidence from a field study. *Academy of Management Journal,* 46**,** 139-149.
Soll, J. B. & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 30**,** 299.
Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica,* 47**,** 143-148.
Taylor, S. E. & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological bulletin,* 103**,** 193.
Vandegrift, D. & Yavas, A. (2009). Men, women, and competition: An experimental test of behavior. *Journal of Economic Behavior & Organization,* 72**,** 554-570.
Weinberg, B. A. (2009). A model of overconfidence. *Pacific Economic Review,* 14**,** 502-515.
Windschitl, P. D. & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 30**,** 198.
Wozniak, D., Harbaugh, W. T. & Mayr, U. (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics,* 32**,** 161-198.

Appendix A Instructions

Experiment 1, Session 1 and 2

INSTRUCTIONS

No Talking Allowed
Now that the experiment has begun, we ask that you do not talk. If you have a question after we finish reading the instructions, please raise your hand and the experimenter will approach you and answer your question in private.

Anonymity
The identity of the participants will not be revealed to other participants at any time during the experiment.

Show-up Fee
If you agree to participate in the experiment you will be given $5, which is yours to keep.
Structure of the Experiment
This experiment is computerized. If you have any problems entering your decision, please alert the experimenter. The experiment involves two sets of decisions. Each set of decisions consists of 12 rounds (i.e. 24 rounds in total). These two sets differ in how the rank is determined.
In the first 12 rounds your rank will be determined by your speed of finishing the mazes (as will be explained later). In the second 12 rounds your rank will be determined randomly. *(Session 2 instructions read: In the first 12 rounds your rank will be determined randomly. In the second 12 rounds your rank will be determined by your speed of finishing the mazes (as will be explained later)).*

In each round you are asked to decide whether to enter the market or not. In the beginning of each round the market capacity “c” for that round will be announced. You can think of “c” as the size of the market. You will also be informed about the number of entrants in the previous round.

Decision Making Task
In each round you start with $10.
If you decide not to enter the market, you earn nothing and lose nothing; your earnings for that round will be $10.
If you decide to enter the market, your payoff in each round will depend on your rank relative to the ranks of other participants who entered the market and on the capacity “c”.

If you entered the market
Your rank and the capacity for that round determine if you are a successful or unsuccessful entrant. If your rank is less than or equal to the capacity, then you are a successful entrant. If your rank is greater than the capacity, then you are an unsuccessful entrant. The unsuccessful entrant will lose the $10 (s)he was given in the beginning of that round. The payoffs of successful entrants as a function of “c” are shown in the table below.

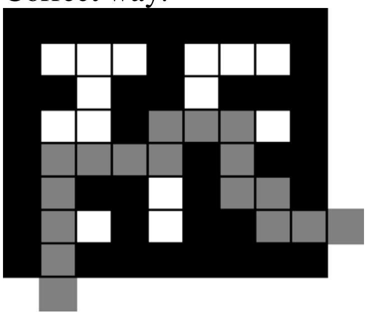
ONE of all 24 rounds will be chosen randomly and your rank and decision in this round will determine your payoff.

Rank	Capacity			
	2	4	6	8
1	33	20	14	11
2	17	15	12	10
3		10	10	8
4		5	7	7
5			5	6
6			2	4
7				3
8				2

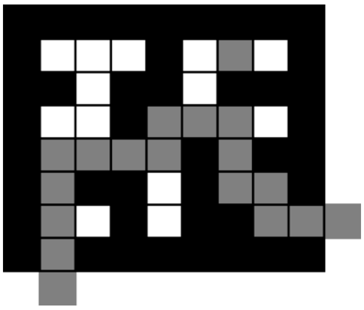
All participants will take part in both sets of decisions in the same order. In each round you will be also asked to estimate the number of people (including you) that you expect to enter the market in that round. If your estimation of the number of entrants is the same as the actual number of entrants in that round, additional $1 will be added to your payoff in that round.

The Maze
After you finish all 24 rounds in the decision making task, you will be given five mazes to solve. You need to find the shortest way from one end of the maze to another. If you have highlighted all the correct squares in the maze, the OK button will pop up. Click OK in order to continue. The participant, who finishes the mazes the fastest, will be ranked number 1. A participant, who is the second fastest, will be ranked number 2 and so on.

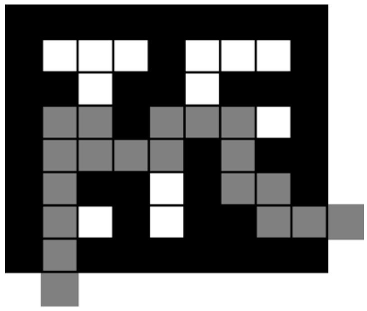
Correct way:



Incorrect way:



Incorrect way:



Example
Suppose "c" is 2 and four participants decide to enter the market. The entrant with rank number 1 earns $33 and the entrant with rank number 2 earns $17. The entrants with rank number 3 and number 4 lose $10, i.e. their payoff for that round will be 0.

Payment of Experiment Earnings
ONE of all 24 rounds will be chosen randomly and your rank and decisions in this round will determine your payoff.
All money will be paid to you in cash at the end of the experiment. Because your decision is private, we ask that you do not tell anyone your decision or your earnings either during or after the experiment. We also ask you to not gather near the lab after you receive your payment.

Are there any questions?

Experiment 1, Sessions 3-8

INSTRUCTIONS

No Talking Allowed
Now that the experiment has begun, we ask that you do not talk. If you have a question after we finish reading the instructions, please raise your hand and the experimenter will approach you and answer your question in private.

Anonymity
The identity of the participants will not be revealed to other participants at any time during the experiment.

Show-up Fee
If you agree to participate in the experiment you will be given $5, which is yours to keep.

Structure of the Experiment
This experiment is computerized. If you have any problems entering your decision, please alert the experimenter. The experiment involves two sets of decisions. Each set of decisions consists of 12 rounds (i.e. 24 rounds in total). These two sets differ in how the rank is determined.
In the first 12 rounds your rank will be determined randomly. In the second 12 rounds your rank will be determined by your score on a quiz (as will be explained later).
(Sessions 4, 6 and 8 read: In the first 12 rounds your rank will be determined by your score on a quiz (as will be explained later). In the second 12 rounds your rank will be determined randomly.)

In each round you are asked to decide whether to enter the market or not. In the beginning of each round the market capacity "c" for that round will be announced. You can think of "c" as the size of the market. You will also be informed about the number of entrants in the previous round.

Decision Making Task
In each round you start with $10.
If you decide not to enter the market, you earn nothing and lose nothing; your earnings for that round will be $10.
If you decide to enter the market, your payoff in that round will depend on your rank relative to the ranks of other participants who entered the market and on the capacity "c".

If you entered the market
Your rank and the capacity for that round determine if you are a successful or unsuccessful entrant. If your rank is less than or equal to the capacity, then you are a successful entrant. If your rank is greater than the capacity, then you are an unsuccessful entrant. The unsuccessful entrant will lose the $10 (s)he was given in the beginning of that round. The payoffs of successful entrants as a function of "c" are shown in the table below.

ONE of all 24 rounds will be chosen randomly and your rank and decision in this round will determine your payoff.

Rank	Capacity "c"			
	2	4	6	8
1	33	20	14	11
2	17	15	12	10
3		10	10	8
4		5	7	7
5			5	6
6			2	4
7				3
8				2

All participants will take part in both sets of decisions in the same order. In each round you will be also asked to estimate the number of people (including you) that you expect to enter the market in that round. If your estimation of the number of entrants is the same as the actual number of entrants in that round, additional $1 will be added to your payoff in that round.

The Quiz
After you finish all 24 rounds in the decision making task, you will be asked to participate in a multiple choice quiz. There are 30 sports & current events questions in the quiz, each question has only one correct answer. You will have 10 minutes to answer all questions. A participant with the most correct answers will be ranked number 1, etc. If two or more participants correctly answered the same number of questions, the ties will be broken by the shorter amount of time taken to answer all questions.
Example
Suppose "c" is 2 and four participants decide to enter the market. The entrant with rank number 1 earns $33 and the entrant with rank number 2 earns $17. The entrants with rank number 3 and number 4 lose $10, i.e. their payoff for that round will be 0.

Payment of Experiment Earnings
ONE of all 24 rounds will be chosen randomly and your rank and decisions in this round will determine your payoff.
All money will be paid to you in cash at the end of the experiment. Because your decision is private, we ask that you do not tell anyone your decision or your earnings either during or after the experiment. We also ask you to not gather near the lab after you receive your payment.
Are there any questions?

Experiment 2

INSTRUCTIONS

No Talking Allowed
Now that the experiment has begun, we ask that you do not talk. If you have a question after we finish reading the instructions, please raise your hand and the experimenter will approach you and answer your question in private.
Anonymity
There are 14 participants (including you) in this experiment. The identity of the participants will not be revealed to other participants at any time during the experiment.

Show-up Fee
If you agree to participate in the experiment you will be given 100 KČ, which is yours to keep.
Structure of the Experiment
This experiment is computerized. If you have any problems entering your decision, please alert the experimenter.
The experiment involves two sets of decisions. Each set of decisions consists of 12 rounds (i.e. 24 rounds in total). These two sets differ in how the rank is determined. In the first 12 rounds your rank will be determined randomly. In the second 12 rounds your rank will be determined by your score (number of correct answers) in an addition task (as will be explained later).
*(Sessions 1, 2, 5, and 6 read: In the first 12 rounds your rank will be determined by your score (*number of correct answers*) in an addition task (as will be explained later). In the second 12 rounds your rank will be determined randomly.*

In each round you are asked to decide whether to enter the market or not. In the beginning of each round the market capacity "c" for that round will be announced. You can think of "c" as the size of the market. You will also be informed about the number of entrants in the previous round.

Decision Making Task
In each round you start with 100 KČ.
- If you decide not to enter the market, you earn nothing and lose nothing; your earnings for that round will be 100 KČ.
- If you decide to enter the market, your payoff in that round will depend on your rank relative to the ranks of other participants who entered the market and on the capacity "c".

If you entered the market
Your rank and the capacity for that round determine if you are a successful or unsuccessful entrant. If your rank is less than or equal to the capacity, then you are a successful entrant. If your rank is greater than the capacity, then you are an unsuccessful entrant. The unsuccessful entrant will lose the 100 KČ (s)he was given in the beginning of that round. The payoffs (in KČ) of successful entrants as a function of "c" are shown in the table below.

ONE of all 24 rounds will be chosen randomly and your rank and decision in this round will determine your payoff.

Rank	Capacity "c"			
	c = 2	c = 4	c = 6	c = 8
1	330	200	140	110
2	170	150	120	100
3		100	100	80
4		50	70	70
5			50	60
6			20	40
7				30
8				20

All participants will take part in both sets of decisions in the same order. In each round you will be also asked to estimate the number of people (including you) that you expect to enter the market in that **round.** If your estimation of the number of entrants is the same as the actual number of entrants in that round, additional 10 KČ will be added to your payoff in that round.

Addition Task
After you finish all 24 rounds in the decision making task, you will be asked to participate in an addition task.
In this addition task, you will be asked to calculate the sum of five randomly chosen two-digit numbers. You will be given 5 minutes to calculate the correct sum of a series of these problems. You cannot use a calculator to determine this sum, however you are welcome to write the numbers down and make use of the provided scratch paper. You submit an answer by clicking the OK button with your mouse.

A participant with the most correct answers will be ranked number 1, etc. If two or more participants correctly answered the same number of questions, the ties will be broken by the shorter amount of time taken to answer the problems.
Example
Suppose "c" is 2 and four participants decide to enter the market. The entrant with rank number 1 earns 330 KČ and the entrant with rank number 2 earns 170 KČ. The entrants with rank number 3 and number 4 lose 100 KČ, i.e. their payoff for that round will be 0.

Payment of Experiment Earnings
ONE of all 24 rounds will be chosen randomly and your rank and decisions in this round will determine your payoff.

All money will be paid to you in cash at the end of the experiment. Because your decision is private, we ask that you do not tell anyone your decision or your earnings either during or after the experiment. We also ask you to not gather near the lab after you receive your payment.
Are there any questions?

Appendix B Control Questions

1. How much would you earn in a round if c=6, you entered and your rank was 5 among the entrants?

2. How much would you earn in a round if c=2, you entered and your rank was 4 among the entrants?

3. How much would you earn in a round if you decided not to enter the market?

4. How many rounds are there in total in this experiment?

Appendix C Auxiliary Tables

Table C1. Market capacity "c" values in CL and Experiment 1.

Round	Session 1	Session 2	Session 3-6	Session 7 and 8
1	2	8	2	4
2	4	4	6	2
3	8	2	4	6
4	6	6	4	8
5	4	4	2	6
6	2	2	6	4
7	8	8	4	2
8	6	6	6	8
9	4	4	2	6
10	6	2	6	4
11	8	8	4	2
12	2	6	2	8

Table C2. Experiment 1 fixed-effects logistic regression of the entry decision, sessions 3-8

	Model 1			Model 2		
Variable	**Coef (R Std. Err)**	**Marginal**	**z-statistic (p-value)**	**Coef (Robust Std. Err)**	**Marginal**	**z-statistic (p-value)**
Intercept	-0.050 (0.170)		-0.29 (0.770)	0.344 (0.497)		0.69 (0.489)
C	-0.038 (0.025)	-0.009	-1.56 (0.118)	-0.029 (0.026)	-0.006	-1.13 (0.259)
$E(\pi_{ijt})$	-0.006 (0.004)	-0.001	-1.50 (0.130)	-0.008* (0.004)	-0.002	-1.74 (0.081)
Self-selection	-0.802*** (0.167)	-0.042	-4.80 (0.001)	-1.234*** (0.200)	-0.106	-6.17 (0.001)
Skill rank	0.078 (0.177)	0.040	0.44 (0.659)	0.084 (0.182)	0.040	0.46 (0.644)
Male	-0.180 (0.173)	0.123	-1.04 (0.299)	-0.239 (0.192)	0.127	-1.24 (0.215)
Age				-0.075*** (0.012)	-0.016	-6.05 (0.001)
Non NZ Nationality				-0.137 (0.133)	-0.029	-1.03 (0.304)
Siblings				-0.033 (0.037)	-0.007	-0.89 (0.372)
Relative Income				-0.076 (0.072)	-0.016	-1.05 (0.291)
City size				0.506*** (0.063)	0.107	8.04 (0.001)
Living with others				0.057 (0.032)	0.012	1.81 (0.071)
Money				0.001 (0.001)	0.001	0.18 (0.858)
Finance study				0.003* (0.002)	0.001	1.77 (0.076)
Rely				-0.074*** (0.022)	-0.016	-3.38 (0.001)
Law				0.770*** (0.164)	0.162	4.69 (0.001)
Humanities				-0.056 (0.210)	-0.012	-0.26 (0.791)
Natural Sciences				0.261* (0.158)	0.055	1.65 (0.098)
Other Social Sciences				0.956*** (0.167)	0.202	5.73 (0.001)
Engineering				0.622*** (0.166)	0.131	3.76 (0.001)
Medical Science				-0.580 (0.413)	-0.122	-1.40 (0.161)
Self-selection *Skill rank	0.087 (0.184)		0.47 (0.636)	0.097 (0.190)		0.51 (0.610)
Self-selection*Male	1.048*** (0.186)		5.63 (0.001)	1.252*** (0.228)		5.48 (0.001)
Skill rank *Male	0.072 (0.179)		0.40 (0.686)	0.081 (0.187)		0.43 (0.666)
Log-likelihood	-1356.111	Pseudo R2	= 0.085	Log-likelihood	-1355.0	Pseudo R2 = 0.086

Sessions 3-8, n=2208, standard errors are not clustered at the session level because of a small number of sessions.
Task Score variable is included in the regression to approximate participants' perceptions of their own skills. Note that the timing of the experiment inhibits causality from performance to market entry. 24 participants entered the market zero times during the entire session.
Description of demographic variables: non New Zealander represents participants who are not from New Zealand; Siblings represents the number of siblings; Relative income represents whether the income is far below average, below avg., avg., above avg., or far above avg. (from 1 to 5). City size, Living with others, Money, and Finance study represent respectively the size of the city from 2000 to 100 000+; number of people in a household, the size of the monthly budget, and the share of monthly expenses one finances alone. Rely is a self-reported variable that indicates on a scale 1 to 9 the reliability of the information provided in a questionnaire with 9 being the most reliable. Law, Humanities, Natural Sciences, Other Social Sciences, Engineering, Medical Science are dummy variables representing the fields of study; the omitted variable is Economics.
Run on StataSE 13.0. Robust standard errors used.
*, **, *** refer to statistical significance at the 10%, 5% and 1% levels, respectively.