Organizing for Entrepreneurship: Field-Experimental Evidence on the Performance Effects of Autonomy in Choosing Project Teams and Ideas

Viktoria Boss (TUHH)
Christoph Ihl (TUHH)
Linus Dahlander (ESMT Berlin)
Rajshri Jayaraman (ESMT Berlin and University of Toronto)

Discussion Paper No. 204

November 29, 2019

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

ORGANIZING FOR ENTREPRENEURSHIP: FIELD-EXPERIMENTAL EVIDENCE ON THE PERFORMANCE EFFECTS OF AUTONOMY IN CHOOSING PROJECT TEAMS AND IDEAS

VIKTORIA BOSS, TUHH
CHRISTOPH IHL, TUHH
LINUS DAHLANDER, ESMT BERLIN
RAJSHRI JAYARAMAN, ESMT BERLIN & UNIVERSITY OF TORONTO

ABSTRACT. Organizations constantly strive to unleash their entrepreneurial potential to keep up with market and technology changes. To this end, they engage employees in practices like corporate crowdsourcing, incubators, accelerators or hackathons. These organizational practices emulate independent "green-field" entrepreneurship by relinquishing hierarchical control and granting employees autonomy in the choices of how to conduct work. We aim to shed light on two such choices that are fundamental in differentiating hierarchical from entrepreneurial modes of organizing work: (1) choosing projects ideas to work on and (2) choosing project teams to work with. Both of these choices are typically pre-determined in hierarchies and self-determined in entrepreneurship. We run a field experiment in an entrepreneurship course carefully designed to disentangle the separate and joint effects of granting autonomy in both choosing teams and choosing ideas compared to a pre-determined base case. Our results show that high autonomy in choosing implies a trade-off between personal satisfaction and objective performance. Self-determined choices along both dimensions promote subjective well-being in a complementary way, but their joint performance impact is diminishing. After ruling out alternative explanations related to differing project qualities and homophilic team choices, the detrimental performance impact of too much choice seems to be related to the implied cognitive burden and overconfidence.

1. INTRODUCTION

Over the last decade, a large body of literature has generated rich insights regarding the antecedents of network formation within companies (see e.g., Kleinbaum et al., 2013; Burt, 2004; Quintane Carnabuci, 2016). One of the issues researchers have explored is how formal structures shape the creation of ties in the workplace (see e.g., Biancani et al., 2014; McEvily Tortoriello, 2014; Clement Puranam, 2018). In so doing, they typically operate on the premise that individuals can choose their collaborators, to build networks that are consistent with their objectives (Burt, 2005). However, it is not clear that autonomy in collaborator choice is necessarily a good thing. For example, Ingram

Date: July, 2019.

Rajshri Jayaraman gratefully acknowledges inancial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119). She and Linus Dahlander are also grateful to NESTA IGL PR230 and the Peter Curtius Foundation for financial support of the field work. We thank seminar and conference participants at the University of Copenhagen, University of Rotterdam, Aarhus University, Max Plank, DRUID conference, Erasmus University, EURAM conference, INSEAD, Singapore Management University, as well as University College London for constructive feedback.

Morris (2007) show that even well-motivated and professionally sophisticated EMBA students struggle to go beyond their comfort zone during mixers, rarely creating ties that extend beyond friends of their friends.

The tension between the need to build workplace networks and potential problems associated with endogenous network choice, raises the intriguing possibility that placing constraints on collaborator choice may actually enhance outcomes. This possibility has real-world precedents. As Burt Merluzzi (2016, pg. 374) note, not all collaborations are a matter of choice: networks are characterized by a "mixture of exogenous assignment and endogenous choice, with the mix playing out differently for different individuals." An understudied question in this context is to what extent formal structures or constraints on collaborator choice may shape the creation of ties in the workplace (see e.g., Biancani et al., 2014; McEvily Tortoriello, 2014; Clement Puranam, 2018), and ultimately, workplace performance.

A related, albeit smaller, stream of literature in organization design has explored how tasks are allocated (Puranam et al., 2014). At one end, hierarchical organizations often assign tasks to workers and at another, more "horizontal" organizations tend to allow for some autonomy in task choice (Barley Kunda, 2001). For instance, in many "open" collaboration contexts, people who engage have the liberty to choose the tasks they want to engage with (Levine Prietula, 2014). Of course, hybrid models of work organization, whereby employees split their time between assigned projects and projects of their own choosing, are increasingly used in the business world. For instance, the company 3M was early to give their employees a day off their daily work to spur innovation within the company. An open question in this context, is whether autonomy in task choice actually improves workplace performance.

This paper brings these two strands of literature together by asking whether allowing for choice in team collaborators or ideas (or both) improves outcomes. This is, at its heart, an organizational design question. In particular, choice along one or both of these dimensions constitutes a fundamental difference between horizontal, entrepreneurial modes of organizing work, and more hierarchical ones. In the former, collaborators and ideas are typically a matter of choice. In the latter they are customarily assigned, although there is a growing move in these organizations to mimic "pure entrepreneurship to various degrees by relinquishing hierarchical control and granting employees more autonomy in the choices of how to go about the work. This includes the time allocation to idea development alluded to above, as well as by developing such practices as incubators, accelerators or hackathons.

Our aim is to shed light on whether allowing, or disallowing, autonomy of choice along these two dimensions, collaborators and ideas, improves performance. We do so by conducting a field experiment in an entrepreneurship course. (Following best practice, we filed a pre-analysis plan before commencing the experiment.) Our experimental setting has three key benefits. First, the entrepreneurship course itself is inspired by the lean startup movement. This means that we are basically looking at an early-stage entrepreneurship-type setting, with horizontal organizational structures, in which autonomy in idea and collaborator choice is the norm. Second, we have access to a large subject pool, of 900 students divided into 300 teams. This sample size furnishes us with the necessary power to make meaningful statistical inference. Third, the output we use for performance evaluation is designed to mimic the first stage of a venture capitalists dealflow (Bernstein et al., 2017; Brooks et al., 2014; Huang Pearce, 2015). In particular, the deliverable after this 14-week course is a pitch deck, which is evaluated by real venture capitalists, business angels and seasoned entrepreneurs. Pitch deck performance is obviously a noisy predictor of ultimate business success (McKenzie (2017) expertly

examines this later-stage outcome). Nevertheless, entrepreneurs generally need to pass this hurdle in order to get their business off the ground.

The experiment itself has a two-by-two design, whereby participants are randomly assigned to one of four treatment groups. The aim of this design is to disentangle whether allowing choice along one of both of the idea and collaborator dimensions, impacts performance. In the first treatment group *Choose team* participants can choose their own team but have to work on a randomly assigned pre-defined idea. In the second treatment group *Choose idea* participants can choose their own idea but are randomly assigned to a team. In the third treatment group *Choose both* participants can choose their own team and their own idea. Finally, in the baseline treatment *Choose neither* participants are randomly assigned to both a team and an idea. In using random assignment as the baseline because, as Clement Puranam (2018, pg. 3880) emphasize, "formal structure - even randomly selected and poorly enforced ones - are the benchmark that self-organizing systems must beat in order to replace traditional structures in our organizational landscapes, and beating this benchmark is not necessarily easy". This baseline also seemed to resonate with accelerator programs and entrepreneurs who, in informal interviews, highlighted what they saw as a large random component in team and task composition based, for example, on coincidental temporal or spatial location.

2. Prior Work

There has been substantial prior work on team composition and its effects on different facets of entrepreneurial performance (see e.g., Cooper Saral, 2013; Horwitz Horwitz, 2007; Steffens et al., 2012). Although entrepreneurship research has begun to pay attention to causal inference using experimental designs (see e.g., McKenzie, 2017; Clingingsmith Shane, 2017), there are few field experiments pertaining to entrepreneurial teams (Mao et al., 2016; Hasan Koning, 2019). This is likely due to the difficulty of finding a field setting with a large group of potential entrepreneurs one with a large enough sample of teams to yield statistically meaningful results. This difficulty is aggravated by the dearth of contexts, which are open to conducting field experiments.

In two important examples of studies that overcome these hurdles, Hoogendoorn and co-authors investigate the effect of team composition on performance based on exogenous assignment to teams. In a first natural field experiment, Hoogendoorn et al. (2013) study undergraduate students who start up a venture as part of their curriculum. This experiment assigns students to teams based on gender and finds that more gender-balanced teams outperform male-dominated teams. In a follow-up experiment, Hoogendoorn et al. (2017) assign students to teams based on measured cognitive abilities. They find that while average team ability has no effect on outcomes, variation in ability within teams does. In particular, teams at intermediate levels of ability dispersion outperform teams with very low and very high ability dispersions.

Our paper draws on a similar field setting but diverges from prior work in two important ways. First, while most, if not all, of the literature on entrepreneurial teams discussed below has explored the effect of team composition on performance, our focus is the effect of autonomy in choosing teams and ideas. Team composition, in this context, may be a consequence of choice – and we will explore this, too – but our focus is on the ramifications of autonomy in choosing teams and ideas on early stage entrepreneurial performance. Second, our focus is on early-stage team performance as captured by pitch decks examined at preliminary stages of a venture capitalists deal flow (Huang Pearce, 2015).

2.1. **Autonomy in Choosing Teams.** The network literature has built impressive inroads to understanding the interactions between formal and informal structures (Gulati Puranam, 2009; Kleinbaum et al., 2013; Biancani et al., 2014; McEvily Tortoriello, 2014). A key feature of this literature is that informal structures do not fully mimic formal structures. Rather, ties in informal networks emerge more organically. Burt Merluzzi (2016), in particular, maintain that networks are both a result through exogenous assignment *and* endogenous choice.

Autonomy in choosing collaborators is ubiquitous in the context of entrepreneurship. Beckman Burton (2008) show that the formation phase casts a long shadow, from the point of inception, to the direction a venture takes, and all the way to its performance in an IPO. Motivated by these findings, the entrepreneurship literature has gone to great lengths to understand which characteristics of teams improve entrepreneurial performance. Entrepreneurial teams are formed for a reason – entrepreneurs choose whom they would like to work with, frequently based on personal preferences or complementary skills. Indeed, endogenous team formation is probably more common in the context of entrepreneurship than it is in paid employment, where employees are often assigned to teams by supervisors based on project needs. Self-selection into teams makes it difficult to establish the causal effects of team composition on outcomes in real-world settings due to unobserved heterogeneity across teams (Hansen et al., 2015).

We seek to explore whether being able to choose your own team members – as opposed to being assigned to a team – has a bearing on performance. There are a number of reasons why team choice might matter for performance. When choosing team members, people may select those with whom they already have relationships, something that naturally fosters familiarity among team members. This affords the opportunity for internal cohesion and coordination of tasks between team members (Reagans et al., 2004). In addition to leveraging existing network ties, team choice facilitates matching based on complementary personality traits, skills or experiences (McPherson et al., 2001). Since the nature of relationships between team members determines the effectiveness with which a task can be handled (Brannon et al., 2013), self-selected teams may well perform better than randomly assigned teams that are, by construction, blind to such complementarities.

At the same time, prioritizing efficiency or performance over personal bonds in team formation is notoriously difficult. Ruef et al. (2003), for example, compared entrepreneurial team assembly mechanisms with randomly assembled teams that could have formed, but did not. They find that "[F]ounders of organizations appear more concerned with trust and familiarity, at this early stage, than with functional competence, leading to a 'competency discount' in founder recruitment" (p. 217). This suggests that choosing team members may have a downside in that people let social expediency take precedence over business efficacy. This tendency to gravitate toward familiar faces is not restricted to entrepreneurial settings. Ingram Morris (2007) traced participants in an EMBA student mixer using RFID tags, to examine who interacted with whom. They find that, despite participants self-declared motive of expanding their network by establishing new ties, the vast majority failed to do so. Instead, they interacted with their friends. Even within larger organizations with well-established organizational hierarchies, there is room for choice in professional interactions (Biancani et al., 2014; Kleinbaum et al., 2013).

While it has been shown that people who are given the option of choosing their own team members are possibly more inclined to choose people they know over strangers, the effect of such a choice on performance can have divergent effects. For one, valuable time and effort can be saved if one is already familiar with team members modus operandi. Time saved can be devoted instead to tackling

the task at hand. Moreover, working with a self-selected group can raise social expectations within the group, leading team members to exert more effort, which in turn translates to better performance.

2.2. **Autonomy in Choosing Ideas.** It has been popular in recent years to hail the people on a team as being more important than the idea per se. This notion, popularized by Ries (2011), has ignited interest among entrepreneurs and venture capitalists, even though rigorous empirical evidence supporting this notion is hard to come by. Although entrepreneurs tend to choose the idea they would like to work on, in many corporate settings tasks are assigned to teams. Even in the entrepreneurship setting, some companies have experimented with providing less autonomy to choose ideas.

Separating the effect of an idea from other factors is not trivial, as an idea shapes a new venture from its inception to its implementation. Some research looks at how an early version of ideas translates into market success (Astebro, 2003; Goldenberg et al., 2001), but this work often captures entrepreneurs who are well into the process. For instance, Goldenberg et al. (2001) study patents and case studies from published books to disentangle why some ideas are more likely to succeed. It seems likely, however, that many ideas die well before they merit patents or book mentions. Along the same lines, Kornish Ulrich (2014) emphasize the importance of separating raw ideas from the work that teams put into advancing an idea. They use data from Quirky, a crowdsourcing platform that organizes weekly tournaments for product ideas where some ideas are then selected and brought to market, to show that the raw idea has a large effect on the outcome.

A key feature of how organizations work is task division and task allocation, whereby managers tell subordinates what to do (see e.g., Puranam et al., 2014), providing limited opportunity for employees to choose the ideas they would like to work on. Because of inefficiency and coordination problems, organizations typically assign employees to tasks (Fama Jensen, 1983). At the same time, the ability of employees to self-select ideas is often seen as a major source of intrinsic motivation (Lovas Ghoshal, 2000). Using data from 23 different countries, Benz Frey (2008) find that self-employed people are much more satisfied than those who are employed. Striking a similar chord, Thompson (2000) notes that the more authority team members have to manage their own work, the more likely they are to be motivated and involved in their work.

This has lead companies such as Github, Valve, Oticon, and Google to experiment with giving people the mandate to self-select the ideas they want to work on. Having the autonomy to choose ideas to work on may thus increase performance by getting people more attached to the ideas they are working on. In their study of the Danish hearing aid company, Oticon, that allowed for employees to choose the ideas they wanted to work on, Lovas Ghoshal (2000) note: "this created an internal ecological environment where employees competed to join and stay with the most interesting projects, and the people responsible for a strategic initiative competed to attract and retain the most talented individuals." So, rather than being assigned ideas to work on, employees were now being given the choice to initiate new ideas to work on, and to join those ideas that they liked. In a critical examination of Oticon, Foss (2003) notes how employees suddenly had the authority to work "just like entrepreneurs in a market setting." While Oticon has since moved away from their complete bottom-up approach, the concept of employees coming up with their own projects is still intact in many entrepreneurial settings. However, not all entrepreneurial settings are like this. The Berlin-based Rocket Internet often decide to pursue an idea, then later assign or recruit a team to the task of exercising the idea.

2.3. **Autonomy in Choosing *Both* Teams *and* Ideas.** The jockey-versus-horse literature disentangles whether it is the team or the business that matters. Gompers et al. (2019) survey 885 institutional venture capitalists and find that they weigh the management team as more important than business-related characteristics such as product or technology. However, Kaplan et al. (2009), for instance, find that investors in startups should place more weight on the business ("the horse") than on the management team ("the jockey"). An important insight that stems from this literature is that teams and ideas have to be considered in tandem, and the difficulty associated with disentangling the relative importance of teams over the idea. Our reasoning has thus far considered having the ability to choose teams and ideas separately, but what is clear from this literature is that these can be pursued in tandem by giving people the ability to choose both teams and ideas. There are two competing explanations for how permitting more choice would affect performance.

One stream of literature on autonomy suggests that more freedom to explore options *positively* affects creative performance (Hackmann, 2002). This situation is often observed in real life, where potential entrepreneurs choose both their teams and their ideas. This environment provides the largest scope for selection, with few formalized structures that constrain choice. The freedom of choice that this affords should, if this argument is correct, translate into higher performance.

Another line of work offers a competing argument, that too much choice can be *detrimental* to performance. Simon (1955) famously argued that maximizing behavior, which requires assessing all possible choices, strains the limits of human cognition. He argued that in the presence of such psychic costs, satisficing – making an acceptable rather than an optimal choice – might well result in higher efficiency. Considering the choice of teams and ideas in tandem open up for new possibilities for thinking about how high early stage entrepreneurial performance comes about. In other words, one can go beyond comparing the extreme treatments of no choice with choice in both dimensions. For instance, conditional on choosing your collaborators, do you outperform those teams that have no choice in their ideas? Or conditional upon choosing ideas, would a chosen team outperform a randomly assigned one? If team choice results in better performance than team assignment, and idea choice results in better performance than idea assignment, it is natural to deduce that allowing for choice in both of these dimensions results in better performance than disallowing it in one or the other dimension. Theoretically, however, the answer to this question is not obvious. Team members who know each other may be too distracted by their social interactions to choose suitable ideas or even work efficiently Ingram Morris (2007). Similarly, they may choose ideas based on the personal preferences of group members rather than the business merit of these ideas. As a result, performance could potentially be enhanced by restricting choice along one, or both, of these dimensions.

3. The Experiment

3.1. **Setting.** In order to explore the effect of choice along these two dimensions – teams and ideas – on entrepreneurial performance, we conduct a natural field experiment. The experiment itself took place in the Business and Entrepreneurship course at a public university in Germany.[1] The university offers a three-year (six-semester) undergraduate degree in various engineering majors, with business as a minor in the curriculum. All undergraduate students at the university attend a mandatory, introductory, semester-long Business and Entrepreneurship course at some stage during their

[1] A pre-analysis plan was registered with the American Economic Association RCT registery. This is available upon request. Ethics approval was obtained from the university prior to the experiment.

undergraduate study. This course is offered each semester and our experiment took place in three successive semesters – referred to hereafter as "cohorts" – in 2016 and 2017. The Business and Entrepreneurship course is divided between lectures and tutorials. Lectures are conducted by a variety of different professors who cover their respective areas of expertise. All students attend the same lecture component but are divided into smaller groups for tutorials. The experiment took place in the context of the tutorial component of this course during which students, organized in entrepreneurial teams, worked on developing and pitching business ideas.

Semesters are 11 weeks long, and 90-minute tutorials take place once a week. Each tutorial is run by one teaching assistant (TA) and one experienced entrepreneur. We will refer to this TA-entrepreneur pair as "tutors" hereafter. In order to prevent experimental effects and in keeping with standard practice in natural field experiments, tutors were unaware of the experiment. In order to accommodate for the large number of students, the same tutorial is taught simultaneously in multiple rooms where each pair of tutors repeats the same tutorial twice on the same weekday to two different groups of students: once in an "early session" and once in a "late session". In cohort 1, there were four tutor pairs teaching a total of eight sessions. Cohorts 2 and 3 both had 10 tutor pairs, amounting to 20 sessions each. There were between six and nine student teams in each session.

In tutorials, tutors guide teams of students through the development of an entrepreneurial pitch deck. As we discuss later, this pitch deck forms the basis upon which entrepreneurial performance is evaluated. The pitch deck is aimed at hypothetical venture capitalists and closely resembles the document real entrepreneurs have to produce in order to get venture funding. It provides an in-depth understanding of the idea, its feasibility, target market, and projected revenue. Over the course of the semester, students develop their pitch deck and in the last session, they give a presentation of their final version.

3.2. **Interventions.** We implemented a two-by-two experimental design described in Table 1. In the *Team choice* treatment dimension, students were able to either choose their own team or they were assigned to a team of three members.[2] In the *Idea choice* treatment dimension, students were able to either choose their own idea or they were randomly assigned to work on one of 15 pre-determined problem statements (ideas). This comprised a brief description of the problem they were to address, along with an indication of how this problem could be resolved.

In the following, we denote the dummy variables that are equal to 1 if teams were allowed to choose their team member and ideas as *Team choice* and *Idea choice*, respectively. We use these dummy variables as independent variables in the regressions. Furthermore, we denote the treatment groups that result from the 2*2 factorial design of the two treatment dimensions as *Choose neither*, *Choose team*, *Choose idea* and *Choose both*, such that the *Choose both* treatment group, for example, is determined by *Team choice* and *Idea choice* both being equal to one. We use these treatment groups for descriptive and predictive group comparisions.

Students were randomly assigned to one of the four treatment groups. In order to ensure adherence to treatment assignment and minimize contamination or experimental effects, the four treatment groups were separated by time and space. Temporal separation was accomplished on the basis of early versus late time slots. The physical environment allowed for spatial separation since five tutorial classrooms

[2] In total, 96.5% of the teams had 3 members, 3.9% had 4, and 0.6% had 2. 4-member teams were necessary to balance out participant numbers not divisible by 3; 4 students decided they would rather work in 2 teams of 2 than 1 team of 4. Our results are robust to the exclusion of 2 and/or 4-member teams.

Table 1. Treatment overview

		Team choice	
		0 = "Assign"	1 = "Choose"
3***Idea choice**	0 = "Assign"	Baseline *"Choose neither"*	Treatment group 1 *"Choose team"*
	1 = "Choose"	Treatment group 2 *"Choose idea"*	Treatment group 3 *"Choose both"*

were in the west wing of the building, and five were in the east wing, with doors on either side of a stairwell separating the two wings. Appendix A gives an example of this temporal and spatial division for cohort 3, and Appendix B shows pictures for the interventions of all four treatments.[3]

Appendix C provides an overview of the substantive content of the 11 tutorial sessions. Depending on their (randomly assigned) treatment group, students were informed of the room and time their tutorial would take place at. So, for example, in cohort 3 described in in Appendix A, students who could choose their own team members were assigned to the early time slots, with those in the *Choose both* treatment in the west wing of the building with tutors 1 to 5, and those in the *Choose team* treatment in the east wing with tutors 6 to 10. Students with no choice in the team dimension were assigned to the late time slot, with the *Choose idea* treatments in the west wing with tutors 1 to 5 and *Choose neither* treatments in the east wing with tutors 6 to10. Early and late time slots were both on either side of noon, and students did not express any preference particular time slots. The east and west wings were, similarly, indistinguishable. In our regressions, we use fixed effects to capture systematic differences across tutors.

In week one, tutorials were dedicated to team formation. During the early session, all students were informed in their room by a tutor that they would have the opportunity to self-select into teams of three. Their choice was restricted to students in the same part of the building (i.e., people in the same treatment). All doors between the east and west wings of the building were closed and guarded by one of the tutors so as to ensure no mixing of treatments. Once all tutors had completed their introductions, students were released to the hallway to find suitable team members. Every two minutes a bell was sounded to encourage students to change partners and talk to new people. Once three people agreed to be in a team together, registered their team. Tutorial rooms were closed once they had reached a maximum capacity of nine teams, thus ensuring that all rooms had approximately the same number of teams. Team formation lasted for approximately half an hour in all three cohorts. Students in the treatments with team assignment (the late time slot in our example) were simply informed of their randomly assigned team members by their tutor in their designated room. We show later that the teams are balanced across treatments.

In week two, tutorials were dedicated to idea formation. Teams without choice in the idea dimension were situated in the east wing. At the beginning of the tutorial, tutors handed out the randomly

[3] Students have historically had no systematic preference for particular time slots. (The earliest slot starts at 9:45 a.m. and the last one ends at 5:15 p.m., so the hours are not inconvenient for full-time students.) The classrooms in the east and west wings are also comparable. Nevertheless, in order to avoid any systematic differences between treatment groups, we randomized the temporal and spatial allocation of treatment groups across cohorts, such that each treatment group was in at least 2 different locations and 2 different time slots over the course of our experiment.

assigned "idea" to each team.[4] Assigned ideas resemble a problem definition. They are sketched out on one page and steer the students in the direction of one possible solution. To deter students from copying from one another, we provided 15 different ideas and no idea was repeated within a given room; Appendix D provides an overview of all 15 ideas. Each of these ideas was extensively pre-tested to ensure both good quality and fit with the objectives of the course. Teams with choice in the idea dimension were situated in the west wing. During their second tutorial, students were guided through an idea generation process. The students were given freedom in their choice of ideas, provided it satisfied three criteria: (1) their idea had to solve a problem or add value; (2) there needed to be a potential target market; and (3) there had to be a way to generate revenue through the sale of the product or service. Ideas and their respective products or services could be adjusted over the course of the class.

The final deliverable for each team was a pitch deck comprising a maximum of 10 slides. Other than this, no formal restrictions were placed on the students.

3.3. **Evaluation.** In order to avoid experimenter effects and mimic real-life entrepreneurial situations, pitch decks were evaluated by practitioners. We had 40 such evaluators, who were practicing entrepreneurs, business angels or venture capitalists with, on average, more than 25 previous pitch deck evaluations and 0.875 founded companies per evaluator. Table 2 summarizes relevant characteristics pertaining to evaluators' expertise. In order to account for systematic variation across evaluators, each pitch deck was scored by three separate evaluators. To avoid negative selection – for example, attracting unsuccessful entrepreneurs in need of money – evaluators did not receive monetary compensation. Instead, the evaluation effort was organized as part of an entrepreneurship event. The event began with minimal social interaction and the distribution of the pitch decks with instructions for the evaluation. Names and pictures of the people in the team were redacted to ensure that pitch decks were evaluated based on merit alone. Evaluators were instructed not to talk to one another and were supervised in order to ensure that there was no communication. Once the evaluations were completed, three keynote lectures took place. At the end of the event, participants had the opportunity to network over an informal wine tasting.

Table 2. Evaluator expertise

	Average per Evaluator
# of startups founded	0.88
# of startups worked for	1.20
# of startups coached	2.51
# of startups funded with own money	0.65
# of startups funded as part of a venture capitalist decision	0.42
# of startups evaluated in the past	25.56
% self-reported qualification for evaluating the pitch decks	0.79

For the evaluation, all evaluators were given a verbal and a written explanation (see Appendix E) of the procedure. The event took place in a large office building where every evaluator had their own

[4]Only one team refused to work on their assigned idea. They were given a different idea from the original list of 15. Our results are robust to the exclusion of this team from the sample.

spot at a large table. The event began by collecting general information on the evaluator and their experience in the field was collected in a written survey. Following this survey, each pitch deck was individually evaluated. For this purpose, all pitch decks were printed out, with an evaluation form as the front sheet (see Appendix E, 3). Each evaluator was given a set of 23 or 24 pitch decks. Pitch decks were randomly assigned to evaluators using a customized assignment problem algorithm such that: (1) the same pitch deck was not evaluated by the same person twice; (2) the same assigned idea (of the 15 such ideas) was not evaluated by the same person twice, and (3) each person evaluated an equal number of pitch decks from all four treatments. We added a second heuristic approach by randomizing the order of the pitch decks each evaluator received. This avoids temporal effects such as being less generous as time progresses (Danziger et al., 2011).

Pitch decks were assessed on a variety of different criteria. Evaluators rated each pitch deck on a 7-point Likert scale, separately, according to: (1) novelty, (2) feasibility, and (3) market potential. In addition, evaluators had to rate (4) the project's likelihood of success and (5) the likelihood of inviting this team for a follow-up meeting both on a percentage scale ranging from 0 to 100 percent. Finally, once all pitch decks had been separately assessed, the evaluators had to allocate (6) a fictional investment budget of one million dollars across the entrepreneurship projects they had to evaluate. Evaluators could choose to spend some, all, or none, of their total (fictional) budget. This measure allowed for a direct comparison across pitch decks and closely mimiced the decisions of real-life venture capitalists. For this purpose, they were provided with stickers of fictional bills amounting to one million dollars. They then had to stick these bills onto on a designated field in the evaluation sheet of each pitch deck (see Appendix E).

These evaluation measurements constitute our performance indicators. The evaluation criteria were derived from previous works by Maxwell (2011) and Dean et al. (2006), and pre-tested. The evaluation concluded with a short exit survey asking the evaluators to judge the overall quality of the pitch decks, and self-assess their qualification to perform such evaluations on a 7-point Likert scale. We account for evaluator as well as order fixed effects in our analyses.

4. Data and Methods

4.1. **Sample.** A total of 940 students enrolled in this class over three cohorts – 173 students in the first cohort, 408 in the second, and 359 in the third. Of these 940 students, 27 dropped out, mainly for personal reasons. This left us with a final subject pool of 913 students constituting a total of 310 teams, 300 of which had three members, 12 with four members and 3 with two members.

To register for the class, participants had to fill out an online entry survey collecting information on their demographics, entrepreneurial experience, preferred team composition, and current skills (see Appendix F). After the final presentation in the last session, participants completed a written exit survey on their teamwork, satisfaction with the team and idea, new entrepreneurial skill development, and overall learning (see Appendix G). Students were unaware that they were part of an experiment. The entry survey has a 100 percent response rate while 94.5 percent took the exit survey; non-response was uncorrelated with treatment assignment.

Table 3 provides summary statistics for relevant student characteristics, including demographic information, familiarity with entrepreneurship, and educational background. Our sample is balanced in terms of predetermined student characteristics across the four treatments.

[b]
Table 3. Balance checks

	Average per Treatment			
	Choose neither	**Choose team**	**Choose idea**	**Choose both**
Personal Characteristics				
Female	0.309	0.248	0.239	0.293
German citizen	0.843	0.853	0.872	0.853
Entrepreneurial efficacy	40.593	39.731	39.829	39.483
Risk seeking	54.034	52.920	50.380	51.177
Confidence	64.047	65.063	66.026	64.289
Entrepreneurial career aspiration	0.148	0.105	0.145	0.121
Entrepreneurial experience	0.068	0.038	0.043	0.034
Number of entrepreneurs they know	4.886	4.668	5.145[a]	4.069
Number of preferred team members	2.407	2.664	2.543	2.556
Study Major				
General Engineering Science	0.102	0.113	0.09	0.086
Constructional and Environmental Engineering	0.148	0.130	0.132	0.116
Bio Process Engineering	0.030	0.042	0.043	0.017
Computer Science	0.119	0.147	0.145	0.147
Electrical Engineering	0.042	0.067[b]	0.021	0.052
Energy and Environmental Engineering	0.051	0.063	0.056	0.060
Computer Engineering	0.076	0.063	0.094	0.095
Logistics and Mobility	0.131	0.109	0.150	0.142
Mechanical Engineering	0.157	0.147	0.137	0.168
Mechatronics	0.025	0.029	0.013	0.026
Naval Architecture	0.034	0.034	0.034	0.030
Process Engineering	0.051	0.038	0.047	0.034

Note: Every treatment was compared to every other treatment using t-tests. a and b denote significance at the 5% level for t-tests for the comparison between treatments *Choose idea* vs. *Choose both* and *Choose team* vs. *Choose idea*, respectively.

In addition to balance checks regarding individual characteristics, we check the balance for ideas among those treatments which involve assigning ideas (i.e. *Choose team* and *Choose neither*). The results of a Pearson's chi-squared test ($\mathcal{X}^2$=5.768, p=0.972) show a balanced assignment of ideas, allaying fears that any difference in treatment effects across these two groups is driven by differences in the assigned idea *per se*. Furthermore, Pearson's chi-squared tests show that both the assignment of the four treatments to the 40 evaluators ($\mathcal{X}^2$=41.749, p=1), and the 15 assigned ideas to evaluators ($\mathcal{X}^2$=141.52, p=1) are random. Our previously described procedure further ensured that no evaluator assessed the same project or assigned idea multiple times.

4.2. **Variables. Dependent Variable**. Our dependent variable comes from the evaluation criteria described earlier, namely *novelty, feasibility, market potential, success potential, invitation probability, and investment*. Apart from information on the founding team (which was redacted on the pitch decks in order to avert bias) prior work describes these as driving funding decisions made by business

angels and venture capitalists (e.g. Carpentier Suret, 2015; Maxwell, 2011). The latter three variables are logged transformed to account for their skewness. All six variables are then standardized using z-scores to make them comparable. The z-scores are calculated by subtracting the baseline *Choose neither* group mean and dividing by the *Choose neither* group standard deviation. Thus, each performance indicator has mean 0 and standard deviation 1 for the baseline *Choose neither* group (for a similar approach see Kling et al., 2007).

All evaluators evaluated their assigned entrepreneurship projects in regard to at least three indicators. In fact, only very few evaluations were missing: one project-evaluator dyad only had three indicators evaluated, 11 dyads had four observed performance indicators and 43 out of 930 dyads had five out of six observed indicators, the remaining dyads had all six performance indicators evaluated. These missing values were imputed at the their respective overall means across all treatment groups (see Kling et al., 2007). Thus, we end up with 930 dyads, 310 projects evaluted by three evaluators each. All in all, the six performance indicators have a high internal consistency and reliability in measuring the project performance in our context (Cronbach's $\alpha = 0.85$).

To draw general conclusions about the experiments results, we therefore aggregate information over the multiple performance indicators. On the one hand, the dependent variable is defined as performance index to be the equally weighted average of z-scores of the performance indicators (see Kling et al., 2007). On the other hand, instead of averaging, we leave the six indicators separate and "stack" them as our dependent variable in the regressions (see Atkin et al., 2017). This method yields the same coefficients as an averaged performance index, but produces more conservative estimates (i.e., higher standard errors). Furthermore, the "stacked" regression appraoch still allows to detect differences in treatment effects for different performance indicators via interaction effects as will be explained below.

Independent Variables. Our main aim is to uncover how the autonomy to choose team members and/or business ideas affects entrepreneurial performance. Our baseline regression specification therefore includes one dummy variable *Team choice* that is equal to 1 if teams are allowed to choose members and 0 otherwise. In terms of our 2X2 design, it is equal to 1 for both the *Choose team* and *Choose both* treatment groups. The other dummy variable *Idea choice* is equal to 1 if teams are allowed to choose ideas and 0 otherwise. Hence, it is equal to 1 for both the *Choose idea* and *Choose both* treatment groups. In addition, we include the interaction term as product of both dummies. It is equal to 1 in the *Choose both* treatment group, when there is autonomy in choice on both dimensions, and zero otherwise. This model is structurally equivalent to a model with three dummies for the respective treatment groups *Choose team*, *Choose idea*, and *Choose both* (with *Choose neither* excluded as baseline). We prefer this "factorial" specification, however, as it allows us to assess more directly whether choice along the team and idea dimensions are complements or substitutes in terms of entrepreneurial performance.

4.3. **Estimation.** Estimation of treatment effects is straightforward in the context of randomized experiments. At its most basic, it involves a comparison of means across the different treatments. The unit of our empirical analysis is the project-evaluator dyad, of which there are 930 (310 projects $\times$ 3 evaluations per project). This permits the inclusion of evaluator fixed effects, which is important because prior research has found that there is large variation in the assessment of early stage ideas (see Boudreau et al., 2016, for a similar approach). By the means of fixed effects, we further control for potential differences in (1) the order in which a specific project was evaluated by a specific evaluator,

(2) the course cohort in which a specific team participated, and (3) the tutor pair that instructed a specific team.[5]

We rely on OLS estimates for our main results due to its ease of interpretation. In order to account for potential correlation among observations from the same team or evaluator, all standard error estimates are clustered at the team and evaluator level. The formal specification of our performance regression model is the following:

$$
\begin{aligned}
y_{ij} = \; & \beta_0 + \beta_1 * (\textit{Team choice})_i + \beta_2 * (\textit{Idea choice})_i + \\
& \beta_3 * (\textit{Team choice})_i * (\textit{Idea choice})_i + \gamma_i + \delta_i + \zeta_{ij} + \eta_j + \epsilon_{ij}
\end{aligned} \tag{1}
$$

where y_{ij} denotes the performance evaluation of team i as assessed by evaluator j, and *Team choice* and *Idea choice*, as described above, are dummy variables denoting autonomy of choice in team members or ideas, respectively. The product of both dummies *Team choice * Idea choice* is supposed to capture the interaction effect of both treatments. γ_i is a vector of cohort fixed effects; δ_i is a vector of tutor fixed effects; ζ_{ij} is a vector of evaluation order fixed effects; η_j is a vector of evaluator fixed effects; and ϵ_{ij} is a random error component.

Our parameters of interest are the $\beta's$. In the simultaneous presence of the interaction term, β_1 captures the effect of autonomy in *Team choice* as the difference between the treatment groups *Choose team* and *Choose neither* as baseline. β_2 captures the effect of autonomy in *Team choice* as the difference between the treatment groups *Choose idea* and the baseline *Choose neither*. β_3 captures the effect of autonomy in both *Team choice* and *Idea choice* and is effectively a difference-in-differences. It captures how the just defined performance effects of *Team choice* and *Idea choice* change when teams also have autonomy along the respective other dimension of choice. In other words, it examines potential complementarities between autonomy in idea and team choice.

Instead of implementing equation 1 separately for each performance indicator, we specify three regression models that aggregate the information of the performave indicators in different ways: Firstly, with the data gathered in "long" format, we regress the stack of all 6 performance indicators on interactions of the treatment variables with dummies referring to each performance indicator. The coefficients on these interactions provide the treatment effects separately for each performance indicator. We also include interactions of the performance indicator dummies with the fixed effects of cohorts, tutors, order and evaluators (see Atkin et al., 2017). The estimated coefficients are identical to those from separate regressions for each performance indicator, but using this specification the standard errors clustered account for project- and evaluator-level correlations within and across performance indicators.

Secondly, in the stacked regression described before, we restrict the coefficients on the treatment effects to be identical across all six performance indicators. This specification yields the same coefficients as an averaged performance index, but produces more conservative estimates (i.e., higher

[5] In each cohort, each (pair of) tutors was responsible for two tutorial sesions – meaning they had students from two different treatments. Tutor pairs were deliberately switched every cohort. Nevertheless, given that 23 tutors were responsible for tutoring 310 teams, tutors were not perfectly balanced across treatments. We therefore control for tutor fixed effects in our estimations.

standard errors) (see Atkin et al., 2017). Thirdly, as a less conservative comparison, we run regression equation 1 with a performance index as dependent variable that is based on an equally weighted average of the six performance indicators.

5. Results

5.1. **Main Results.** Table 4 shows OLS parameter estimation results. Model 2 shows that the *Team choice* treatment marginally outperforms the baseline treatment by an increase of 0.206 standard deviations ($p = 0.084$ compared to $p = 0.015$ in Model 3). Model 1 shows that this treatment effect is significant for the three performance indicators *Market Potential*, *Success Potential* and *Invitation Probability*. The highest impact refers to *Invitation Probability* with an increase of 0.274 standard deviations if teams can be chosen. In terms of unstandardized and unlogged values, this implies an increase of 5.2 percentage points from 7.4% average invitation probablity in the control group *Choose neither* to 12.6% in the *Choose team* treatment condition – a relative increase of 69.8%.

The *Idea choice* treatment is highly significant ($p < 0.01$) across all model specifications and performance indicators (except for *Feasability*). On average the effect implies an increase of 0.411 standard deviations across all performance indicators. *Idea choice* has the highest impact on *Investment* with an increase of 0.526 standard deviations if ideas can be chosen and are not assigned. In terms of monetary value, this implies an increase of 2,913 (fictional) Dollar from 1,239 Dollar average investment in the control group *Choose neither* to 4,152 Dollar average investment in the *Choose idea* treatment condition – a relative increase of 235%.

Lastly, the interaction term between the *Team choice* and the *Idea choice* treatment is significantly negative with an average decrease of -0.508 standard deviations across all performance indicators ($p = 0.017$ in Model 2 as compared to $p = 0.001$ in Model 3). Model 1 shows that this interaction effect is at least marginally significant at $p < 0.1$ for all performance indicators (except for *Newness* with $p = 0.174$). The largest effect size of this interaction effect is associated to *Success Potential* with a decrease of -0.612 standard deviations. The negative interaction effect implies that there is a performance penalty if teams get to choose *both* their team member *and* their ideas. The autonomy of choosing *both* teams *and* ideas do not provide complementary benefits. The benefits of choosing *either* do not even add up. Teams that get to choose *both* teams *and* ideas actually suffer from a significant performance penalty compared to the situation of choosing *either*.

To assess how the *Choose both* treatment condition compares to the *Choose neither* condition, we have to substract the penalty of choosing *both* from the individual benefits if choosing *either* teams *or* ideas. In the case of *Success Potential*, the net benefit corresponds to $0.220 + 0.409 - 0.612 = 0.017$ standard deviations. Expressed in percentge points for the probability of success, this means an increase of only 0.29 percentage points from an average success probability of 11.82% in the *Choose neither* control group to 12.12% in the *Choose both* treatment group. Hence, the individual benefits of choosing *either* teams *or* ideas are entirely offset by negative complementarities between both choices.[6]

[6] The differences between the *Choose neither* control group and the *Choose both* treatment group are not significant on average across all performance inidcators. This becomes evident from regression analyses that instead of the factorial specification use a specification with three dummy variables each referring to a treatment group. These results are available upon request.

Table 4. Regression results for the treatment effects on performance

	Stacked Performance Indicators with Interactions	Stacked Performance Indicators	Performance Index
	(1)	(2)	(3)
Team choice (β_1)		0.206*	0.206**
		(0.119)	(0.085)
Idea choice (β_2)		0.411***	0.411***
		(0.130)	(0.099)
Team choice * Idea choice (β_3)		−0.508**	−0.508***
		(0.213)	(0.151)
Team choice (β_1) * Newness	0.088		
	(0.143)		
Team choice (β_1) * Feasibility	0.225		
	(0.155)		
Team choice (β_1) * Market Potential	0.222**		
	(0.112)		
Team choice (β_1) * Success Potential	0.220**		
	(0.106)		
Team choice (β_1) * Invitation Probability	0.274***		
	(0.104)		
Team choice (β_1) * Investment	0.208		
	(0.156)		
Idea choice (β_2) * Newness	0.507***		
	(0.152)		
Idea choice (β_2) * Feasibility	0.213		
	(0.159)		
Idea choice (β_2) * Market Potential	0.417***		
	(0.117)		
Idea choice (β_2) * Success Potential	0.409***		
	(0.127)		
Idea choice (β_2) * Invitation Probability	0.394***		
	(0.130)		
Idea choice (β_2) * Investment	0.526***		
	(0.169)		
Team * Idea choice (β_3) * Newness	−0.325		
	(0.239)		
Team * Idea choice (β_3) * Feasibility	−0.542*		
	(0.279)		
Team * Idea choice (β_3) * Market Potential	−0.525***		
	(0.197)		
Team * Idea choice (β_3) * Success Potential	−0.612***		
	(0.215)		
Team * Idea choice (β_3) * Invitation Probability	−0.556***		
	(0.191)		
Team * Idea choice (β_3) * Investment	−0.491*		
	(0.292)		
Performance Indicator FE	Yes	Yes	No
Cohort FE	Yes	Yes	Yes
Mentor FE	Yes	Yes	Yes
Evaluation Order FE	Yes	Yes	Yes
Evaluator FE	Yes	Yes	Yes
Observations	5,580	5,580	930
R^2	0.379	0.376	0.408

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors reported in parentheses are clustered at the team and evaluator level. β_1 is a dummy variable equal to 1 for treatments with choice in teams; β_2 is a dummy variable equal to 1 for treatments with choice in ideas; β_3 is the interaction term of the two. Model 1 is a stacked regression showing treatment effects separately for all six performace indicators. Model 2 is a stacked regression restricting treatment effects to be equal across performance indicators. The interaction effects allowing the treatment effects to vary across performance indicators are jointly significant at a marginal level ($F(5052, 15) = 1.5461; p = 0.08061$). Model 3 collapses the performance indicators into a single averaged index in an unstacked regression.

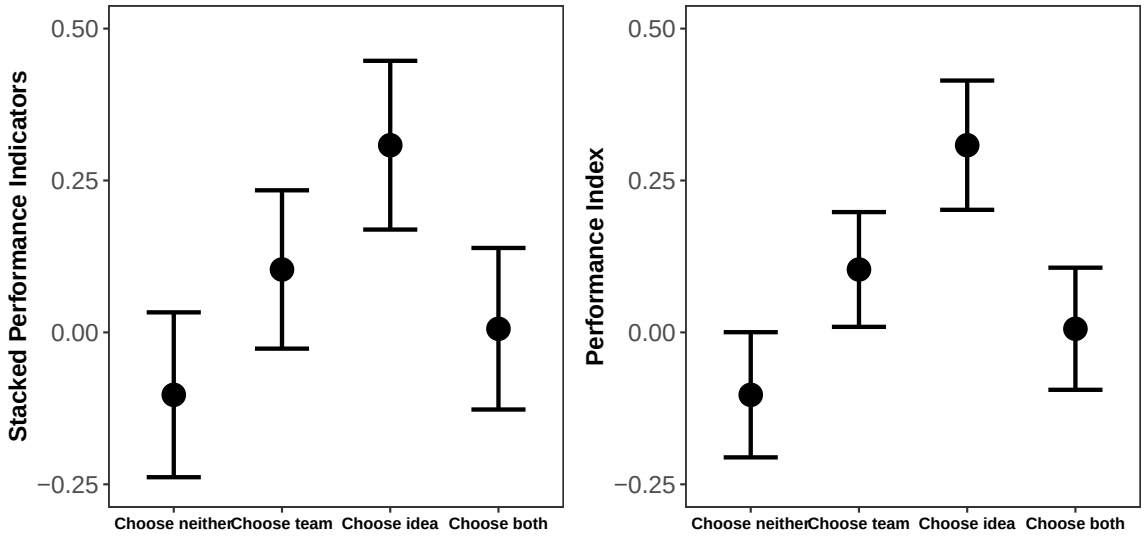


Figure 1. Conditional treatment group means (predicted from models 2 and 3 with 95% confidence intervals)

Figure 1 visualizes the predicted treatment effects from the OLS regression models 2 and 3 along with their 95% confidence intervals. These conditional means mirror those in the raw data, but conditioning on cohort, tutor, evaluation order, and evaluator fixed effects throughout our analysis increases precision and further limits potential confoundedness of our treatment effect estimates. It again becomes evident that Model 3 using the performance index yields less conservative, i.e. larger confidence intervals.

Both models show that the baseline treatment with no choice of either team members or idea (*Choose neither*) results in the lowest performance. Being able to choose the idea but not one's team (*Choose idea*), by contrast, consistently results in the highest performance. The performance of treatment groups in which teams could choose their members but not the idea (*Choose team*) is somewhere in between the *Choose neither* and *Choose idea* treatment groups. Teams that could choose along both dimensions (*Choose both*) have performance outcomes statistically similar to the baseline (*Choose neither*) condition.

5.2. **Specification Checks.** In this section, we report the results of specification checks, which address three main concerns. First, it is reasonable to wonder whether assigned ideas performed worse than endogenously generated ones simply because assigned ideas were of poor quality. In order to allay this concern, we conducted an independent quality check of assigned and chosen ideas. In particular, we randomly selected 15 of the generated ideas and had each written up twice as a paragraph by four different externals in order to mimick the idea sketches handed out for the 15 assigned ideas. In a second step, another person combined these paragraphs into one, making sure that none of the write ups were biased. We then took the 30 idea paragraphs and had them evaluated on Mechanical Turk along the same performance criteria used in our main analysis (*Newness*, *Market potential* and *Success potential*; leaving out *Investment* because judging a possible investment should be done on

the basis of a pitch deck rather than the simple idea itself). A total of 20 evaluators each evaluated 10 (out of the 30) ideas, leaving us with 200 total evaluations of 15 randomly chosen endogenous ideas and the 15 exogenous ideas. For our analysis, we include a *Choose idea* dummy together with evaluator fixed effects to account for systematic differences between evaluators and we cluster the standard errors at the evaluator and idea level (idea fixed effects are not possible since our key predictor, *Choose idea*, is invariant within ideas). We do not find any statistically significant differences between ideas that were assigned or chosen for all three criteria (results not shown). This indicates that our findings are not driven by systematic differences in idea quality across assigned versus chosen idea treatments, *per se*.

Second, although we report OLS estimates in our main results for ease of interpretation, the fact that there are limited dependent variables does technically warrant different estimation methods. We deal with this by estimating Tobit and ordered Probit models. The results, presented in Appendix **??**, confirm our main results.

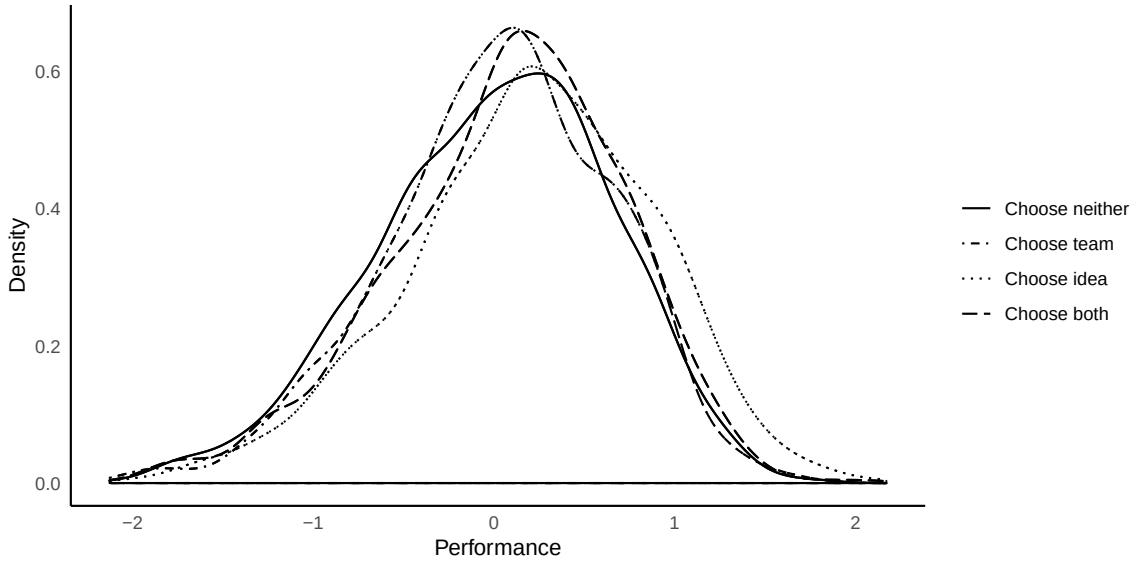


Figure 2. Kernel density estimates of predicted performance per treatment group (based on model 2)

Finally, a common concern in the entrepreneurship and innovation literature is the fact that results might be driven by the tails of the distribution, meaning that very few highly innovative results skew the results in one direction (or likewise that few terrible outcomes diminish positive results). In order to address this concern, we use heteroscedastic regressions to analyze the dispersion. The results confirm that both the *Team choice* treatment ($p < 0.0448$) and the *Idea choice* treatment ($p < 0.0357$) also increase the variance in outcomes next to the means. The interaction effect of choosing both teams and ideas, however, marginally reduces outcome dispersion ($p < 0.0516$). We also analyzed potential outcome dispersion effects visually. We hence estiamtes the kernel density of predicted performance (measured with the stacked performance indicators) based on regression model 2. Figure 2 shows kernel densities of predicted performance differentiated according to treatment group. These results suggest that *Choose idea* treatment increases the average performance by shifting the

action to the right-tail of the distribution, while *Choose idea* and *Choose both* seem to avoid very bad performance outcomes.

5.3. **Understanding how autonomy in choice matters.** Choice is likely to affect outcomes through four main channels: (i) team composition, (ii) team work, and (iii) satisfaction, and (iv) confidence. Team composition refers to who is on a team. Team work refers to how work is organized and distributed in a team. Satisfaction and confidence are subjective assessments. They pertain, respectively, to whether team members are personally satisfied with their team member and (separately) project idea, and whether they are confident in their project's business value. In this section, we explore whether there is any systematic variation along these four channels across the different choice treatments. We measure these channels in a variety of different ways; variable descriptions and summary statistics are provided in Table 5. Underlying each of these channels is a recognition that people make choices for a reason and these choices, in turn, have ramifications for intermediate inputs that determine a team's ultimate performance.

Team composition. Homophily is commonly manifested in endogenous team formation. Given the choice, people tend to work with others who they know, or with whom they share similar characteristics (McPherson et al., 2001). Table 6 confirms that homophily in terms of observed characteristics is manifested in our experimental treatments which allowed for *Team choice*. These teams are significantly less diverse with respect to gender (column 2), study majors (column 4) and tenure at the university (column 5); there is roughly a third less variation in terms of gender and majors and two-thirds less variation in majors in endogenously formed teams relative to the no choice baseline. The fact that there are no statistically significant differences between the treatments in terms of age (column 1) and nationality (column 3) reflects the fact that there is little underlying variation in these variables from this sample of undergraduate German university students.

Table 5. Variable definitions for the different channels with descriptive statistics

width=totalheight=keepaspectratio

Variable name	Variable definition	Data source	Mean	SD	Min	Max
Team composition						
Homophily						
Age (SD)	Standard deviation (SD) of age	Baseline survey	1.62	1.18	0.00	7.78
Gender (SD)	Blau Index (= probability that two randomly selected individuals in the team are of the same gender) calculated by $1 - \sum p_i^2$ where p is the proportion of group members in a given category and i is the number of different catego of the feature across all groups.	Baseline survey	0.22	0.22	0.00	0.50
Nationality (Count)	Number of distinct nationalities / number of team members	Baseline survey	0.47	0.20	0.25	1.00
Majors (Count)	Number of distinct study majors / number of team members	Baseline survey	0.72	0.23	0.33	1.00
Tenure (SD)	SD of semester of study	Baseline survey	0.99	0.87	0.00	4.04
Sorting						
Ability (SD)	SD of final exam grade (distinct from entrepreneurial performance)	Exam	0.10	0.06	0.00	0.32
Self-confidence (SD)	SD of self-reported confidence in own performance in class	Baseline survey	0.69	0.43	0.00	2.08
Aspiration (SD)	SD of self-reported aspiration to engage in entrepreneurship in the future	Baseline survey	20.74	11.41	0.00	57.17
Efficacy (SD)	SD of self-reported entrepreneurial efficacy	Baseline survey	0.89	0.48	0.00	2.59
Network						
Ties (Count)	Number of wishes from the baseline survey that were fulfilled during team formation / number of team members	Baseline survey & Administrative records	0.08	0.17	0.00	1.00
Acquaintances (Mean)	Mean number of team members participants were previously acquainted with	Endline survey	0.50	0.63	0.00	2.00
Teamwork						
Attendance (Mean)	Mean course attendance	Administrative records	9.77	1.02	5.50	11.00
Attendance (SD)	SD of course attendance	Administrative records	1.02	1.17	0.00	6.35
Attrition (Dummy)	Dummy variable = 1 if a team member quit the project	Administrative records	0.08	0.27	0.00	1.00
Collaboration (Mean)	Mean of team members' rating of whether project tasks were solved together vs. separately among team members	Endline survey	5.30	1.07	1.00	7.00
Workload (Total)	Sum of team members' stated workload - 100	Endline survey	33.86	30.95	-17.00	156.00
Workload (SD)	SD of team members' stated workload	Endline survey	12.43	8.59	0.00	35.20
Satisfaction						
Team satisfaction (Mean)	Mean of team members' identification with team	Endline survey	5.25	1.06	2.33	7.00
Idea satisfaction (Mean)	Mean of team members' identification with idea	Endline survey	4.56	1.12	1.50	7.00
Confidence						
Perceived potential (Mean)	Mean of team members' ratings of project's success potential on a scale from 0-100%	Endline survey	48.32	20.30	1.67	100.00
Commitment (Mean)	Mean of team members' likelihood to pursue project on a scale from 0-100%	Endline survey	12.25	14.23	0.00	80.00

Table 6. Team composition

	Dependent variable:								
	Age	Gender	Nationality	Majors	Tenure	Ability	Self-confidence	Aspiration	Efficacy
	(SD)	(SD)	(Count)	(Count)	(SD)	(SD)	(SD)	(SD)	(SD)
	OLS	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Team choice (β_1)	−0.020	−0.085**	−0.053	−0.269***	−0.447***	−0.006	−0.339***	−4.242**	−0.165
	(0.096)	(0.037)	(0.052)	(0.026)	(0.173)	(0.011)	(0.103)	(1.741)	(0.101)
Idea choice (β_2)	−0.333***	−0.013	−0.059	−0.063	0.062	−0.021	−0.206**	−2.283	−0.162
	(0.104)	(0.032)	(0.048)	(0.040)	(0.175)	(0.013)	(0.097)	(2.065)	(0.105)
Team choice * Idea choice (β_3)	0.384***	0.039	0.096	0.154***	−0.128	0.018	0.556***	3.018	0.246
	(0.118)	(0.055)	(0.096)	(0.048)	(0.290)	(0.020)	(0.180)	(3.210)	(0.159)
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mentor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	310	310	310	310	310	269	310	310	310
R^2	0.060	0.125	0.057	0.226	0.171	0.064	0.075	0.078	0.046

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors reported in parentheses are clustered at the tutor level.

In columns 6 through 9 of Table 6, we continue by examining whether homophily is manifested in the team choice treatments through assortative matching along potentially unobserved dimensions including ability, self-confidence, self-reported aspirations to engage in future entrepreneurship, and self-reported entrepreneurial efficacy. The evidence for homophily along these dimensions in the team choice treatments is weak. Teams that had autonomy in *Team choice*, i.e. the treatment groups *Choose team* and *Choose both*, are more homogeneous with regard to entrepreneurial aspiration than the *Choose idea* and the baseline *Choose neither* treatment group ($p < 0.01$). Teams with autonomy in *Team choice* are also more homogenous in terms of self-confidence, but only if they did not also have autonomy in *Idea choice*. There is no evidence of assortative matching along the lines of ability or entrepreneurial efficacy.

Prior network ties amd acquaintance with people often guide and possibly bias a person's choice of team members. We capture this through two separate measures. First, in the baseline survey, we asked people to name up to five people whom they would like have on their entrepreneurial team; this list represents people with whom they have prior frienship and/ or working ties. Following treatment assignment, we calculate the number of such preferred ties that was actually realized. Second, in the endline survey, we asked participants to indicate how many of their team members they knew beforehand. This allows us to capture whether, given the choice, people are more likely to gravitate to someone with whom they have a tie than a stranger.

The last two columns of Table 6 show that network ties are significantly more prevalent and stronger in team choice relative to team assignment treatments. Teams that had autonomy in *Team choice*, i.e. the treatment groups *Choose team* and *Choose both*, are 80% more likely to be acquainted and are 94% more likely to belong to a team containing a member from their baseline (5-person) wish list. The network map in Figure 3 shows this graphically. We find far more realized wishes (ties) for participants in the two team choice treatments (indicated in red) than for assigned team treatments in cohort 3. The same pattern holds in the other two cohorts (not shown).

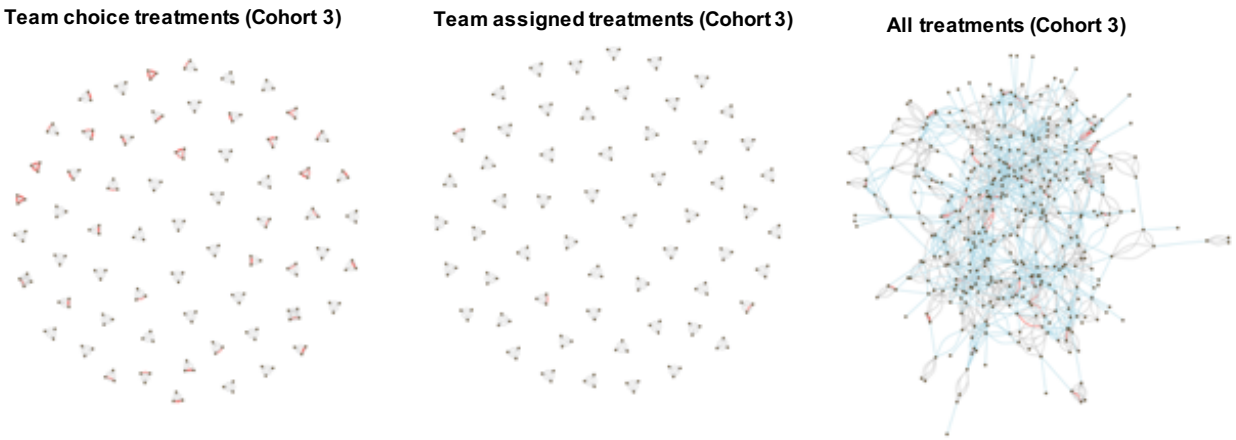


Figure 3. Network graphs of collaboration ties (wanted / not wanted; realized / unrealized)

Table 6 underscores three main findings concerning the effect of team choice on team composition. First, subjects in our experiment tend to exhibit homophily in terms of observed characteristics. Since others have shown that lack of diversity along these dimensions is associated with worse entrepreneurial outcomes, this tendency may well hinder team performance (e.g., Cooper Bruno, 1977; Horwitz Horwitz, 2007; Steffens et al., 2012). Second, there is little compelling evidence of assortative matching in the team choice treatments along unobserved dimensions – such as ability or proclivity for entrepreneurship – that would *a priori* contribute to entrepreneurial performance. On the one hand, this is not surprising given that these attributes are difficult to observe. On the other it does suggest that, at least in our setting, that teams may not be formed on the basis of such complementary performance-related attributes. Finally, our results indicate that when given the choice, people tend to work with friends at the possible expense of choosing the best possible collaborators. While prior ties can be positive in reducing frictions among people (Reagans, 2011), they can also lead to complacent teams being formed between people who are friends rather than agents aspiring to achieve a goal (Ruef et al., 2003).

Teamwork. Table 7 explores how members contribute to teamwork after teams have been assembled. We measure this in four ways. First, course attendance, which is a proxy for effort. Second, dropout which measures participation in the team effort. Third, self-reported collaboration within the team. Fourth, the distribution of the project workload. The results are mixed. Columns 1 to 2 indicate that there are no significant differences across treatments in terms of commitment to the team, as measured by class attendance and drop out. At the same time, there is evidence of a more collaborative spirit among teams with autonomy in *Team choice*, i.e. the treatment groups *Choose team* and *Choose both* exhibit a higher average collaboration quality than than the *Choose idea* and the baseline *Choose neither* treatment group ($p < 0.05$). Teams in the *Choose team* treatment group also overstate the combined workload to a significantly lower extend, where there is no effect on the workload diversity among team members. In sum, Table 7 provides some evidence of there being more collaborative teamwork in treatments which allow for team choice. Unfortunately, this does not seem to readily translate into improved team performance in these treatments.

Satisfaction. If choice along both the idea and team dimensions is detrimental to performance, then why is this extent of choice the prevalent model in entrepreneurial teams? One possibility is that entrepreneurial teams face a trade-off between personal satisfaction and team performance. Table 8 explores this by examining treatment effects with respect to teams' satisfaction with their members (Column 1) and the idea (Column 2). It shows that choice lends satisfaction along the dimension where choice is afforded. *Team choice* has a significant effect on team satisfaction ($p < .01$) but has, as expected, no effect on the satisfaction with the idea. Autonomy in *Team choice* leads to 13% higher team satisfaction when comparing the treatment groups *Choose team* and *Choose neither* and a 8% higher team satisfaction when comparing the treatment groups *Choose both* and *Choose neither*. *Idea Choice* has a significant effect on idea satisfaction ($p < .01$) but has, as expected, no effect on the satisfaction with the team. Autonomy in *Idea choice* leads to 18% higher idea satisfaction when comparing the treatment groups *Choose idea* and *Choose neither* and a 22% higher idea satisfaction when comparing the treatment groups *Choose both* and *Choose neither*. If the penalty for greater choice is entrepreneurial performance, these results suggest that the reward is personal satisfaction. This points to a potential rationale for why so many entrepreneurial projects develop organically in terms of both team composition and idea generation, even if the odds of them failing are higher as a result.

Table 7. Teamwork

	Dependent variable:					
	Attendance (Mean)	Attendance (SD)	Attrition (Dummy)	Collaboration (Mean)	Workload (Total)	Workload (SD)
	OLS	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)	(5)	(6)
Team choice (β_1)	−0.113	−0.329	−0.088	0.530**	−8.932**	−1.907
	(0.390)	(0.362)	(0.060)	(0.254)	(3.676)	(1.220)
Idea choice (β_2)	−0.090	0.098	0.017	0.213	−1.495	−1.202
	(0.363)	(0.324)	(0.053)	(0.269)	(4.036)	(1.277)
Team choice * Idea choice (β_3)	0.394	−0.045	−0.019	−0.366	7.386*	1.234
	(0.658)	(0.605)	(0.095)	(0.491)	(4.414)	(1.736)
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Mentor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	310	310	310	309	310	310
R^2	0.101	0.080	0.124	0.135	0.114	0.080

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors reported in parentheses are clustered at the tutor level.

Table 8. Satisfaction and confidence

	Dependent variable:			
	Team Satisfaction	Idea Satisfaction	Perceived Potential	Commitment
	(Mean)	(Mean)	(Mean)	(Mean)
	OLS	*OLS*	*OLS*	*OLS*
	(1)	(2)	(3)	(4)
Team choice (β_1)	0.769***	0.110	−6.518	−0.201
	(0.220)	(0.178)	(4.720)	(1.877)
Idea choice (β_2)	0.079	0.859***	−0.087	−0.925
	(0.236)	(0.183)	(4.397)	(2.401)
Team choice * Idea choice (β_3)	−0.403	0.226	20.242**	9.047***
	(0.402)	(0.284)	(7.908)	(2.553)
Cohort FE	Yes	Yes	Yes	Yes
Mentor FE	Yes	Yes	Yes	Yes
Observations	309	309	309	309
R^2	0.101	0.320	0.210	0.175

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors reported in parentheses are clustered at the tutor level.

Confidence. The last two columns of Table 8 gauge the effect of choice on the subjective assessment of a project's potential to succeed (Column 3) and team members' stated intent to pursue the project in the future. These two measures capture team member's confidence in their project quality. The results indicate that only the interaction term between *Team choice* and *Idea Choice* is significant ($p < 0.05$ and $p < 0.01$, respectively). This means that only teams in the the *Choose both* treatment group are significantly more confident in the project compared to the baseline *Choose neither* treatment group. Subjective assessment of project potential is 23% higher, and stated intent to pursue the project in the future is 39% higher than in the baseline no choice treatment. This juxtaposition between greater satisfaction and higher confidence on the one hand and poorer performance on the other hand is striking. It suggests that a key ingredient in the paradox of choice may be that it generates overconfidence.

In sum, we find that allowing for team choice results in more homogeneity in team composition along observable dimensions, without the benefits of assortative matching along unobserved dimensions which are likely to foster team performance. At the same time, given the option, individuals are more likely to choose team members who belong to their network. In other words, homophily in team composition generates teams whose members resemble one another and probably like one another, without the benefits of complementary skills or traits. This is manifested in the fact that endogenously selected teams do not seem to display markedly more efficient team work. This raises the question of why endogenously formed teams and organically developed ideas are the rule rather than the exception. A simple, but plausible, rationale suggested by our data is that giving people choice makes them happier in the dimension along which choice is afforded. At the same time, choice on both dimensions generates confidence, or perhaps even overconfidence given the negative effects on performance relative to other choice treatments.

6. Discussion

Autonomy in choosing tasks and collaborators has been at the heart for organizational theorists for decades (Puranam et al., 2014). Inside traditional organizations, there is usually little room for employees to choose their collaborators and tasks they want to work on. Entrepreneurship, in contrast, often provide more room for founders to self-assemble teams and the ideas they want to execute. Between the two polar extremes, there are also different hybrids allowing some choice (Baumann et al., 2018). This paper sought to address this question through a field experiment, whose major appeal is that it allows for causal inference while having sufficiently large sample sizes to allow for meaningful statistical inference: our experiment involved 900+ participants and over 300 teams. In order to address our main research question, subjects in a two-by-two experimental design were randomly assigned to one of four treatments in which they (i) chose both their team and the idea to pursue; (ii) chose their own team members but not the idea; (iii) chose their own idea but not their team; or (iv) chose neither their team nor the idea.

We find that freedom of choice on either or both of these dimensions at least weakly dominates having no choice on either of these dimensions. In other words, autonomy in choice matters and generally increase pitch-deck performance. At the same time, how much choice and what kind of choice one has also matters. We find that allowing for choice on the idea dimension and disallowing it on the team dimension results in the best performance outcomes. This result is striking because it indicates that, once you allow teams the freedom to choose their own idea, a randomly assigned team would perform better than one where people choose their collaborators.

Our results show that high autonomy in choosing implies a trade-off between personal satisfaction and objective performance. Self-determined choices along both dimensions promote subjective well-being in a complementary way, but their joint performance impact is diminishing. After ruling out alternative explanations related to differing project qualities and homophilic team choices, the detrimental performance impact of too much choice seems to be related to the implied cognitive burden and overconfidence. Our results yield implications for the literature on organizational design of entrepreneurial orientation in firms (Foss et al., 2015) as we highlight the importance and limits of two organizational design levers that can help to promote entrepreneurial culture in firms. On the other hand, we speak to the behavioral economics literature on entrepreneurship (Benz Frey, 2008) as we experimentally validate procedural utility to arise from the freedom of choice in entrepreneurship, but also to potentially limit entrepreneurial performance.

We argue that choice is likely to affect outcomes through four main channels: (i) team composition, (ii) team work, (iii) team satisfaction, and (iv) confidence. With respect to team composition, we find that endogenously formed teams more homophilous on many observable dimensions and tend to rely on pre-existing networks. However, homophily is not exhibited in terms of unobserved characteristics that would seem conducive to entrepreneurial success. This, in turn, is borne out in our finding that subjects in the team choice treatments do not seem to be working more effectively as a team.

If endogenously formed teams perform worse than assigned teams, why is the former so frequently observed in practice? Our results point to two answers. The first is that, conditional on being assigned to an idea, self-selected teams perform at least weakly better than assigned teams: some choice is better than none. The second is that choice increases satisfaction. People are more satisfied with their idea, when they came up with it, and they are more satisfied with their team when they get to decide

who is in it. Unfortunately, conditional on being able to choose ones idea, the latter does not translate into higher performance. People may be happier when they choose their team members, but when they also get to choose their ideas, their confidence in the merits of their project is not commensurate to its actual performance. A stylized fact in entrepreneurship research is that entrepreneurs often persist in their endeavors despite having low average returns (albeit being over-represented in the tails of the distribution) (Hamilton, 2000). One explanation is that entrepreneurs prefer autonomy and are thus willing to forsake part of their income (Moskowitz Vissing-Jørgensen, 2002). Our result support this conjecture by showing that autonomy in the choice of both teams and ideas makes people in this treatment happier; convinces them that their idea has potential; and makes them more likely to pursue it in the future.

Choosing ideas matters for performance: teams that endogenously choose the idea they wanted to work on consistently showed higher performance. (We show in robustness checks that this is not driven by the possibility that assigned ideas per se were of low quality.) In some sense this echoes conventional wisdom: the thought bubble over peoples heads in comic strips typically contains a light bulb, and not a bunch of team members. Nevertheless, this is a noteworthy finding in light of recent discussions from practitioners who hail the team as the most important ingredient of entrepreneurial success (see e.g., Blank, 2007). It also substantiates the work of Kornish Ulrich (2014), who emphasize the relative importance of ideas generated by teams in very early stages of entrepreneurship.

Our findings indicate that some choice is better than no choice, but there is such a thing as having "too much choice". There are two main reasons for why choice along both the idea and team dimension may be detrimental to team performance. First, giving participants the freedom to choose along multiple dimensions might make the task of developing a business plan – a task that is highly complex to begin with – even too complex. Overwhelmed participants might not have been able to efficiently allocate their resources, thus slacking in overall performance. Second, there is likely to be a tradeoff between personal benefits and business performance. We find that self-selected teams are more cohesive, homogeneous, and happier. Yet, it stands to reason that the focus of a high performing team should not be on satisfaction, but rather on efficiency, characterized by effective working relationships with a clear division of labor (Hoogendoorn et al., 2017). Developing a business idea with people you do not go for after-work drinks with likely shifts the focus on getting the job done, leading to a higher performance.

The findings in this paper are pertinent to the "professionalization of entrepreneurship", particularly through incubator and accelerator programs. Most accelerators and incubators give aspiring entrepreneurs choice on both the idea and team dimensions. Occasionally – for example, in Berlins Rocket Internet – they are afforded choice in neither dimension. Our results indicate that there may be a more effective third way, namely, assigning teams but letting them freely choose their idea. This precludes assortative matching and may detract from personal happiness generated from social interactions. But it is likely to generate the kind of environment in which better ideas can flourish and be translated into more successful entrepreneurial team performance.

6.1. **Limitations and future research.** The experimental design and main results presented in the paper followed a pre-analysis plan, which was pre-registered at the Social Science Registry of the American Economic Association. This set up enables to draw causal inference without being susceptible to p-hacking. In addition, conducting our field experiment among this group of business students

permitted a large enough sample size to allow for statistically meaningful inference. Still, this paper has several limitations.

Our experimental subjects comprised students in an entrepreneurship class who were tasked with producing an entrepreneurial pitch deck. This context mirrored that used by Hoogendoorn et al. (2013, 2017), who conduct field experiments with roughly 50 teams of business students. The rationale for this field setting was to have over 300 teams a sample size large enough to permit meaningful statistical inference. While our experimental design mimicked a real entrepreneurial setting by using real investors and entrepreneurs in the evaluation of these pitch decks, a drawback of this setting is that the subjects themselves were students, and not real-world entrepreneurs. This places natural limitations on the potential generalizability of our findings.

Another challenge is the potential for contamination across treatment groups. We sought to minimize this by carefully separating experimental conditions across time and space. In addition, neither tutors nor students were aware that they were participating in an experiment. Nevertheless, we cannot rule out the possibility that there was some communication across teams in different treatment groups.

Our approach compares teams who are permitted choice in choosing team members and ideas with those who do not have that choice. The control group in this context consists of randomly formed teams and randomly allocated ideas. In real life, organizations do not form teams at random – they form teams in the hope of achieving superior performance. There are several ways in which such teams could form, such as through maximizing functional and social diversity (Lamaze Van Knippenberg, 2014) or allowing team members to have some prior connections (Reagans et al., 2004). An interesting future line of research would be compare different such baselines rather than randomly assigned teams.

REFERENCES

Astebro T (2003) The return to independent invention: Evidence of unrealistic optimism, risk seeking or skewness loving? *The Economic Journal* 113(484):226–239, ISSN 00130133, URL http://dx.doi.org/10.1111/1468-0297.00089.

Atkin D, Khandelwal AK, Osman A (2017) Exporting and Firm Performance: Evidence from a Randomized Experiment*. *The Quarterly Journal of Economics* 132(2):551–615, ISSN 0033-5533, URL http://dx.doi.org/10.1093/qje/qjx002.

Barley SR, Kunda G (2001) Bringing Work Back In. *Organization Science* 12(1):76–95, ISSN 10477039, URL http://dx.doi.org/10.1287/orsc.12.1.76.10122.

Baumann O, Bergenholtz C, Frederiksen L, Grant RM, Köhler R, Preston DL, Shane S (2018) Rocket Internet: organizing a startup factory. *Journal of Organization Design* 7(1), URL http://dx.doi.org/10.1186/s41469-018-0037-2.

Beckman CM, Burton MD (2008) Founding the future: Path dependence in the evolution of top management teams from founding to IPO. *Organization Science* 19(1):3–24, ISSN 1047-7039, URL http://dx.doi.org/10.1287/orsc.1070.0311.

Benz M, Frey BS (2008) The value of doing what you like: Evidence from the self-employed in 23 countries. *Journal of Economic Behavior and Organization* 68(3-4):445–455, ISSN 01672681, URL http://dx.doi.org/10.1016/j.jebo.2006.10.014.

Bernstein S, Korteweg A, Laws K (2017) Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment. *Journal of Finance* 72(2):509–538, ISSN 15406261, URL http://dx.doi.org/10.1111/jofi.12470.

Biancani S, McFarland DA, Dahlander L (2014) The Semiformal Organization. *Organization Science* 25(5):1306–1324, ISSN 15265455, URL http://dx.doi.org/10.1287/orsc.2013.0882.

Blank SG (2007) *The four steps to the epiphany*. ISBN 1411601726, URL http://dx.doi.org/22.

Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science* 62(10):2765–2783, ISSN 0025-1909, URL http://dx.doi.org/10.1287/mnsc.2015.2285.

Brannon DL, Wiklund J, Haynie JM (2013) The varying effects of family relationships in entrepreneurial teams. *Entrepreneurship: Theory and Practice* 37(1):107–132, ISSN 10422587, URL http://dx.doi.org/10.1111/j.1540-6520.2012.00533.x.

Brooks AW, Huang L, Kearney SW, Murray FE (2014) Investors Prefer Entrepreneurial Ventures Pitched by Attractive Men. *Proceedings of the National Academy of Sciences of the United States of America* 111(12):4427–4431, ISSN 10916490, URL http://dx.doi.org/10.1073/pnas.1321202111.

Burt RS (2004) Structural Holes and Good Ideas. *American Journal of Sociology* 110(2):349–399.

Burt RS (2005) *Brokerage and Closure: An Introduction to Social Capital* (New York, United States: Oxford University Press), 1 edition, ISBN 978-0-19-924914-5, URL http://dx.doi.org/https://doi.org/10.1093/esr/jcm030.

Burt RS, Merluzzi J (2016) Network Oscillation. *Academy of Management Discoveries* 2(4):368–391, URL http://dx.doi.org/10.5465/amd.2015.0108.

Carpentier C, Suret JM (2015) Angel group members' decision process and rejection criteria: A longitudinal analysis. *Journal of Business Venturing* 30(6):808–821, ISSN 08839026, URL http:

//dx.doi.org/10.1016/j.jbusvent.2015.04.002.

Clement J, Puranam P (2018) Searching for Structure: Formal Organization Design as a Guide to Network Evolution. *Management Science* 64(8):3879–3895, ISSN 15265501, URL http://dx.doi.org/10.1287/mnsc.2017.2807.

Clingingsmith D, Shane S (2017) Training aspiring entrepreneurs to pitch experienced investors: Evidence from a field experiment in the United States. *Management Science* ISSN 0025-1909, URL http://dx.doi.org/10.1287/mnsc.2017.2882.

Cooper AC, Bruno AV (1977) Success among high-technology firms. *Business Horizongs* 20(2):16–22, URL https://papers.ssrn.com/sol3/papers.cfm?abstract{_}id=1505206.

Cooper DJ, Saral KJ (2013) Entrepreneurship and team participation: An experimental study. *European Economic Review* 59:126–140, URL http://mpra.ub.uni-muenchen.de/25144/.

Danziger S, Levav J, Avnaim-Pesso L (2011) Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America* 108(17):6889–6892, ISSN 1091-6490, URL http://dx.doi.org/10.1073/pnas.1018033108.

Dean DL, Hender JM, Rodgers TL, Santanen EL (2006) Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation. *Journal of the Association for Information Systems* 7(10):646–699, URL https://www.researchgate.net/profile/Douglas{_}Dean2/publication/256001578{_}Identifying{_}Good{_}Ideas{_}Constructs{_}and{_}Scales{_}for{_}Ide
links/545ac0ff0cf2c46f66438840.pdf.

Fama EF, Jensen MC (1983) Separation of ownership and control. *Source: Journal of Law and Economics* 26(2):301–325, URL http://www.jstor.org/stable/725104.

Foss NJ (2003) Selective intervention and internal hybrids: Interpreting and learning from the rise and decline of the Oticon spaghetti organization. *Organization Science* 14(3):331–349, URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.7308{&}rep=rep1{&}type=pdf.

Foss NJ, Lyngsie J, Zahra SA (2015) Organizational design correlates of entrepreneurship: The roles of decentralization and formalization for opportunity discovery and realization. *Strategic Organization* 13(1):32–60, ISSN 1741315X, URL http://dx.doi.org/10.1177/1476127014561944.

Goldenberg J, Lehmann DR, Mazursky D (2001) The idea itself and the circumstances of Its emergence as predictors of new product success. *Management Science* 47(1), URL http://dx.doi.org/10.1287/mnsc.47.1.69.10670.

Gompers PA, Gornall W, Kaplan SN, Strebulaev IA (2019) How do venture capitalists make decisions? *Journal of Financial Economics* (xxxx), ISSN 0304405X, URL http://dx.doi.org/10.1016/j.jfineco.2019.06.011.

Gulati R, Puranam P (2009) Renewal through reorganization: The value of inconsistencies between formal and informal organization. *Organization Science* 20(2):422–440, ISSN 10477039, URL http://dx.doi.org/10.1287/orsc.1090.0421.

Hackmann JR (2002) *Leading teams: Setting the stage for great performances* (Harvard Business Press), URL https://books.google.de/books?hl=en{&}lr={&}id=snfoCQAAQBAJ{&}oi=fnd{&}pg=PR5{&}dq=leading+teams+setting+the+stage{&}ots=SAj0vy{_}8-h{&}sig=-wo5huIfpKTWahCwuXVDPCrn7FU{#}v=onepage{&}q=leadingteamssettingthestage{&}f=false.

Hamilton BH (2000) Does entrepreneurship pay? An empirical analysis of the returns of self-employment. *The Journal of Political Economy* 108(3):604–631, URL https://www.researchgate.net/profile/Barton{_}Hamilton/publication/24104134{_}Does{_}Entrepreneurship{_}Pay{_}An{_}Empirical{_}Analysis{_}of{_}th links/00b4952a751bd81090000000.pdf.

Hansen Z, Owan H, Pan J (2015) The impact of group diversity on class performance: evidence from college classrooms. *Education Economics* 23(2):238–258, ISSN 0964-5292, URL http://dx.doi.org/10.1080/09645292.2013.813908.

Hasan S, Koning R (2019) Prior ties and the limits of peer effects on startup team performance. *Strategic Management Journal* (January):1394–1416, ISSN 10970266, URL http://dx.doi.org/10.1002/smj.3032.

Hoogendoorn S, Oosterbeek H, Van Praag M (2013) The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science* 59(7):1514–1528, URL http://pubsonline.informs.orghttp://dx.doi.org/10.1287/.

Hoogendoorn S, Parker SC, van Praag M (2017) Smart or diverse start-up teams? Evidence from a field experiment. *Organization Science* 28(6):1010–1028, ISSN 1047-7039, URL http://dx.doi.org/10.1287/orsc.2017.1158.

Horwitz SK, Horwitz IB (2007) The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management* 33(6):987–1015, URL http://dx.doi.org/10.1177/0149206307308587.

Huang L, Pearce JL (2015) *Managing the Unknowable: The Effectiveness of Early-stage Investor Gut Feel in Entrepreneurial Investment Decisions*, volume 60. ISBN 0001839215597, URL http://dx.doi.org/10.1177/0001839215597270.

Ingram P, Morris MW (2007) Do people mix at mixers? Structures, homophily, and the "Life of the Party". *Administrative Science Quarterly* 52(4):14–26, URL http://journals.sagepub.com/doi/pdf/10.2189/asqu.52.4.558.

Kaplan SN, Sensoy BA, Strömberg P (2009) Should Investors Bet on the Jockey or the Horse. *The Journal of Finance* 64(1):75–115, ISSN 00221082, URL http://dx.doi.org/10.1111/j.1540-6261.2008.01429.x.

Kleinbaum AM, Stuart TE, Tushman ML (2013) Discretion within constraint: Homophily and structure in a formal organization. *Organization Science* 24(5):1316–1336, URL http://dx.doi.org/10.1287/orsc.1120.0804.

Kling JR, Liebman JB, Katz LF (2007) Experimental analysis of neighborhood effects. *Econometrica* 75(1):83–119, URL http://dx.doi.org/10.1111/j.1468-0262.2007.00733.x.

Kornish LJ, Ulrich KT (2014) The importance of the raw idea in innovation: Testing the sow's ear hypothesis. *Journal of Marketing Research* 51(1):14–26, URL https://pdfs.semanticscholar.org/ce90/92701b6be2de4ccc69ba23d511089e66fe18.pdf.

Lamaze A, Van Knippenberg D (2014) Teams in pursuit of radical innovation: A goal orientation perspective. *Academy of Management Review* 39(4):423–438, URL http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0{&}sid=eb64164d-936a-4155-b80b-dfbe3e788be0{%}40sessionmgr102.

Levine SS, Prietula MJ (2014) Open Collaboration for Innovation: Principles and Performance. *Organization Science* 25(5):1414–1433, ISSN 15265455, URL http://dx.doi.org/10.1287/orsc.2013.0872.

Lovas B, Ghoshal S (2000) Strategy as guided evolution. *Strategic Management Journal* 875–896, URL https://flora.insead.edu/fichiersti{_}wp/inseadwp1998/98-37.pdf.

Mao A, Mason W, Suri S, Watts DJ (2016) An experimental study of team size and performance on a complex task. *PLOS ONE* 11(4):e0153048, ISSN 1932-6203, URL http://dx.doi.org/10.1371/journal.pone.0153048.

Maxwell A (2011) Business angel decision making. *Journal of Business Venturing* 26(2):212–225, URL https://uwspace.uwaterloo.ca/bitstream/handle/10012/6484/Maxwell{_}Andre?sequence=1.

McEvily B, Tortoriello M (2014) More Formally: Rediscovering the Missing Link between Formal Organization and Informal Social Structure. *The Academy of Management Annals* 8(1):299–345.

McKenzie D (2017) Identifying and Spurring High-growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *American Economic Review* 107(8):2278–2307, ISSN 00028282, URL http://dx.doi.org/10.1257/aer.20151404.

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–444, URL http://www.leonidzhukov.net/hse/2017/networkscience/papers/McPherson{_}HomophilyInSocialNetworks.pdf.

Moskowitz TJ, Vissing-Jørgensen A (2002) The returns to entrepreneurial investment: A private equity premium puzzle? *American Economic Review* 92(4):745–748, URL https://www.researchgate.net/profile/Tobias{_}Moskowitz/publication/4980985{_}The{_}Returns{_}to{_}Entrepreneurial{_}Investment{_}A{_}Private{_}Eq
links/55ed6e6b08ae65b6389f4e90/The-Returns-to-Entrepreneurial-Investment-A-Pri
pd.

Puranam P, Alexy O, Reitzig M (2014) What's "New" about New Forms of Organizing? *Academy of Management Review* 39(2):162–180, ISSN 03637425, URL http://dx.doi.org/10.5465/amr.2011.0436.

Quintane E, Carnabuci G (2016) How Do Brokers Broker? Tertius Gaudens, Tertius Lungens, and the Temporality of Structural Holes. *Organization Science* 27(6):1343–1360, ISSN 15265455, URL http://dx.doi.org/10.1287/orsc.2016.1091.

Reagans R (2011) Close encounters: Analyzing how social similarity and propinquity contribute to strong network connections. *Organization Science* 22(4):835–849, URL https://dspace.mit.edu/bitstream/handle/1721.1/69851/ce{_}final.pdf?sequence=1.

Reagans R, Zuckermann E, McEvily B (2004) How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative Science Quarterly* 49(1):101–133, URL http://www.jstor.org/stable/4131457.

Ries E (2011) *The lean startup* (New York, NY: The Crown Publishing Group), URL https://itunes.apple.com/de/book/the-lean-startup/id422540072?mt=11.

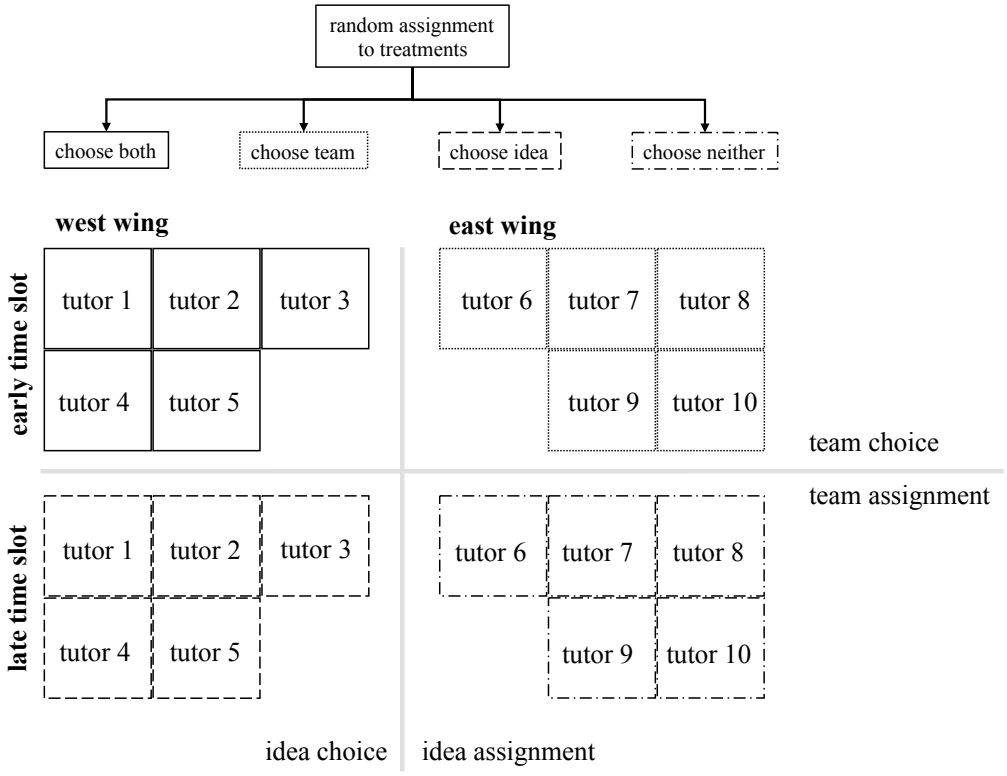
Ruef M, Aldrich HE, Carter NM (2003) The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review* 68(2):195–222, URL https://www.researchgate.net/profile/Howard{_}Aldrich/publication/228264239{_}The{_}Structure{_}of{_}Founding{_}Teams{_}Homophily{_}Strong{_}Tie
links/09e41511a9f2910ff0000000.pdf.

Simon HA (1955) A Behavioral model of rational choice. *The Quarterly Journal of Economics* 69(1):99–118, URL http://dx.doi.org/10.2307/1884852.

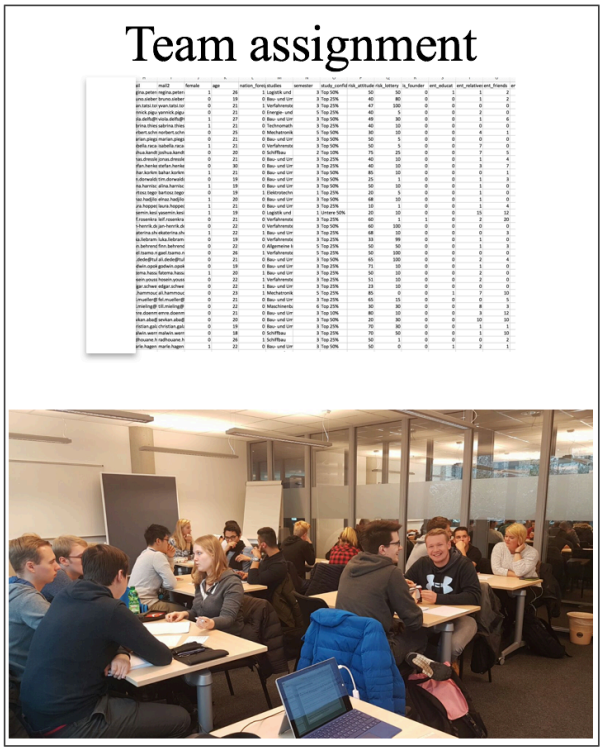
Steffens P, Terjesen S, Davidsson P (2012) Birds of a feather get lost together: new venture team composition and performance. *Small Business Economics* 39(3):727–743, URL https://s3.amazonaws.com/academia.edu.documents/43753872/New{_}Venture{_}Team{_}Composition{_}Persistence20160315-26435-e9bhtx.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A{&}Expires=1521551695{&}Signature=A6dkvqqYVv6gJ{%}2BSAD{%}2Fe2g{%}2BQaDQ4{%}3D{&}responsei.

Thompson LL (2000) *Making the team: A guide for managers* (Upper Saddle River, NJ: Prentice Hall), URL http://www.leighthompson.com/images/books/mtt/TableofContentsandPreface.pdf.

APPENDIX A. TEAM AND IDEA RANDOMIZATION



APPENDIX B. PICTURES OF TEAM AND IDEA SESSIONS



APPENDIX C. SESSION OVERVIEW

Session	Title	Event
1	Team building	Intervention for endogenous or exogenous team composition
2	Idea specification	Intervention for endogenous or exogenous topic
3	Customer	Regular class
4	Vision	Regular class
5	Prototyping	Regular class
6	Market analysis	Regular class
7	Marketing	Regular class
8	Pricing	Regular class
9	Finances	Regular class
10	Feedback	Peer and teacher feedback on pitch deck draft
11	Final presentation	10min final pitch from each team; exit survey afterwards

APPENDIX D. PRE-DETERMINED IDEAS

01) Veganism is the (mostly ethically motivated) complete renunciation of animal products, particularly when it comes to food. Vegan food and a sustainable, healthy and possibly regional diet have gained considerable attention in recent years. A growing interest in veganism is reflected in the increasing demand for meat substitute products and plant-based alternatives. The number of vegan restaurants has increased considerably in recent years and there is a growing demand. Those who choose a life style free of animal-based products, and who like to cook for themselves but have little time or not enough experience often miss suitable opportunities and offers.

02) Spontaneous visits to pubs, restaurants, and bars can get tricky due to lack of space. This is especially the case when a larger group is looking to sit together. One way to deal with this problem would be a group reservation, but this requires a certain degree of planning and spontaneous visits to a restaurant are impossible. Another interesting point in this context is the fact that larger tables are ”under-occupied”. This means that, for example, a table for four persons is only used by two.

03) During sightseeing or a city outing, you would usually learn about cultural and historical aspects of your destination. Unfortunately, you do not know the place and guided tours can take too long. In addition, guides tours typically concentrate on one part of the city/sights and you may find yourself in a throng of other tourists, wedged into a historical building or a monument. Therefore, people may prefer to discover sights or a city independently, but lack the historical/cultural knowledge about the place.

04) This may sound familiar: It is the end of the month, and the data plan on your smartphone is exhausted. It takes ages for a website to load. Mobile service providers advertise increasingly favorable tariffs, which at the same time offer more and more data volumes. But why not use free hotspots, which are now available in every city? The downside is that to log in and to enter the required data repeatedly is often time- consuming.

05) In modern society, health in general and a healthy diet in particular are becoming more and more important. For many, a healthy juice is part of every breakfast. However, sugar and industrial additives are often added to such juices. In addition to physical needs for allergy sufferers, there is also great potential through fun during customizing your own product. There is a limited range of juice varieties, which can impose unnecessary restrictions on allergy sufferers or consumers who prefer more choice.

06) Job interviews usually have several stages. The entire application process is very complex and costs time and money. One way of avoiding these costs could be to shorten application periods and, for example, retrieve important information from cover letters and CVs. However, this way, employers often miss out on a comprehensive impression of an applicant. Video interview are becoming increasingly popular in application processes, as they reduce travel time. Still, it is necessary to agree on a specific date, which can also be cost-relevant and time-intensive in the first stages of an application.

07) Individual design is becoming more and more popular, be it in fashion, interior design, etc.. Commissioning a professional designer is very costly, so unique, customized designs are only available to

those on a high income. Apart from exclusively designed products, there is a great potential of talented and capable young designers, who can support a large variety of favorable, aesthetic life styles with their flair and individuality, while building their own portfolio.

08) It is a common saying that a dog is a man's best friend. Nevertheless, keeping a dog means taking responsibility and many dogs end up in the animal shelter. Their owners realize that they simply do not have the time or capacity to take care of them, particularly during holidays, as often pets cannot be taken along.

09) Geocaching is a modern form of treasure hunting. Equipped with a Global Positioning System (GPS) receiver and the coordinates of a "treasure" (for example from online forums), the treasure hunter can locate what others have hidden in unusual places. Another example are Live Escape Games, where small groups of people are (virtually) locked in a room or several rooms and have to escape their jail again within a given time (usually 60 minutes) with the help of hidden notes and objects. Both activities are becoming increasingly popular, for people of all ages. Still, there is no combination of the two.

10) Moving something from A to B proves often more difficult than transporting people. Small items can be conveniently sent by mail and other service providers, bulkier objects are more difficult to dispatch. At the same time, there is much, albeit underused, capacity when it comes to privately owned cars. Here, it is quite usual to offer someone a lift, but not so much when it comes to transporting items.

11) In the past fifteen year, the amount of plastic waste in Germany has more than doubled. Most of the plastic waste ends up in the ocean posing a serious threat to sea life. Inevitably, plastic then ends up on our plates when we eat seafood. Advertising material, such as posters and flyers, which are only short-lived, are good examples of such plastic waste. Using degradable materials would help to counteract this problem.

12) Many public places lack opportunities to buy fresh, healthy snacks, although there is definitely a demand. But what people get is unhealthy snacks, such as crisps, sweets, and sugary drinks from vending machines. Healthy alternatives to conventional vending-machine snacks is what is needed. A balanced offer of in-between meals in the fast way is still missing. Instead of fatty and sugary snacks, appropriate alternatives are required.

13) Shipping boxes in all shapes and sizes have become very popular in recent years. You can have anything delivered at home in the appropriate box: books, cosmetics, food. But what is missing here is interaction between people. Customers do not only want the purchased article as such, they also want communicate about it, try out something new, share it with friends.

14) For more and more people, the work- life balance is very important. But with the immense variety and different possibilities out there, it gets increasingly difficult to find the right balance between life and work. Annoying and unpopular activities such as food shopping, restaurant booking, but also the birthday present for the aunt etc. could happily be given to somebody else.

15) Becoming a university or college student mostly entails moving to another city or country even. Everything is new, exciting, and strange. What students often miss is a contact center, where they can get information about living, working, studying, events, travelling, and get advice on how to live on a low budget.

APPENDIX E. EVALUATION PROCEDURE

This evaluation was presented to the evaluators during their evaluation. They were given each section in numerical order.

1) Introduction

Dear evaluator,

we have 23-24 pitch decks for your review. These were created by students as part of a university competition. The aim of this review is to reward the best pitch decks. First of all, we would like to ask you to provide some information about yourself and your experience in the field of start ups.

Afterwards, please rate each pitch deck in the order of the stack. All decks are enclosed with an evaluation form stapled to the front. Each pitch deck should be rated in the categories "novelty of the idea", "feasibility of the idea", "market potential of the idea" and "implementation as a pitch deck". The assessment is on a scale of 1 (really bad) to 7 (really good). In addition, you should estimate the probabilities for inviting the team and the success of a company founded on the basis of the respective pitch deck as a percentage.

After completing the individual evaluations of each pitch deck, please decide on possible investments in the respective start ups in a second step. For this, we provide you with an investment budget of 1 Mio. Startup-$. You should distribute this money on the pitch decks according to your preferences. Please stick the appropriate investment amount to the evaluation form of the respective pitch deck. If you do not want to use part of your budget, leave the stickers aside. You will then receive another short final questionnaire. We ask that you do not talk to the other evaluators during the evaluation. If you have questions, you can always contact us directly.

We sincerely thank you for your participation and wish you a lot of fun!

2) Evaluator information

Please provide the following information:

Name

Gender

Year of birth

Highest academic degree (and field)

Current job (and field)

I have founded ___ startups.

I have worked in ___ startups.

I have coached ___ startups.

I have participated in ___ courses and training in the field of startups.

I have co-funded ___ startups with my personal capital.

I was involved in deciding on the financing of ___ investor-initiated startups.

I have already rated ___ startups.

3) Evaluation of each individual pitch deck

Please rate this pitch deck along the following criteria using a 7-point scale (1 = really bad, 7 = really good):

Novelty

Feasibility

Market Potential

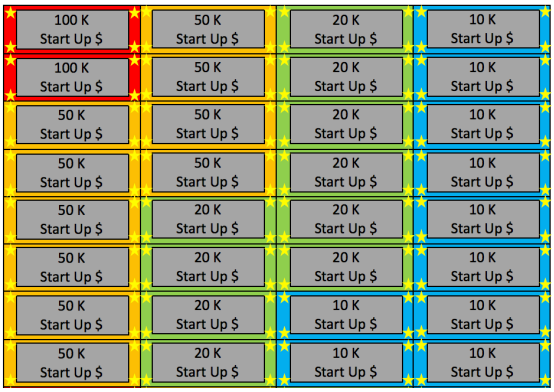
Implementation as a pitch deck

Personal interest in the idea

What is the likelihood that you will invite the start-up team for a personal interview? *(0-100%)*

What is the probability of success of a start-up based on this pitch deck? *(1-100%)*

Please allocate your budget of 1 Mio. Startup-$ *on sticky paper between your bunch of pitch decks.*



4) Exit questions

Did you understand the task? *(y/n)*

Did you feel qualified to evaluate these pitch decks? *(1) not at all — (7) absolutely*

APPENDIX F. BASELINE SURVEY

The following survey was given to the participants before the experiment. It was mandatory for participation in the class and had no impact on the random assignment of participants to treatment - despite the survey being masked as a team assignment exercise. This course of action, as well as the experiment in its entirety, was approved by the responsible ethics committee of the university. This is a translation of the original survey from German:

1) Introduction

Dear student,

The goal of this class ("Project Entrepreneurship") is to show you how to successfully build a startup by working on a practical project in a team. To ensure your succes, we try to assign suitable project teams and tasks to you. We would like to ask you for some information in this 10-minute survey. Before starting with the first exercise, you will be informed about your allocation to a team.

ATTENTION: We would like to point out that completing this survey is mandatory for project participant. Leaving the class during the semester is not possible. If you complete the project, but do not participate directly in the exam, we will gladly include your project grade in the future.

Thank you for your support.

2) Personal Details

Please state your full name.

Please state your e-mail address.

Please state your gender.

Please state your year of birth.

Please state your educational background.

Please state your current semester.

Which performance range do you see yourself in at the end of the project compared to your fellow students? *(Top 5% Top 10%, Top 25%, Top 50%, Bottom 50%)*

How do your rate your own risk behavior? *(0) extremely risk averse — (100) extremely risk seeking*

How much would you spend on a lottery ticket with a 10% chance to win 1000€? *(integer values)*

Have you ever become self-employed or founded your own company? *(y/n)*

Have you ever participated in entrepreneurship focused courses, trainings or competitions? *(y/n)*

How many of your relatives are or have been self-employed or have established a company? *(integer values)*

How many of your friends are or have been self-employed or have founded a company? *(integer values)*

How many newly founded companies have your worked in or managed? *(integer values)*

Please answer the following questions on a scale from 1 (not true at all) to 7 (completely true). I feel capable of: *starting and running a successful company; defining realistic objectives for a company; designing products and services that solve current problems; doing a market research; setting a profitable price for a product or a service; making decisions under uncertainty and risk; creating products and services that meet the needs of current customers; identifying potential sources of funding for investments; keeping an overview of the business costs.*

3) Your motivation

What kind of work will you most likely pursue after graduation? *(large company, medium/small company, consultancy, academia, entrepreneur)*

How likely will you start a business or become self-employed after graduation? *((0) completely unlikely — (100) completely sure)*

4) Team composition

Please name up to 5 preferred team members.

Please state how you know each of your preferred team members. *(We know each other from university. We know each other from outside university.)*

Please state how close you are with each of your preferred team members. *(We are loose acquaintances. We are acquaintances. We are friends. We are good friends.)*

Please rank the following entrepreneurial tasks from highest to lowest according to your preference. *(Strategic planning and vision; Product design; Technical development; Marketing and sales; Finance)*

APPENDIX G. ENDLINE SURVEY

This survey was presented to the participants in paper form after the final presentation. Each participant was uniquely identified by the tutors and given their corresponding survey. This survey was masked as the teaching evaluation. Some of the questions are repeated from the entry survey to monitor students' progress. This is a translation of the original survey from German:

1) Introduction

Dear student,

in order to improve this class ("Project Entrepreneurship"), we ask you to participate in this 10-min survey. The goal of this survey is to identify possible areas for future improvements for following semesters. Your answers have no impact on your grade and are treated anonymously.

Thank you for your support.

2) Teaching goals

Please answer the following questions on a scale from 1 (not true at all) to 7 (completely true). I feel capable of: *starting and running a successful company; defining realistic objectives for a company; designing products and services that solve current problems; doing a market research; setting a profitable price for a product or a service; making decisions under uncertainty and risk; creating products and services that meet the needs of current customers; identifying potential sources of funding for investments; keeping an overview of the business costs.*

What kind of work will you most likely pursue after graduation? *(large company, medium/small company, consultancy, academia, entrepreneur)*

Which performance range do you see yourself in at the end of the project compared with your fellow students? *(Top 5% Top 10%, Top 25%, Top 50%, Bottom 50%)*

How likely will you start a business or become self-employed after graduation? *((0) completely unlikely — (100) completely sure)*

3) Team work

Please answer the following questions on a scale from 1 (not true at all) to 7 (completely true). The members of my team: *(had good work relationships; had good friendship relationships; enjoyed spending work time together; enjoyed spending free time together; were good colleagues; were good friends)*

How much of the workload did you contribute to the team workload? *(percentage value)*

With how many of your team members would you work again? *(all/some/none)*

How many of your team members did you know before this class? *(all/some/none)*

How did you distribute tasks within your team? *((1) we worked completely individually – (7) we worked completely collaboratively)*

4) Idea

Please answer the following questions on a scale from 1 (not true at all) to 7 (completely true). Our project idea: *(had me excited.; had me motivated; made it possible for me to identify myself with the idea; was suited for this class; matched my previous expertise; matched with my team)*

How do you rate the success potential of a company based on your idea? *(percentage)*

How likely would you found a company based on your idea? *(percentage)*

5) Satisfaction How happy were you with your team? *(1) extremely dissatisfied — (7) extremely satisfied*

How happy were you with your project idea? *(1) extremely dissatisfied — (7) extremely satisfied*

There were a number of additional questions similar to those typically asked in a teaching evaluation, but these were not used for this experiment, because they pertained to topics such as rooms, times, and lectures.