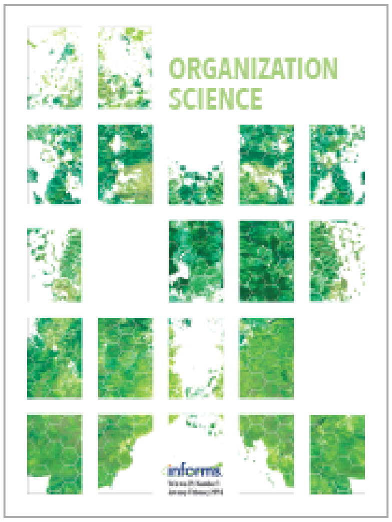


Organizing Entrepreneurial Teams: A Field Experiment on Autonomy over Choosing Teams and Ideas

Viktoria Boss, Linus Dahlander, Christoph Ihl, Rajshri Jayaraman

informs
https://pubsonline.informs.org/journal/orsc

ORGANIZATION SCIENCE
Vol. 34, No. 6, November–December 2023, pp. 2097–2118
ISSN 1047-7039 (print), ISSN 1526-5455 (online)

Organizing Entrepreneurial Teams: A Field Experiment on Autonomy over Choosing Teams and Ideas

Viktoria Boss,[a] **Linus Dahlander,**[b] **Christoph Ihl,**[a] **Rajshri Jayaraman**[b,c]

[a] Institute of Entrepreneurship, Hamburg University of Technology, 21073 Hamburg, Germany; [b] ESMT Berlin, 10178 Berlin, Germany; [c] University of Toronto, Toronto, Ontario M5S 3K7, Canada

Contact: viktoria.boss@tuhh.de (VB); linus.dahlander@esmt.org, https://orcid.org/0000-0003-3527-7440 (LD); christoph.ihl@tuhh.de, https://orcid.org/0000-0002-0842-5201 (CI); rajshri.jayaraman@esmt.org, https://orcid.org/0000-0002-7942-7052 (RJ)

Received: September 15, 2019
Revised: June 30, 2020; January 16, 2021; June 28, 2021
Accepted: July 26, 2021
Published Online in Articles in Advance: November 5, 2021

https://doi.org/10.1287/orsc.2021.1520

Abstract. Scholars have suggested that autonomy can lead to better entrepreneurial team performance. Yet, there are different types of autonomy, and they come at a cost. We shed light on whether two fundamental organizational design choices—granting teams autonomy to (1) choose project ideas to work on and (2) choose team members to work with—affect performance. We run a field experiment involving 939 students in a lean startup entrepreneurship course over 11 weeks. The aim is to disentangle the separate and joint effects of granting autonomy over choosing teams and choosing ideas compared with a baseline treatment with preassigned ideas and team members. We find that teams with autonomy over choosing either ideas or team members outperform teams in the baseline treatment as measured by pitch deck performance. The effect of choosing *ideas* is significantly stronger than the effect of choosing *teams*. However, the performance gains vanish for teams that are granted full autonomy over choosing *both* ideas and teams. This suggests the two forms of autonomy are substitutes. Causal mediation analysis reveals that the main effects of choosing ideas or teams can be partly explained by a better match of ideas with team members' interests and prior network contacts among team members, respectively. Although homophily and lack of team diversity cannot explain the performance drop among teams with full autonomy, our results suggest that self-selected teams fall prey to overconfidence and complacency too early to fully exploit the potential of their chosen idea. We discuss the implications of these findings for research on organizational design, autonomy, and innovation.

History: This paper has been accepted for the *Organization Science* Special Issue on Experiments in Organizational Theory.

Funding: The authors are grateful for funding from the Innovation Growth Laboratory, the Peter Curtius Foundation, and the Deutsche Forschungsgemeinschaft [Grant CRC TRR 190].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/orsc.2021.1520.

Keywords: autonomy • teams • ideas • entrepreneurial performance • field experiment

1. Introduction

Companies aim to inspire innovation and entrepreneurship throughout their organizations (Burgelman 1983, Kanter 1985). Ample research on autonomy at work has shown that an employee's opportunity to have a say in how to do their work increases creativity and innovation (Amabile and Gitomer 1984) and entrepreneurial behavior (Lumpkin et al. 2009). This literature underscores that innovation and entrepreneurship are fostered when teams have autonomy over day-to-day tasks and a sense of ownership regarding their work and ideas (Pelz and Andrews 1966, Bailyn 1985). Individuals with autonomy are more likely to create unconventional, ground-breaking ideas (Miner 1994) and generate novel inventions (Gambardella et al. 2020).

At the same time, organization scholars have questioned whether granting autonomy is the best way to organize for innovation and entrepreneurship (Shimizu 2012, Clement and Puranam 2018). Granting complete autonomy to everyone on a team can give rise to coordination problems, distracting from the company's strategic direction (Van de Ven 1986, Simon et al. 1999, Shimizu 2012, Gambardella et al. 2020). The challenge of knowing when and how to provide autonomy is not just an academic exercise. Companies such as Valve and 3M grant their employees time off from their daily work schedule to encourage innovation within the company (Biancani et al. 2014, Lovas and Ghoshal 2000, Puranam and Håkonsson 2015). A lesser-known example is the Swedish music-streaming service Spotify, which organizes its teams into so-called "tribes," which are further divided into "squads," that work independently on different functional areas. Their ambition is to provide autonomy at every level by allowing employees to make more timely decisions, thereby accelerating teams' innovation performance.

Although the literature has noted that autonomy has advantages and disadvantages, it has remained rather silent on how autonomy is actually implemented along particular dimensions and how that matters for performance. Although there are a few studies trying to disentangle the effects of autonomy on the firm level (Lumpkin et al. 2009) and the individual level (Gambardella et al. 2020), empirical evidence on the team level is still scarce. To further disentangle autonomy over entrepreneurial teams, we take an organizational design perspective and focus on two interrelated problem dimensions: (1) task division, which refers to the breakdown of the organization's goals into contributory tasks, and (2) task allocation, which pertains to the assignment of these tasks to individual members within the organization (Puranam et al. 2014). In our conception of autonomy, the two organizational problem dimensions can be solved either through managerial assignment, that is, without autonomy, or through entrepreneurial self-selection, that is, with autonomy. Entrepreneurial teams that are granted autonomy can choose a business idea to work on and/or team members to work with, instead of being assigned to an idea and team members.

Prior work has investigated the effects of these two dimensions of autonomy, albeit separately and not necessarily in the context of entrepreneurial teams. Scholars have studied the allocation of tasks in the context of new organizational forms (Puranam et al. 2014). For example, they have documented that more "horizontal" organizations tend to allow for some autonomy over choosing tasks (Barley and Kunda 2001) as do more "open" collaboration contexts (Levine and Prietula 2014). Research has shown that autonomy over task selection can foster motivation (Gambardella et al. 2020) and facilitate better matches between tasks and team members' skills and interests, as people get to choose their pet projects (Criscuolo et al. 2014). According to the qualitative work of Bailyn (1985), choosing a task to work on refers to strategic autonomy (what to do) rather than operational autonomy (how to do it). On the other hand, research on self-selection of collaboration partners has shown that individuals tend to prefer close friends as coworkers (Ingram and Morris 2007). Although this focus on familiarity and trust can lead to lower complementarity of team competencies (Ruef et al. 2003), it may also improve cohesion and coordination of tasks between team members (Reagans et al. 2004).

Although autonomy can increase motivation and help create teams that work well together, it may also fail to push people outside their comfort zone. If autonomy leads individuals to choose topics and team members they are too familiar with, it might reduce performance. This raises the question of whether allowing or disallowing autonomy of choice in team members or ideas improves team performance. Disentangling these two types of autonomy allows an examination of whether one type of autonomy matters more than the other, and whether the two types are complements or substitutes. If they are complements, then granting autonomy over both dimensions would lead to the highest performance. If they are substitutes, then it would imply that granting autonomy over one dimension may be detrimental to the performance effect of autonomy over the other dimension. Finally, disentangling the two sources of autonomy allows for a more fine-grained account of the underlying mechanisms through which autonomy operates.

To achieve this objective, we conduct a field experiment in an entrepreneurship course at a German university. The experiment has a two × two factorial design, in which one factor pertains to the autonomy over choosing (versus being assigned) team members and the other factor pertains to the autonomy over choosing (versus being assigned) ideas. Participants are randomly assigned to one of four resulting treatment groups. In the first treatment group, *Choose team*, participants can choose their own team but have to work on a randomly assigned, predefined 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. Following best practices, we received ethics approval and filed a preanalysis plan before starting the experiment.

The design allows us to compare two fundamentally different approaches to organizing entrepreneurship and innovation. At one extreme, assigning people to teams and ideas reflects managerial, hierarchical organizing (i.e., the baseline or "business as usual"). At the other extreme, the self-selection condition in which participants have the autonomy to choose both team members and ideas is meant to mimic "green field" entrepreneurship, where entrepreneurs have full freedom to choose. The ideal context for our experiment would, of course, be a real organizational setting with actual entrepreneurial teams. The university setting, however, allows us to strike a balance between control and realism. In terms of control, we can cleanly execute the treatments and achieve a sample size of 939 students divided into the 310 teams needed for our experimental design and statistical inference, even in the tails of the performance distribution.

In terms of realism, the course structure and content are inspired by the lean startup approach, which these days is also applied in corporate contexts to inspire entrepreneurship and innovation (Hampel et al. 2020). This course culminates after 11 weeks in a pitch deck that details a business proposal, which is evaluated by practicing venture capitalists, business angels, and seasoned entrepreneurs. Pitch deck performance is a noisy predictor of ultimate business success (McKenzie (2017) expertly examines this later-stage outcome), but it mimics the first-stage performance evaluation by venture capitalists, which entrepreneurs need to master to get their businesses off the ground (Brooks et al. 2014, Huang and Pearce 2015, Bernstein et al. 2017). Furthermore, we do not necessarily have to assume that the sampled students behave and react like experienced entrepreneurs, but more like employees in an organizational setting, in which entrepreneurship is inspired by the means of organizational practices (Hoogendoorn et al. 2013).

We use random assignment in the baseline treatment because, as Clement and Puranam (2018, p. 3880) emphasizes, "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." We know from prior studies that there is a significant random component in how companies work, for example, basing their team and task composition on coincidental temporal or spatial location (Liu et al. 2016). Furthermore, work on new organizational forms (Puranam et al. 2014) shows that managerial task allocation has limitations when tasks are novel and complex, and there are many potential candidates with hidden skill sets. Therefore, the approximation of managerial organization through random task allocation may well be justified in the context of assigning novel tasks in business development projects in a larger organization, in which managers can only imperfectly draw on employees' prior knowledge and skill sets.

Our work makes two main contributions. The first speaks to the role of autonomy over innovation. Many scholars have argued that providing autonomy is an important ingredient in innovation (Pelz and Andrews 1966, Amabile and Gitomer 1984). Rather than looking at individual-level outcomes, as most prior work on autonomy has done, we turn our attention to the effects on team performance. We contribute to this literature by distinguishing between two kinds of autonomy—choosing teams versus choosing ideas—which has to date been collapsed into a single dimension as a choice/no-choice dichotomy. We find that making this distinction matters. Our results show that teams with autonomy over choosing either ideas or team members outperform teams in the baseline treatment with assignment. The effect of choosing ideas is significantly stronger than the effect of choosing teams. By carefully tracing the teams and their activities with baseline and follow-up surveys over the course of the experiment, we further point to some candidate mechanisms that account for this result. Causal mediation analysis reveals that the positive main effects of choosing either ideas or teams can partly be explained by a better match of ideas with team members' interests and prior network contacts among team members, respectively. However, the performance gains vanish for teams that are granted full autonomy over choosing both ideas and teams. This suggests that there is a substitution effect between the two forms of autonomy. Thus, there are downsides to autonomy that can be easily overlooked (Gambardella et al. 2020), and these sometimes outweigh its advantages. In fact, conditional on choosing ideas, we find that teams constructed randomly outperform teams that form through choice. Whereas homophily and lack of team diversity cannot explain the performance drop among teams with full autonomy, our results suggest that self-selected teams fall prey to overconfidence and complacency too early to fully exploit the potential of their chosen idea.

Thus, our second contribution pertains to the organization of entrepreneurial teams. Teams play a key role in entrepreneurial success, but we know little about which organizational design choices maximize chances of success. By drawing insights from extant theory on the role of team autonomy in successful entrepreneurial teams, we provide a caveat that self-organizing entrepreneurial teams may not translate into higher performance. This is because granting autonomy over both teams and ideas can lead to too much confidence too early on in a project, thus triggering an upward "efficacy-performance spiral" (Lindsley et al. 1995), whereby highly familiar, homogenous team members can get trapped both in terms of their cognition and motivation, resulting in inertia and complacency. For the organizational designer, this implies that one must consider providing some autonomy while being mindful that allowing people to choose team members can reduce their performance.

2. Theory and Hypotheses

There is ample prior work on team composition and its effects on different facets of entrepreneurial performance (Horwitz and Horwitz 2007, Steffens et al. 2012, Cooper and Saral 2013). This literature explores how entrepreneurs, or even intrapreneurs, can self-select collaboration partners. Although entrepreneurship research has begun to pay attention to causal inference using experimental designs (McKenzie 2017, Clingingsmith and Shane 2018), there are few field experiments pertaining to entrepreneurial teams (Mao et al. 2016, Hasan and Koning 2019). This is likely because of the difficulty of finding a field setting with a meaningfully sized group of potential entrepreneurs—one with a large enough sample of teams to yield statistically meaningful results (Camuffo et al. 2020).

In two important examples of studies that overcome these hurdles, Hoogendoorn and coauthors investigate the effect of team composition on performance based on exogenous assignment to teams. In a first such 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, although average team ability has no effect on outcomes, variation in ability within teams does. Specifically, 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 an important way. Most of the literature on entrepreneurial teams discussed has explored the effect of team composition on performance, whereas our focus is the effect of autonomy over choosing teams and ideas. Team composition, in this context, may be a consequence of choice, but our focus is on the ramifications of autonomy over choosing teams and ideas on early-stage entrepreneurial performance. In what follows, we elaborate on how autonomy over choice of team members and ideas affects team performance, building on hypotheses from our preanalysis plan. Our focus is on early-stage team performance as captured by pitch decks examined at preliminary stages of a venture capitalist's deal flow (McKenzie 2017, Clingingsmith and Shane 2018). In our post hoc analysis, we then disentangle the different underlying channels through which autonomy operates.

2.1. Autonomy over Choosing Teams

A first dimension of autonomy is choosing team members, or the self-selection of collaboration partners. This possibility has real-world precedents. As Burt and Merluzzi (2016, p. 374) note, not all collaborations are a matter of choice: collaborations in many workplaces are characterized by a "mixture of exogenous assignment and endogenous choice, with the mix playing out differently for different individuals." Autonomy in choosing collaborators is ubiquitous in the context of entrepreneurship, but larger companies have also experimented with more lateral forms of organizing (Dahlander and O'Mahony 2011, Kleinbaum et al. 2013, Biancani et al. 2014). In general, the ability to choose collaborators 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 because of unobserved heterogeneity across teams (Hansen et al. 2015).

Research on autonomy shows that providing such operational autonomy makes people more innovative and better able to generate ground-breaking ideas (Bailyn 1985) and that two potential mechanisms can explain how autonomy over choice of collaborators affects entrepreneurial team performance. First, autonomy enables people to 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). Teams that self-select can take advantage of greater levels of familiarity among the members and work together more effectively (Huckman et al. 2009, Cattani et al. 2013), resulting in higher performance.

Second, choosing team members allows for picking members based on complementary personality traits, skills, or experiences (McPherson et al. 2001) independently of whether they had a prior relationship. For example, one can imagine a situation where people sort themselves into teams based on their ambition, which is likely to determine the effectiveness with which a task can be handled (Brannon et al. 2013). In this situation, self-selected teams can outperform teams assigned by managers who have less-informed insight into such complementarities.

There are also potential dangers of autonomy (Criscuolo et al. 2014, Gambardella et al. 2020). A natural presumption has been that when autonomy is granted, individuals can choose collaborators with whom their objectives coincide (Burt 2005). This has, of course, come under scrutiny. For example, scholars have observed a tendency toward relational inertia, whereby people stick to existing collaborators even when better matches are available (Seabright et al. 1992, Gulati and Gargiulo 1999). Breaking away from a circle of friends can be difficult; Ingram and Morris (2007) show that even well-motivated EMBA students struggle to go beyond their comfort zone during mixers, rarely creating ties that extend beyond friends of their friends. The implication is that autonomous team formation may lead people to overweigh personal ties at the expense of other factors. Ruef et al. (2003) 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" (Ruef et al. 2003, p. 217). This raises the possibility that placing constraints on collaborator choice may enhance outcomes.

To synthesize, people who are given the option of choosing their own team members are inclined to choose people they know over strangers. This leads to (1) greater familiarity in the team and (2) increased complementarities in skills, knowledge, and ambition, which translate into higher performance. There are some dangers arising from (3) inherent preferences to pick friends over better potential matches. Overall, however, prior work suggests that the advantages of familiarity and complementarities outweigh the disadvantages of choosing teams. Teams that can self-select would potentially have fewer conflicts and operate more smoothly compared with teams that are assigned. We thus hypothesize the following.

Hypothesis 1. *Autonomy over choice of team members leads to higher entrepreneurial team performance.*

2.2. Autonomy over Choosing Ideas

A second dimension of autonomy enables people to self-select ideas or tasks to work on (Bailyn 1985). As the study of Burgelman (1983) illustrates, this is important even in large companies, which use organizational design elements such as "autonomous strategic behavior" to create an environment for innovation. Of course, autonomy over idea choice provides a different set of challenges than autonomy over team composition. Traditionally, work within organizations is characterized by task division and task allocation, with managers telling subordinates what to do (Puranam et al. 2014). The rationale for this is to improve efficiency and enable coordination (Fama and Jensen 1983). By the same token, task assignment may demotivate people, and this has especially negative consequences in the context of tasks that require innovation and creativity.

There are two primary mechanisms through which autonomy over choosing ideas leads to higher performance. First, it can be a major source of intrinsic motivation (Lovas and Ghoshal 2000) and important in attracting and retaining talent (Sauermann and Cohen 2010). For example, Benz and Frey (2008) find that self-employed people are much more content than the employed, which they attribute to more autonomy and involvement. 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. In essence, this literature suggests that autonomy results in intrinsic motivation from a greater sense of ownership, which boosts performance.

This has led companies such as Github, Valve, and Oticon to experiment with letting people self-select the ideas they work on. Having the autonomy to choose ideas may increase performance, because people become more attached to their choices. In their study of the Danish hearing aid company Oticon, which gave employees the autonomy to choose ideas to work on, Lovas and Ghoshal (2000, p. 890) 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." In a critical examination of Oticon, Foss (2003) notes that employees suddenly had the authority to work like entrepreneurs in a market setting. Although Oticon has since moved away from their complete bottom-up approach, the concept of employees who come up with their own projects is still intact in other settings. For instance, studies of open source and Wikipedia note that some autonomy is given to people to choose tasks (Lee and Cole 2003, Gallus 2017, Klapper and Reitzig 2018).

Second, choosing ideas facilitates matching skills to tasks that would lead to higher performance. When building teams, organizational designers often choose members based on how their skills, expertise, or preferences contribute to the overall objective (Hackman and Oldham 1975, Moreland and Argote 2003). Allowing for autonomy over idea choice facilitates the organic development of this matching, because those ideas that are most likely to succeed, based on team members' characteristics, can be selected by the team. In other words, idea choice allows people to creatively identify ideas that best match a team's interests, backgrounds, and skills. Idea choice autonomy overcomes the challenge of information asymmetry regarding workers' characteristics faced by organizational designers when assigning people to teams (MacCormack et al. 2012). With autonomy, the skill and preference matching is left to the team.

In sum, the previous arguments suggest that autonomy over idea choice (1) increases intrinsic motivation and (2) facilitates a closer match to the team's skills and preferences. These two effects outweigh the potential dark sides of idea choice arising from people not going outside their comfort zone (Pierce et al. 2001). We thus hypothesize the following.

Hypothesis 2. *Autonomy over choice of ideas leads to higher entrepreneurial team performance.*

2.3. Autonomy over Choosing Both Teams and Ideas

Thus far, our reasoning has considered the ability to choose teams and ideas separately. The two choices, however, open the intriguing possibility of considering them in tandem. The implicit notion in the literature is that both types are granted simultaneously (Criscuolo et al. 2014). Put differently, the theoretical separation between autonomy over choice of teams and ideas allows us to theorize whether, conditional on choosing collaborators, teams that have autonomy over idea choice outperform teams that do not, or whether, conditional on choosing ideas, a self-selected team outperforms an a priori assigned one. We elaborate on the mechanisms that explain why autonomy over the choice of teams and ideas are *complements* versus *substitutes*.

On the one hand, there are arguments that granting autonomy over both teams and ideas would be *complementary* and result in the highest performance. Much research on autonomy suggests that more freedom to explore options positively affects creative performance (Hackmann 2002). We often observe this situation in real life, where more entrepreneurial settings allow people to choose both their teams and ideas. This environment provides the largest scope for selection, with few formalized structures that constrain choice. Autonomy then allows for choosing team members with prior ties, resulting in familiarity, coherence, and smooth collaboration, and choosing an idea that matches the team members' skills and preferences. In other words, autonomy over both dimensions encapsulates all the benefits from both autonomy over choosing team members and choosing the idea, which could lead to the highest performance. Although choosing an idea can be dominated by one individual's contribution, the elaboration of that idea into a compelling business proposal requires a concerted team effort. Recent research has shown that creative teams need to develop "collective psychological ownership" around an individual team member's idea to be successful and that this requires preventing or mitigating team conflicts (Gray et al. 2020). Arguing in favor of complementarity, the autonomy to choose team members can achieve exactly that: preventing or mitigating conflicts among more familiar and trusted team members, who also need to agree and commit to a chosen idea.

On the other hand, research suggests that granting too much autonomy has a cost, which implies that choosing teams and ideas can be *substitutes*. Bernstein et al. (2016), for instance, claim that many studies on "flat" organizations "take an extreme position" and that granting autonomy overlooks the many challenges that exist. One must consider what happens if more homogenous teams get to choose ideas, as a result of granting more autonomy. Teams with a high degree of familiarity may be too distracted by their social interactions to choose suitable ideas or even work efficiently. Furthermore, they may choose ideas and additional information based on their common experience rather than look for higher business merit "further away" from their comfort zone. As a result, this leads to a cognitive lock-in where more homogeneous teams with high levels of prior ties lack the diversity of perspective and imagination to turn a novel idea into a compelling business proposal (Shin et al. 2012). In such an "echo chamber," team members think too much alike and reinforce each other's beliefs and opinions too early and easily.

Granting autonomy can increase teams' perceptions of efficacy, resulting in a positive motivational effect. However, a number of researchers have raised concerns that teams in which efficacy rises above a critical threshold too early are likely to enter into an "efficacy-performance spiral" (Lindsley et al. 1995) and fall prey to overconfidence and complacency (Gist 1987, Sitkin 1992, Lindsley et al. 1995, Knight et al. 2001, Goncalo et al. 2010, Rapp et al. 2014). This can lead to the allocation of insufficient effort toward task completion. In the context of entrepreneurial teams, early feelings of satisfaction and easy triumphs, such as having your friend on the team or coming up with your own idea, can reduce follow-on effort and experimentation, as well as search for and attention to external feedback (Sitkin 1992). A self-selected team becomes more optimistic, and as a result, the team fails to update for new information and learns less (Amore et al. 2021). Especially among familiar teammates, there is a tendency toward internal self-assurance rather than paying attention to external task demands and performance standards (Moore and Healy 2008, Rapp et al. 2014). As a result, performance could potentially be reduced by granting autonomy over choice along both these dimensions.

Combining the arguments for choice of teams and ideas as being complements versus substitutes, one can conclude that there are plausible arguments for both. A priori, it is difficult to separate which one dominates. If the effects are substitutes, then it suggests that granting complete autonomy leads to lower performance than only granting autonomy over either team members or ideas. If they are complements, then teams who are granted authority in both dimensions would outperform the rest. This leads us to formulate a competing hypothesis.

Hypothesis 3. *Autonomy over choice of team members and autonomy over choice of ideas are (a) complements or (b) substitutes in their effect on entrepreneurial performance.*

3. Experiment

3.1. Setting

To explore the effect of choice of teams and ideas on entrepreneurial performance, we conduct a field experiment. A preanalysis plan was registered at the American Economic Association RCT registry

(AEARCTR-0001179), and ethics approval was obtained from the university. The experiment itself was conducted in a natural field setting, within the context of a compulsory Business and Entrepreneurship course at a public university in Germany. More precisely, 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 point during their undergraduate study. This course is offered each semester; 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. A variety of professors cover their respective areas of expertise in lectures, attended by all students, who then separate into smaller groups for tutorials. The experiment took place in the context of the tutorial component of this course, during which students, organized into entrepreneurial teams, worked on developing and pitching business ideas.

Semesters are 11 weeks long, and 90-minute tutorials take place once each week. Each tutorial is run by one teaching assistant (TA) and one experienced entrepreneur. We will refer to this TA-entrepreneur pair as a "mentor" hereafter. To prevent experimental effects and in keeping with standard practice in field experiments, mentors were unaware of the experiment. To accommodate the large number of students, the same tutorial is taught simultaneously in multiple rooms, where each mentor 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 mentors teaching a total of eight sessions. Cohorts 2 and 3 both had 10 mentors, amounting to 20 sessions each. There were between six and nine student teams in each session.

In the tutorials, mentors 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 practicing entrepreneurs must 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 their final presentation.

3.2. Interventions

The experiment had a two × two factorial design, described in Table 1. In the *Team* treatment dimension, students were able to either choose their own team or they were assigned to a team of three members. In the *Idea* treatment dimension, students were able to either choose their own idea or they were randomly assigned to work on one of 15 predetermined 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. We denote the treatment groups that result from the two × two factorial design of the two treatment dimensions as *Choose neither*, *Choose team*, *Choose idea*, and *Choose both*. Students were randomly assigned to one of the four resulting treatment groups.

Table 1. Treatment Overview

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

To ensure adherence to treatment assignment and minimize contamination, the four treatment groups were separated by time and space. Temporal separation was accomplished based on early versus late time slots. The physical environment allowed for spatial separation as five tutorial classrooms were in the west wing of the building and five were in the east wing, with doors on either side of a stairwell separating them. Section A of the online appendix gives an example of this temporal and spatial division for Cohort 3.

Students had no particular preference for any time slot, with the first starting at 9:45 a.m. and the last ending at 5:15 p.m. The classrooms in the east and west wings are comparable. Nevertheless, 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 two different locations and two different time slots over the course of our experiment.

Section B of the online appendix 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. Therefore, for example, in Cohort 3 described in Section A of the online appendix, 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 mentors 1 to 5, and those in the *Choose team* treatment in the east wing with mentors 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 mentors 1 to 5 and *Choose neither* treatments in the east wing with mentors 6 to 10. Early and late time slots were on either side of noon, and students did not express any preference for particular time slots. The east and west wings were, similarly, indistinguishable. In our regressions, we use fixed effects to capture systematic differences across mentors. In the first week, tutorials were dedicated to team formation. During the early session, all students were informed, in their room by a mentor, 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 monitored by one of the mentors to ensure no mixing of treatments. Once all mentors had completed their introductions, students were released into 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 on a team together, they 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 30 minutes 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 mentor in their designated room.

In the second week, 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, mentors handed out the randomly assigned ideas to each team. Assigned ideas resembled a problem definition. They were sketched out on one page and steered 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; Section C of the online appendix provides an overview of all 15 ideas. Each of these ideas was extensively pretested to ensure 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: (1) solve a problem or add value; (2) reach a potential target market; and (3) 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 with a maximum of 10 slides. Other than this, no formal restrictions were placed on the students.

3.3. Evaluation

To avoid experimenter effects and to 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. 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 on 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 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.

For the evaluation, all evaluators were given a verbal and a written explanation (see Section D of the online appendix) of the procedure. The event took place in a large office building where every evaluator had their own spot at a large table. The event began by collecting general information on the evaluator and their relevant experience in a written survey. Following this survey, each pitch deck was individually evaluated. All pitch decks were printed, with an evaluation form as the cover sheet (see Section D of the online appendix). 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 ideas) was not evaluated by the same person twice, and (3) each person evaluated an

Table 2. Evaluator Expertise

Characteristic	Average per evaluator
No. of startups founded	0.88
No. of startups worked for	1.20
No. of startups coached	2.51
No. of startups funded with own money	0.65
No. of startups funded as part of a venture capitalist decision	0.42
No. of startups evaluated in the past	25.56
No. self-reported qualification for evaluating the pitch decks	0.79

equal number of pitch decks from all four treatments. We added a second heuristic by randomizing the order of the pitch decks each evaluator received. This averted the potential for temporal effects such as being less generous as time progresses (Danziger et al. 2011).

Pitch decks were assessed on six main criteria. Evaluators rated each on a seven-point Likert scale, separately, according to (1) novelty, (2) feasibility, and (3) market potential. As well, on a scale from 0% to 100%, evaluators had to assess (4) the project's likelihood of success and (5) the likelihood of inviting the team for a follow-up meeting. Finally, once all pitch decks had been separately assessed, the evaluators had to allocate (6) a fictional investment budget of $1 million across the entrepreneurship projects they evaluated. 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 mimicked the decisions of practicing venture capitalists. For this purpose, they were provided with stickers of fictional bills totaling $1 million dollars. They then had to distribute this "money" across different pitch decks by sticking the bills onto on a designated field in the pitch deck's evaluation sheet (see Section D of the online appendix).

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

Our experimental design deviates from a natural field experiment, because our subjects were students rather than organizational employees or practicing entrepreneurs. Although most of our students were close to career choices of naturally engaging in entrepreneurial activities and recent research has illustrated that the differences between students and managers are often exaggerated (Fréchette 2011, Bolton et al. 2012), the fact that our subjects are students means that the usual caveats regarding generalizability and external validity of this study's findings to real-world settings are warranted. That being said, our field experiment was designed to mimic a real organizational and entrepreneurial environment to the extent possible. The information provided in form of the training that the subjects brought to the experimental task, the task itself to develop a new venture project in form of a pitch deck, students' required commitment of time and effort, and the anticipated performance feedback very much resemble the procedures of today's early-stage entrepreneurial bootcamps and incubation programs. The evaluation was run in a dedicated environment with real venture capitalists and entrepreneurs who regularly evaluate pitch decks. Lastly, the experiment was run at a university with a pronounced entrepreneurial culture, where students regularly work on actual business challenges. The overaching goal of achieving real-world impact was also the framing for the entrepreneurship course in which the experiment was conducted. In light of these efforts, we classify our experiment as a *framed field experiment* according to the taxonomy of Harrison and List (2004, p. 1014).

4. Data and Methods

4.1. Sample

A total of 939 students enrolled in this class over three cohorts: 173 students in the first cohort, 408 in the second, and 359 in the third. These students ended up on a total of 310 teams. The majority, 295 teams, had three members, 12 teams had four members, and 3 teams had two members (to balance out participant numbers not divisible by three). Of these 939 students, 25 dropped out during the course. This left us with a final subject pool of 914 students on 310 teams, 274 of which had three members, 10 had four members, and 26 had two members. Team size and attrition are potentially endogenous to our treatments and can impact team performance, which we investigate further.

To register for the class, participants had to fill out an online baseline survey collecting information on their demographics, entrepreneurial experience, preferred team composition, and current skills (see Section E of the online appendix). After the final presentation in the last session, participants completed a written ending survey on their teamwork, satisfaction with the team and idea, new entrepreneurial skill development, and overall learning (see Section F of the online appendix). Students were unaware that they were part of an experiment. The baseline survey had a 100% response rate, whereas 891 of 914 students (97.5%) took the endline survey. Nonresponse was uncorrelated with treatment assignment. Missing values in the endline survey were imputed at the team means. Our main results are robust to the exclusion of observations, with missing values in the ending survey.

Section G of the online appendix provides summary statistics of balance checks regarding important pretreatment variables related to student demographics, as well as their task preferences, team member wishes, entrepreneurial exposure, self-efficacy, intention, and self-confidence, all obtained from the baseline survey. Overall, random assignment has worked well: student characteristics are sufficiently balanced across treatments. These pretreatment characteristics will serve as a basis to create measures of (endogenous) team compositions in the further analyses.

In addition to balance checks regarding individual characteristics, we check the balance of ideas among those treatments that involve assigning ideas (i.e., *Choose team* and *Choose neither*). The results of a Pearson's chi-squared test (χ^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 (χ^2 = 41.749, p = 1) and the 15 assigned ideas to the 40 evaluators (χ^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

4.2.1. Dependent Variable. Our dependent variables come 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 (Maxwell 2011, Carpentier and Suret 2015). The latter three variables are log-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 a mean of zero and a standard deviation of one for the baseline *Choose neither* group (for a similar approach, see Kling et al. 2007).

All evaluators assessed their assigned entrepreneurship projects regarding at least three indicators. In fact, only a handful evaluations were missing. One project-evaluator dyad had only three indicators evaluated; 11 dyads had four observed performance indicators; and 43 of 930 dyads had five of six observed indicators. The remaining dyads had all six performance indicators evaluated. Missing values were imputed at their respective means across all treatment groups (Kling et al. 2007). Our main results are also robust to the exclusion of observations with missing values in the evaluations. Thus, we end up with 930 dyads comprising 310 projects evaluated by three evaluators each. All in all, the six performance indicators have high internal consistency and reliability in measuring the project performance in our context (Cronbach's α = 0.85).

We start by running separate regressions for each performance indicator. We also define two aggregate performance measures to draw more general conclusions about the experimental results. First, we construct a performance index equal to an equally weighted average of z-scores of all six performance indicators (Kling et al. 2007). Second, instead of averaging, we "stack" the six performance indicators separately into one combined performance variable (Atkin et al. 2017), leading to a long data format with 5,580 observations (310 projects × 3 evaluations per project × 6 performance indicators), and include six dummy variables referring to a particular performance indicator. This stacked model allows us to detect differences in treatment effects for different performance indicators via interaction effects between treatment and performance indicator dummies, just as the separate regressions do. Mechanically constraining the interaction effects in this stacked model to zero yields the same coefficients obtained with the average performance index described previously, with the advantage that the estimates in the constrained stacked model produce more conservative estimates of statistical significance (i.e., higher coefficient standard errors).

4.2.2. 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 three dummy variables, each referring to one of the treatment groups: *Choose team*, *Choose idea*, and *Choose both*, with *Choose neither* excluded as the baseline.

4.3. Estimation

Estimation of treatment effects in randomized experiments basically involves comparing means across the different treatments. We condition the estimation of differences in means across treatments on some fixed effects to further limit confounding and increase precision. Our unit of analysis is the project-evaluator dyad, of which there are 930 (310 projects × 3 evaluations per project). This allows the inclusion of evaluator fixed effects, which is important because prior research has found wide variation in the assessment of early-stage ideas (see Boudreau et al. 2016 for a similar approach). By including 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 mentor who instructed a specific team. In each cohort, each mentor was responsible for two tutorial sessions, which implies they had students from two different treatments. Mentors were deliberately switched every cohort. Nevertheless, given that 23 mentors were responsible for tutoring 310 teams, mentors were not perfectly balanced across treatments. We therefore control for mentor fixed effects in our estimations.

We rely on ordinary least squares (OLS) estimates for our main results because of its ease of interpretation. To account for potential correlation among observations from the same team or evaluator, all standard error estimates are clustered at the team and evaluator levels. The formal specification of our performance regression model is the following

$$y_{ij} = \beta_0 + \beta_1 \times (Choose\ team)_i + \beta_2 \times (Choose\ idea)_i + \beta_3 \times (Choose\ both)_i + \boldsymbol{\gamma}_i + \boldsymbol{\delta}_i + \boldsymbol{\zeta}_{ij} + \boldsymbol{\eta}_j + \epsilon_{ij}, \quad (1)$$

where y_{ij} denotes the performance evaluation of team i as assessed by evaluator j, and *Choose team*, *Choose idea*, and *Choose both*, as described previously, are dummy variables denoting the treatment groups in which participants had the autonomy to choose team members, ideas, or both, respectively. Our parameters of interest are the β's, which capture treatment effects relative to the baseline of no autonomy over choice of team or idea. We control for potentially systematic differences in performance across different environments by including fixed effects for cohorts γ_i, mentors δ_i, evaluation order ζ_{ij}, and evaluators η_j; ϵ_{ij} is a random error component.

We estimate Equation (1) for each performance indicator separately. We also estimate it for the two aggregate performance indicators described earlier. First, with the data gathered in long format, we regress the stack of all six performance indicators on interactions of the treatment variables with the six dummies referring to each performance indicator. The F test with the null that the interaction terms are jointly equal to zero reveals that the treatments have heterogeneous effects on different performance measures. Furthermore, we obtain more conservative estimates of treatment effects as the clustered standard errors in this specification account for project- and evaluator-level correlations within and across performance indicators. We also interact the performance indicator dummies with the fixed effects of cohorts, mentors, order, and evaluators (Atkin et al. 2017).

Second, we regress an equally weighted performance index on the right-hand side of Equation (1). If this more parsimonious specification yields results that are substantively similar to those obtained from richer and more conservative specifications, we will stick with it in the causal mediation analysis. The files required for replicating the study and the analyses reported in the paper are avaliable under the Open Science Framework (https://osf.io/urjpe/) together with some instructions provided in Section K of the online appendix.

5. Results

5.1. Main Results

Table 3 reports the means (and variances) across treatment groups and a Kruskall-Wallis test of joint significance of mean differences assuming unequal variances. The sample sizes in each group reflect team-evaluator dyads. Although these descriptives are informative as a first overview of treatment effects, we aim to better capture the nonindependence of the repeated and cross-nested observations with respect to mentors, evaluators, evaluation order, and cohorts in regression analyses. Table 4 shows OLS parameter estimation results. Models 1–6 show the results for each performance indicator separately. Models 7 and 8 show the results for the stacked performance variable and the performance index, respectively.

The *Choose team* treatment effect is (marginally) significant overall. It has significant effects on *Market Potential* and *Invitation Probability*, whereas for *Success Potential*, it is marginally significant. The effect sizes in terms of the original measurement scales are as follows: *Market Potential* increases by 0.35 Likert scale points from 3.38 in the *Choose neither* treatment to 3.73 in the *Choose team* treatment. *Invitation Probability* increases by 59% from 23.91% to 38.05%. *Success Potential* increases by 31% from 22.99% to 30.15%.

The *Choose idea* treatment effect is strongly significant overall. It is also significantly larger than the *Choose team* treatment effect, except for *Feasibility* and *Invitation Probability*. The effect size is largest for *Novelty*, where the point estimate indicates an increase of 0.77 Likert scale points from 3.00 in the *Choose neither* treatment to 3.78 in the *Choose idea* treatment. *Investment* increases by 120% from $14,103 in the *Choose neither* treatment to $31,092 in the *Choose idea* treatment. The *Choose idea* treatment effect is, relatively, smallest for *Invitation Probability*, which increases by 92% from 23.91% to 46.02%.

Table 3. Mean (and Variance) Comparison Across Treatment Groups (Team-Evaluator Dyad Level)

	Choose neither (N = 234)	Choose team (N = 234)	Choose idea (N = 231)	Choose both (N = 231)	Overall p
Novelty	0.00 (1.00)	0.09 (1.04)	0.47 (1.16)	0.23 (1.13)	<0.001
Feasibility	0.00 (1.00)	−0.01 (1.00)	0.02 (1.02)	−0.04 (1.01)	0.937
Market potential (log)	0.00 (1.00)	0.03 (1.01)	0.14 (1.14)	0.02 (1.13)	0.466
Invitation probability (log)	0.00 (1.00)	0.08 (0.96)	0.14 (0.99)	0.11 (1.00)	0.447
Success potential	0.00 (1.00)	−0.07 (1.03)	0.07 (1.03)	−0.05 (1.04)	0.494
Investment (log)	0.00 (1.00)	0.09 (1.09)	0.36 (1.25)	0.18 (1.18)	0.005
Performance index	0.00 (1.00)	0.05 (1.03)	0.27 (1.13)	0.10 (1.10)	0.040

Table 4. Regression Results for the Treatment Effects on Performance

	Dependent variable							
	Novelty (1)	*Feasibility* (2)	*Market potential* (3)	*Success potential (log)* (4)	*Invitation probability (log)* (5)	*Investment (log)* (6)	*Stacked performance variable* (7)	*Performance index* (8)
Choose team (β_1)	0.088	0.225	0.223**	0.205*	0.266***	0.180	0.198*	0.198**
	(0.143)	(0.155)	(0.111)	(0.108)	(0.103)	(0.156)	(0.118)	(0.084)
Choose idea (β_2)	0.507***	0.213	0.418***	0.381***	0.375***	0.499***	0.399***	0.399***
	(0.152)	(0.159)	(0.116)	(0.131)	(0.132)	(0.166)	(0.129)	(0.099)
Choose both (β_3)	0.270**	−0.103	0.115	0.004	0.108	0.233**	0.104	0.104
	(0.119)	(0.092)	(0.088)	(0.057)	(0.079)	(0.103)	(0.085)	(0.068)
Performance indicator dummies	No	No	No	No	No	No	Yes	No
Cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mentor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluation order fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	930	930	930	930	930	930	5,580	930
R^2	0.295	0.296	0.397	0.562	0.526	0.227	0.376	0.407
p value: Choose team vs. idea	0.00***	0.89	0.04**	0.02**	0.12	0.00***	0.02**	0.00***
p value: Choose team vs. both	0.18	0.04**	0.40	0.06*	0.09*	0.74	0.43	0.27
p value: Choose idea vs. both	0.09*	0.03**	0.01***	0.00***	0.01**	0.11	0.01**	0.00***

Notes. Standard errors reported in parentheses are clustered at the team and evaluator levels. β_1, β_2, and β_3 are dummy variables equal to one for all participants in the treatments *Choose team*, *Choose idea*, and *Choose both*, respectively, with the treatment *Choose neither* as baseline comparison. Models 1–6 are OLS regression of each performance variable run separately on the treatments and controls. Model 7 is an OLS regression of all six performance variables stacked into one outcome variable. For each treatment, the interaction effects with dummy indicators for the six performance variables are restricted to be equal. These (restricted) interaction effects allowing the treatment effects (not) to vary across performance variables are jointly significant ($F(5052, 15) = 1.9361; p = 0.0161$). Model 8 averages the performance variables to an index in an unstacked regression.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The *Choose both* treatment effect is not significant overall. We detect significant effects only for *Novelty* and *Investment*. In the *Choose both* treatment, *Novelty* increases by 0.41 Likert scale points from 3.00 in the *Choose neither* treatment to 3.41. *Investment* increases by 45% from \$14,103 to \$20,399. Overall, the *Choose both* treatment is significantly inferior to the *Choose idea* treatment (except for *Investment*), and marginally inferior to the *Choose team* treatment, at least for the performance indicators *Feasibility*, *Success Potential*, and *Invitation Probability*.

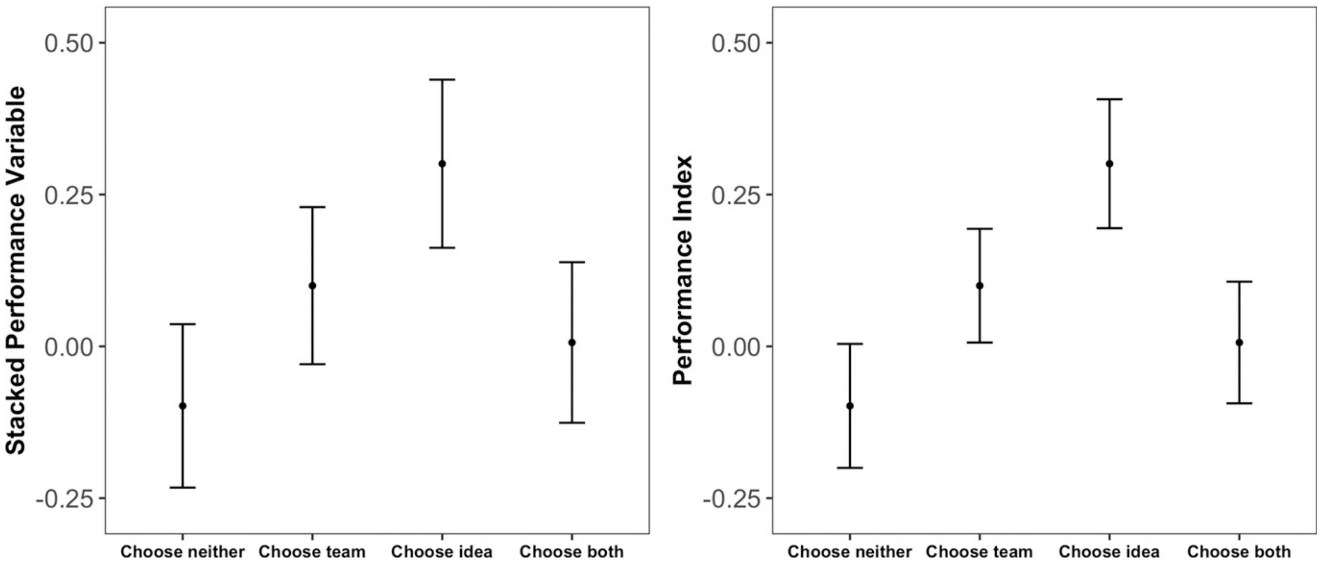
Figure 1 shows the predicted treatment effects from the OLS regression Models 7 and 8, along with their 95% confidence intervals. It becomes evident that Model 8, using the performance index, yields less conservative, that is, smaller, confidence intervals. Both models' predictions confirm graphically that the baseline treatment *Choose neither* results in the lowest performance. The *Choose idea* treatment, in contrast, consistently results in the highest performance. The performance of the *Choose team* treatment group is somewhere between the *Choose neither* and *Choose idea* treatment groups. Teams that could *Choose both* have performance outcomes statistically similar to the baseline *Choose neither* condition. Section H of the online appendix provides these plots for all performance indicators separately.

The results indicate that, contrary to what was hypothesized, the positive performance effects of *Choose team* and *Choose idea* are not complementary. On the contrary, they do not even add up. Both dimensions of autonomy over choice inhibit the others' special benefits when jointly present.

5.2. Robustness Checks

The robustness checks presented here address two concerns. First, it is reasonable to ask whether assigned ideas performed worse than endogenously generated ones simply because the assigned ideas were of poor quality. To allay this concern, we conducted an independent quality check of assigned and chosen ideas. Specifically, we randomly selected 15 of the generated ideas and had each written up twice in a paragraph by four different individuals hired expressly for this purpose to mimic the idea sketches handed out for the 15

Figure 1. Conditional Treatment Group Means (Predicted from Models 7 and 8 with 95% Confidence Intervals)



assigned ideas. In a second step, another person combined these paragraphs into a single document to ensure uniformity in the write-ups. We then had the 30 idea paragraphs evaluated on Mechanical Turk along the same performance criteria used in our main analysis (*Novelty*, *Market Potential*, and *Success Potential*, leaving out *Investment* because judging a possible investment should be done based on a pitch deck rather than simply the idea itself. A total of 20 evaluators each assessed 10 (of the 30) ideas, leaving us with a total of 200 evaluations of the 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 among evaluators and clustered the standard errors at the evaluator and idea level. (Idea fixed effects are not possible because 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 any of the three criteria ($p < 0.01$; see Section I of the online appendix). This indicates that our findings are not driven by ex ante differences in idea quality across assigned versus chosen idea treatments.

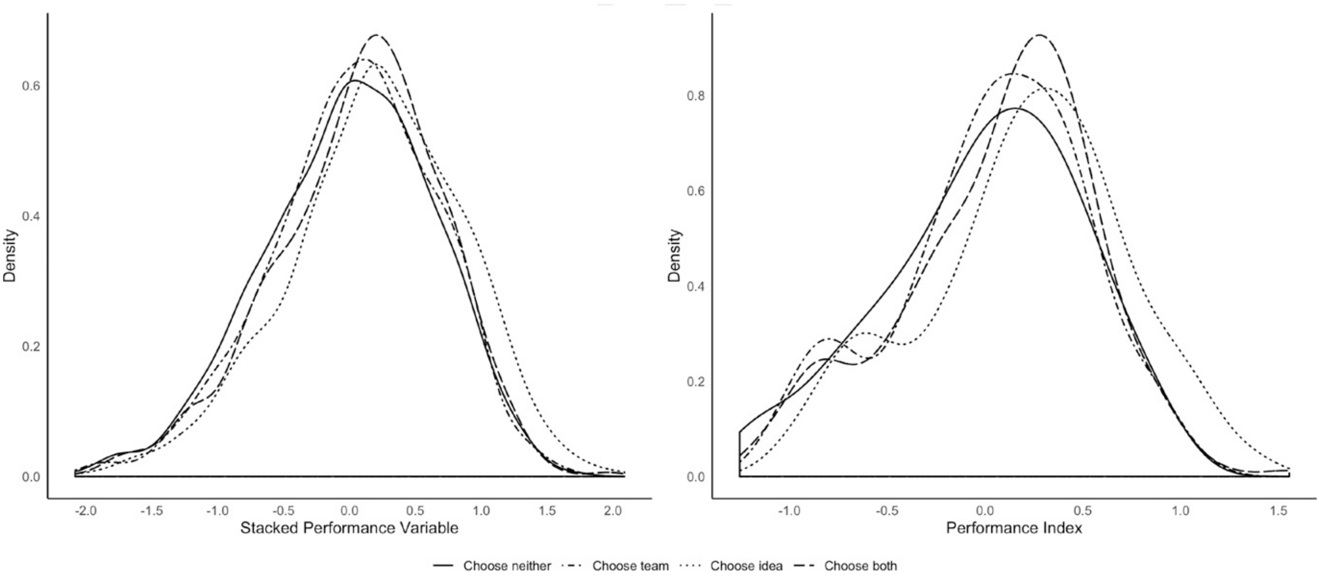
Second, a common concern in the entrepreneurship and innovation literature is the fact that average treatment effects might be driven by the tails of the distribution, meaning that very few highly innovative results skew findings in one direction (or likewise, that a few terrible outcomes diminish positive results). We address this concern first graphically. Figure 2 shows kernel densities of performance differentiated according to treatment group predicted from Models 7 and 8. These results suggest that the *Choose idea* treatment increases the average performance by shifting the action to the right tail of the distribution, while *Choose team* seems to avoid bad performance outcomes. Section H of the online appendix provides these plots for all performance indicators separately.

To investigate this further analytically, we use heteroscedastic regressions to analyze the dispersion. The results confirm that both the *Choose team* treatment ($p < 0.0448$) and the *Choose idea* treatment ($p < 0.0357$) also increase the variance in outcomes around the means. Furthermore, we use quantile regression analysis to explore heterogeneity in treatment effects in different quantiles of the performance distribution. Figure 3 shows the results from a quantile regression for the overall performance index. It confirms that the *Choose team* treatment is working more on the left-tail of the performance distribution. It can shield against bad outcomes between the 31st and 54th percentiles. The *Choose idea* treatment is effective in the right-tail of the performance distribution. It can boost performance in the region between the 26th and 89th percentiles. The *Choose both* treatment is also oriented toward the upside, but to a much lesser extent.

5.3. Understanding How Autonomy over Choice Matters: Causal Mediation Analysis

Thus far, we have analyzed overall treatment effects. The goal of this section is to decompose this into indirect and direct effects. By *indirect effects* we mean that treatment effects operate through intermediate mechanisms or channels. Concretely, autonomy over the choice of team members and ideas is likely to affect outcomes through two main channels: team composition and teamwork. The former pertains to the characteristics and prior relationships of team members. The latter refers to how (well) work is organized and

Figure 2. Kernel Densities of Performance per Treatment Group (Predicted from Models 7 and 8)

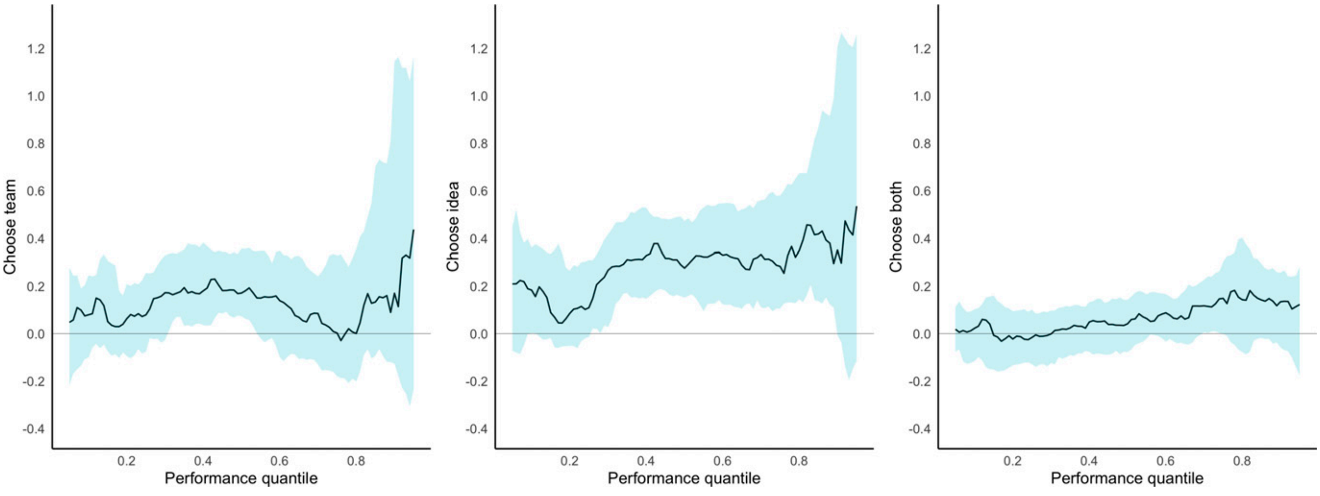


distributed in a team. Team composition and the quality of teamwork are complex constructs that we seek to operationalize with a variety of measures. For these measures to be causal mediators, they need to be impacted by our treatments in the first place. In Section J of the online appendix, we verify whether this is, in fact, the case by exploring how the posttreatment variables we consider as mediators are affected by the treatments. Section J.1 explores differences in mean outcomes pertaining to team composition and teamwork, and the remainder of Section J of the online appendix presents regression estimates of treatment effects on mediators, which we summarize briefly here.

5.3.1. Team Composition. Subjects in the *Choose team* treatment can select team members with whom they share similar characteristics or whom they know. Section J.2 confirms that homophily in terms of observed characteristics is manifested in our experimental treatments, which allow for autonomy over team choice, namely, the *Choose team* and *Choose both* treatments. These teams are significantly less diverse with respect to gender (column 1), tenure at the university (column 4), and study majors (column 5). The fact that there are no statistically significant differences between the treatments in terms of nationality (column 2) and age (column 3) reflects the fact that there is little underlying variation in these variables from this sample.

The last column in Section J.2 shows that team member relations based on prior network ties are significantly greater in the two treatments that allow for autonomy over team choice. Teams in the prevalent

Figure 3. (Color online) Treatment Effects over Quantiles in the Performance Distribution



Choose team and *Choose both* treatments have 15% and 13% more ties, respectively, realized from a list of five social contacts they expressed a preference to work with in the baseline survey prior to treatment assignment. In a three-person team with six potential directed ties, this increase corresponds to almost one additional realized tie that was desired in advance. In short, prior network ties steer participants' choices of team members.

Section J.3 of the online appendix reveals that assortative matching based on observable characteristics also partly leads to higher homogeneity among team members in terms of more latent traits. In our context, entrepreneurial traits and skills are likely correlated with observable demographics such as tenure and study major. Accordingly, diversity with respect to entrepreneurial intent (column 4) and self-confidence about own performance (column 5) is lower among teams in the *Choose team* treatment. Following the logic of homophilous team formation, participants who can choose team members are less strategic about seeking complementarity in skills and interests among team members. This is manifested in our setting by a lower task preference complementarity (column 1) in the *Choose team* treatment, which is calculated with the Hungarian assignment algorithm (Kuhn 1955, Munkres 1957) based on participants' preferences for certain subjects in a business plan. In Section J.4, we test whether the sorting tendencies induced by the treatments led not only to lower variances in team members' entrepreneurial traits, but also to changes in the means of their entrepreneurial traits. We find no systematic evidence in this regard.

5.3.2. Teamwork. In Section J.5 of the online appendix, we explore how teams evaluate their teamwork ex post, contingent on treatment. As a precursor to good teamwork, we measure how well the chosen project idea matched their interests and capabilities. We find that the match is significantly higher in the autonomy of idea choice treatments: the coefficient estimates in column 1 indicate a 1-point Likert scale increase in the *Choose idea* treatment and a 1.5-point increase in the *Choose both* treatment. Treatment effects pertaining to teamwork quality (column 2) are not significantly different, but collaboration intensity (column 2) is (around half a Likert scale point) higher in the autonomy of team choice treatments: *Choose team* and *Choose both*. Although course attendance, as one possible proxy for effort, does not show any significant differences (column 5), the extent to which team members overstate their contribution to the teamwork (column 4) is on average 11 percentage points lower in the *Choose team* treatment. Finally, satisfaction with teamwork is high enough in the *Choose both* treatment to reduce the likelihood of a team member leaving a team by 9 percentage points (column 6).

Overall, the treatment effects on intermediate outcomes are in line with theoretical priors. The key question we wish to answer here is whether these intermediate outcomes constitute indirect paths through which performance is ultimately affected. The *average causal mediation effect (ACME)* quantifies these indirect paths. It examines how changes in the mediator variable, induced by the treatments, affect performance while (counterfactually) holding the treatment effects on performance fixed. ACME is identified under the *sequential ignorability* assumption (Imai et al. 2010), which states that both the treatments and the mediator are statistically independent with respect to performance once we control for pretreatment covariates. Although exogeneity of treatments is assured by our randomized controlled design, we support the exogeneity of mediators by also controlling for pretreatment variables unaffected by the treatment. Specifically, we control for team averages of pretreatment entrepreneurial exposure, efficacy, and intent, as well as self-confidence, in our causal mediation analysis. Although these team-level averages are largely unaffected by the treatments in our setting, technically they arise after team formation, that is, after treatment. Therefore, as robustness checks, we also run causal mediation analyses at the team level without these controls and at the individual level, where these controls can be integrated as clearly exogeneous. The team level results reported in Table 5 are robust across these three specifications. Standard errors are clustered at the team level and determined with a quasi-Bayesian Monte Carlo method based on normal approximation with 1,000 simulations (Imai et al. 2010).

We find systematic causal mediation effects for only two mediators. First, we find evidence that the realization of prior network ties mediates the *Choose team* and *Choose both* performance treatment effects. This mediation holds in particular for the performance indicators *Success Potential* and *Investment*. The proportion of the overall treatment effect that is mediated by realization of prior ties in the *Choose team* treatment is about 18.4%; for *Success Potential* and *Investment* the corresponding proportions are 20.5% and 34.1%, respectively. Second, the average fit between the project idea and team members' interests and skills significantly mediates the *Choose idea* and *Choose both* treatments. This mediation is especially strong for the performance indicators *Market Potential*, *Success Potential*, *Invitation Probability*, and *Investment*. The proportion of the total *Choose idea* effect mediated by team-idea fit ranges from 13.2% to 16.5%. For *Choose both*, the proportion of the overall performance effect mediated by team-idea fit is nearly 60%.

Two conclusions can be drawn from the causal mediation results. First, the mediation effects explain

Table 5. Results from Causal Mediation Analysis

Mediator	Performance indicator	Treatment	ACME	p value	Proportion mediated
Network ties (ratio)	Performance index	Choose team	0.039	0.076	0.184
		Choose both	0.032	0.060	0.261
	Success potential	Choose team	0.044	0.054	0.205
		Choose both	0.037	0.058	−0.074
	Investment	Choose team	0.078	0.034	0.341
		Choose both	0.068	0.036	0.280
Idea team fit (mean)	Performance index	Choose idea	0.052	0.024	0.132
		Choose both	0.069	0.032	0.580
	Market potential	Choose idea	0.071	0.016	0.165
		Choose both	0.094	0.012	0.671
	Success potential	Choose idea	0.051	0.034	0.137
		Choose both	0.070	0.050	0.131
	Invitation probability	Choose idea	0.064	0.018	0.160
		Choose both	0.086	0.012	0.667
	Investment	Choose idea	0.073	0.030	0.143
		Choose both	0.095	0.042	0.382

only a small proportion of total treatment effects. The larger proportions run through direct paths. Heckman and Pinto (2015) point out that direct effects can be attributed to two sources: (1) changes in unobserved inputs induced by the treatments and (2) unobserved changes in the way inputs are transformed into outputs. In our setting, the latter source corresponds to unobserved changes in teamwork processes that allow teams to better translate team members' (pretreatment) motivations and capabilities into performance. Teams that are formed endogenously, based on similarity or familiarity, may work together more efficiently, coordinate more smoothly, and have lower communications costs. Our (self-reported) measures of teamwork quality may capture this inadequately.

The treatments in our setting are also likely to induce unobserved changes in team members' inputs. In line with our theoretical arguments, the autonomy of choosing can have intrinsic value related to the process of choosing itself rather than its instrumental outcomes (such as "I can work on a project and with friends I like"). This procedural utility (Benz and Frey 2008) may be an intrinsic source of motivation that inspires additional effort among team members. Our posttreatment questions regarding motivation are nonverifiable and prone to measurement error. Against this background, we attribute (parts of) the direct treatment effects to intrinsic motivation inspired by autonomy of choice, which inspires additional effort, and could explain the performance effect.

A second conclusion is that we do not find any convincing mediation effects to explain why performance drops in the *Choose both* compared with the *Choose idea*, and to some degree also the *Choose team*, treatments. There is some weakly significant evidence in the data that a lack of team diversity in the number of study majors hinders teams in fully exploiting the potential of developing and elaborating their own project ideas. However, this does not fully explain the performance drop. This leads us to an alternative explanation based on overconfidence, which may be triggered in the *Choose both* treatment, where the degree of autonomy is highest.

5.3.3. Overconfidence. Granting autonomy to individuals and teams increases their perceptions and feelings of efficacy and confidence. As we argued earlier, this can have a positive motivational effect. However, a number of researchers have raised concerns that teams in which efficacy rises above a critical threshold are likely to fall prey to overconfidence, exhibiting complacency and a lack of focus (Gist 1987, Sitkin 1992, Lindsley et al. 1995, Knight et al. 2001, Goncalo et al. 2010, Rapp et al. 2014). This can lead to the allocation of insufficient effort toward task completion. The difficulty is in empirically distinguishing between the positive motivational effects of confidence and the complacency of overconfidence. In fact, Knight et al. (2001) conclude that "studies need to identify where healthy confidence leaves off and foolish overconfidence begins" (p. 336).

Initial satisfaction and easy triumphs, such as having your friend on the team or coming up with your own idea, can trigger an upward "efficacy-performance spiral" (Lindsley et al. 1995) that can ultimately lead to overconfidence and complacency (Moore and Healy 2008, Goncalo et al. 2010). Early success can reduce follow-on effort and experimentation, as well as search for and attention to external feedback (Sitkin 1992). Especially among familiar teammates, there is a tendency toward internal self-assurance rather than paying attention to external task demands and performance standards (Moore and Healy 2008, Rapp et al. 2014). This line of argument points to the possibility that teams in our *Choose both* treatment group

Table 6. Perceived Performance and Overconfidence

	Dependent variable			
	Evaluator: performance index (1)	*Evaluator: success potential (in %)* (2)	*Team: success potential (in %)* (3)	*Overconfidence: share of team members overplacing own project* (4)
Choose team (β_1)	0.182**	4.463*	−6.594	−0.288
	(0.082)	(2.574)	(4.509)	(0.214)
Choose idea (β_2)	0.383***	9.375***	−1.162	−0.320
	(0.097)	(2.816)	(4.373)	(0.213)
Choose both (β_3)	0.103	2.100	13.840***	0.404***
	(0.066)	(1.489)	(2.620)	(0.114)
Entrepreneurial exposure (mean)	0.062	2.186	0.606	−0.081
	(0.040)	(1.407)	(1.701)	(0.073)
Entrepreneurial self-efficacy (mean)	0.030	0.339	2.325	0.093
	(0.037)	(0.950)	(1.783)	(0.075)
Entrepreneurial intent (mean)	−0.003**	−0.090*	0.119	0.008**
	(0.001)	(0.049)	(0.084)	(0.003)
Pre-confidence (mean)	−0.001	−0.324	3.677**	0.171**
	(0.032)	(1.214)	(1.651)	(0.081)
Cohort fixed effects	Yes	Yes	Yes	Yes
Mentor fixed effects	Yes	Yes	Yes	Yes
Evaluation order fixed effects	Yes	Yes	No	Yes
Evaluator fixed effects	Yes	Yes	No	Yes
Observations	930	930	310	930
R^2	0.411	0.506	0.287	0.391

Note. Standard errors reported in parentheses are clustered at the team and evaluator levels.
$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

were experiencing too much confidence too soon, which might have reduced their follow-on effort.

To provide initial empirical evidence for this explanation based on overconfidence, we disentangle the amount of posttreatment confidence that has been instilled in teams through the treatments by controlling for pretreatment sources of confidence. Our measure of posttreatment confidence is the same that we used for the evaluators to judge a project's *Success Potential*, so these two measures are closely comparable. A common way to measure overconfidence in the entrepreneurship literature is through overplacement (Moore and Healy 2008, Gutierrez et al. 2020), which refers to an erroneous belief in one's performance or abilities compared with a reference group. We operationalize this construct in our context as the proportion of team members on a given team whose own judgement about their project's *Success Potential* is placed above the judgment they received from a given evaluator with respect to the same performance indicator.

Table 6 shows the results of this analysis. Model 1 regresses the overall performance index on the treatments, but this time with the team-level controls of prior entrepreneurial confidence included. The base treatment effects on performance still hold, even with these additional controls. Model 2 regresses only the performance indicator *Success Potential* on treatments and controls. The evaluators' assessment of *Success Potential* in the *Choose team* and *Choose idea* treatment groups are 4.5 and 9.4 percentage points higher, respectively. Model 3 shows the results for *Success Potential* as perceived by the teams. Relative to the baseline, teams' confidence in their own projects remains unchanged in the *Choose team* and *Choose idea* treatments, but increases sharply in the *Choose both* treatment group, by nearly 14 percentage points. Rather than an incremental increase in confidence together with levels of autonomy, we see a sudden jump in confidence that occurs only at very high levels of autonomy. Model 4 shows that this jump in confidence leads to more than 40 percentage points of members of teams in the *Choose both* treatment group, placing their own project higher than the evaluation they eventually received from evaluators.

6. Discussion

Autonomy has been at the heart of organizational theory for decades (Pelz and Andrews 1966, Amabile and Gitomer 1984, Bailyn 1985). It has, however, typically been analyzed along a single dimension or assumed that autonomy is about giving complete freedom. By studying two fundamental ways in which autonomy can be granted either by choosing teams or choosing ideas, we garner new insights about the effects of autonomy on entrepreneurial team performance. These insights are gleaned from a field experiment whose major appeal is that it allows for causal inference while having a sufficiently large sample size to allow for meaningful statistical inference: our experiment involved 939 participants and more than 310 teams in a lean startup course over an 11-week period. The participants were randomly assigned to one of four treatments in a two-by-two experimental design 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. Our findings suggest that teams with autonomy over choosing either ideas or team members outperform teams in the baseline treatment as measured by pitch deck performance. The effect of choosing *ideas* is significantly stronger than the effect of choosing *teams*. We find, however, that the two forms of autonomy are substitutes. We elaborate on the implications of these findings below.

6.1. Theoretical Implications

Instead of exploring individual-level outcomes, which is often done in the autonomy literature, we turn our attention to outcomes at the team level. This shift is important as it provides caveats to the literature on entrepreneurial teams, which often studies teams where more autonomy is the norm. By comparing teams with exogenous variation in autonomy along two dimensions, we can ask additional questions of theoretical importance. Our design enables us to draw conclusions about the effectiveness of each dimension separately, and whether autonomy is desirable along both dimensions simultaneously. Our findings suggest that autonomy on either dimension dominates having choice on neither dimension in terms of pitch deck performance. This is consistent with the literature on autonomy (Pelz and Andrews 1966, Amabile and Gitomer 1984, Bailyn 1985, Criscuolo et al. 2014, Gambardella et al. 2020). At the same time, what kind of autonomy 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.

A striking result is 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. In fact, the performance gains vanish for teams that are granted full autonomy over choosing both ideas and teams. This suggests that there is a substitution effect between the two forms of autonomy. Our results show that this overall pattern (i.e., a substitution effect between choosing an idea and choosing the team) is present for all six outcome variables. This pattern is most pronounced for the outcome measure of feasibility and success potential and a little less pronounced for novelty and investment probability. A plausible interpretation is that the novelty of a chosen idea may be driven by individual team members' contributions, whereas development of this idea as a compelling business proposition (in terms of feasibility and success potential) involves more of a team effort. If the team does not function properly together, the potential of a novel idea may not be fully exploited.

Using causal mediation analyses (Imai et al. 2010, Heckman and Pinto 2015), we further explore channels through which the main treatment effects operate. Detailed baseline and endline surveys, with nearly complete response rates, allowed us to investigate numerous potential mechanisms. Regarding autonomy over choosing teams, we find that 18% of the main treatment effect was driven by choosing team members from one's prior network. The causal mediation analysis also reveals that treatment effects resulting from autonomy over idea choice is driven by a better match of skills to team members. More precisely, the proportion of the total *Choose idea* effect mediated by team-idea fit ranges from 13.2% to 16.5%. Moreover, for the people granted the most autonomy, in the *Choose both* treatment, the proportion of the overall performance effect mediated by team-idea fit is nearly 60%. The nil effects pertaining to other mediators are noteworthy. Importantly, although autonomy of team choice promotes homophily and sorting on skills, these intermediate outcomes do not have a mediating effect on the performance treatment effect.

Overall, the mediation effects explain a relatively small proportion of total treatment effects. Rather, the treatment effects run through direct paths. Heckman and Pinto (2015) point out that direct effects can be attributed to two sources: (1) changes in unobserved inputs induced by the treatments and (2) unobserved changes in the way inputs are transformed into outputs. An implication of the causal mediation analysis is that a large chunk of the variance cannot be explained by the plausible mechanisms we explored. Identification of mechanisms is often harder in field experiments than laboratory experiments, with their short time spans and controlled environments. Teams in our experiment were followed in their natural setting over an 11-week period, during which many unobserved factors around team dynamics and satisfaction no doubt evolved.

Although granting autonomy can have a positive motivational effect, our results regarding the substitution effect between the two forms of autonomy adds to recent research on the costs of autonomy (Criscuolo et al. 2014, Gambardella et al. 2020). In teams granted autonomy along both dimensions, confidence may rise above a critical threshold and lead to overconfidence, accompanied by complacency and a lack of focus (Gist 1987, Sitkin 1992, Lindsley et al. 1995, Knight et al. 2001, Goncalo et al. 2010, Rapp et al. 2014). By measuring overconfidence through overplacement (Moore and Healy 2008, Gutierrez et al. 2020), where we compare teams' own judgments with those of evaluators, we can see systematic differences between the treatments. The teams granted the most autonomy over the *Choose both* treatment are the most overconfident. One plausible interpretation is that teams care about the process as much as the outcome. Indeed, research on procedural justice has illustrated that people care about procedures and being involved in the process rather than just the outcomes (Lind and Tyler 1988). In entrepreneurship, it is entirely possible that having autonomy over choice can generate the procedural utility (Benz and Frey 2008) that generates confidence or even overconfidence. This was evident in our experiment: the treatment with the most autonomy was overconfident as group measured by overplacement and were likely to continue with pitch decks that investors deemed subpar. 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 for it (Moskowitz and Vissing-Jørgensen 2002). In other words, too much autonomy may have a downside.

6.2. Managerial Implications

The findings in this paper are pertinent to the "professionalization of entrepreneurship," particularly through incubator and accelerator programs, and for organizations such as Valve and Github, which have experimented with "boss-less" organizations to provide lots of autonomy. Most accelerators and incubators give aspiring entrepreneurs choice on both the idea and team dimensions. Other companies are reducing autonomy, or providing full autonomy for a limited time period, as 3M does. Our results indicate that granting autonomy over choosing ideas leads to the highest performance. This precludes assortative matching and may detract from personal happiness generated from social interactions. However, it is likely to generate the kind of environment in which better ideas can flourish and translate into more successful entrepreneurial team performance.

6.3. Limitations and Future Research

The experimental design and main results presented here followed a preanalysis plan, which was preregistered at the Social Science Registry of the American Economic Association. This setup enables us to draw causal inference without being susceptible to "p-hacking". However, our choices in design of the experiment introduce tradeoffs, which we elaborate on here.

Although our framed field experiment (Harrison and List 2004) was designed to mimic a real organizational and entrepreneurial environment for the experimental task and evaluation to the largest possible extent, a drawback of our setting is that the subjects themselves were students and not practicing employees or entrepreneurs. Although most of our students were close to these career choices and recent research has illustrated that the differences between students and managers are often exaggerated (Fréchette 2011, Bolton et al. 2012), the fact that the experimental subjects are students means that the usual caveats regarding generalizability of this study's findings to real-world organizational and entrepreneurial settings are warranted.

The pragmatic advantages to the experimental setting were threefold. First, it allowed us to conduct a field experiment with a large-enough sample size to allow for statistically meaningful inference to detect effect sizes that are reasonable across the distribution. Second, it is difficult to find settings that allow for experimentation to occur, especially with hundreds of participants. Third, as organizational theorists know, it takes time for the dust to settle and organizational issues to emerge. As a result, we wanted to know how autonomy plays out over a longer time period, where coordination problems and conflicts may arise. Our setting allowed us to study outcomes of pitch decks after 11 weeks rather than immediate outcomes in a laboratory. Of course, the challenge that arises over this relatively long period of observation is that the dynamics of teams and ideas evolve over time, generating mechanisms that may not be observed and hence cannot be accounted for in the causal mediation analysis.

A generic challenge in experiments is the potential for contamination across treatment groups. We sought to minimize this by carefully separating experimental conditions across time and space. Moreover, neither mentors 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. We kept all communication with the students, and only one asked why the students had different amounts of autonomy.

Our approach compares teams who are permitted to choose 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, which organizational theorists have suggested is a tough baseline to beat (Clement and Puranam 2018). In real life, organizations do not form teams at random; they form teams in the hope of achieving superior performance. There are several ways such teams could form, such as through maximizing functional and social diversity (Lamaze and Van Knippenberg 2014) or allowing team members to have some prior connections (Reagans et al. 2004). An interesting future line of research would be to compare different such baselines rather than randomly assigned teams or comparing teams formed through self-selection with those assigned by managers. This would be a question of great importance for scholars of self-organizing operations (Puranam et al. 2014, Bernstein et al. 2017).

Last, students who were given the autonomy to choose their collaborators only had 60 minutes to find their team members, with many teams actually forming well before this deadline: on average, a team was composed in 13 minutes. Although this time constraint is comparable to entrepreneurial mixer events, many employees with autonomy over their collaborators have more time on their hands. Future research could focus on the influence of the amount of time spent on choosing your collaborators.

6.4. Conclusion

We demonstrate the importance of considering autonomy over choice of team members and ideas jointly and separately and show that more autonomy does not uniformly lead to higher entrepreneurial team performance as measured by pitch deck performance. These effects operate through the mechanisms of better matching of ideas to team members' skills for choosing ideas and letting personal networks take precedence over the best collaborators for choosing team members.

Acknowledgments

The authors thank the special issue editors, three superb reviewers, Michael Bikard, Gianluca Carnabuci, Thorsten Grohsjean, Helge Klapper, Robin Kleer, Hannes Lampe, Matt Marx, Markus Perkmann, Henning Piezunka, Keyvan Vakili, Alexander Vossen, and seminar and conference participants at Aarhus University, DRUID, Erasmus University, ESMT Berlin, EURAM, Hamburg University of Technology, INSEAD, the online Organization Science special issue conference, Oxford University, Singapore Management University, University College London, and the University of Copenhagen for constructive feedback. Dimitri Graf and Albert Thieme provided excellent research assistance. All errors are the authors' own.

References

Amabile TM, Gitomer J (1984) Children's artistic creativity: Effects of choice in task materials. *Personality Soc. Psych. Bull.* 10(2): 209–215.

Amore MD, Garofalo O, Martin-Sanchez V (2021) Failing to learn from failure: How optimism impedes entrepreneurial innovation. *Organ. Sci.* 32(4):940–964.

Atkin D, Khandelwal AK, Osman A (2017) Exporting and firm performance: Evidence from a randomized experiment. *Quart. J. Econom.* 132(2):551–615.

Bailyn L (1985) Autonomy in the industrial r&d laboratory. *Human Resource Management* 24(2):129–146.

Barley SR, Kunda G (2001) Bringing work back in. *Organ. Sci.* 12(1): 76–95.

Benz M, Frey BS (2008) The value of doing what you like: Evidence from the self-employed in 23 countries. *J. Econom. Behav. Organ.* 68(3-4):445–455.

Bernstein S, Korteweg A, Laws K (2017) Attracting early-stage investors: Evidence from a randomized field experiment. *J. Finance* 72(2):509–538.

Bernstein E, Bunch J, Canner N, Lee M (2016) Beyond the holacracy hype. *Harvard Bus. Rev.* 94(7):13.

Biancani S, McFarland DA, Dahlander L (2014) The semiformal organization. *Organ. Sci.* 25(5):1306–1324.

Bolton GE, Ockenfels A, Thonemann UW (2012) Managers and students as newsvendors. *Management Sci.* 58(12):2225–2233.

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 Sci.* 62(10):2765–2783.

Brannon DL, Wiklund J, Haynie JM (2013) The varying effects of family relationships in entrepreneurial teams. *Entrepreneurial Theory Practice* 37(1):107–132.

Brooks AW, Huang L, Kearney SW, Murray FE (2014) Investors prefer entrepreneurial ventures pitched by attractive men. *Proc. Natl. Acad. Sci. USA* 111(12):4427–4431.

Burgelman RA (1983) Corporate entrepreneurship and strategic management: Insights from a process study. *Management Sci.* 29(12):1349–1364.

Burt RS (2005) *Brokerage and Closure: An Introduction to Social Capital*, 1st ed. (Oxford University Press, New York).

Burt RS, Merluzzi J (2016) Network oscillation. *Acad. Management Discovery* 2(4):368–391.

Camuffo A, Cordova A, Gambardella A, Spina C (2020) A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Sci.* 66(2):564–586.

Carpentier C, Suret JM (2015) Angel group members' decision process and rejection criteria: A longitudinal analysis. *J. Bus. Venturing* 30(6):808–821.

Cattani G, Ferriani S, Mariani MM, Mengoli S (2013) Tackling the "galácticos" effect: Team familiarity and the performance of star-studded projects. *Industria Corporate Change* 22(6):1629–1662.

Clement J, Puranam P (2018) Searching for structure: Formal organization design as a guide to network evolution. *Management Sci.* 64(8):3879–3895.

Clingingsmith D, Shane S (2018) Training aspiring entrepreneurs to pitch experienced investors: Evidence from a field experiment in the United States. *Management Sci.* 64(11): 5164–5179.

Cooper DJ, Saral KJ (2013) Entrepreneurship and team participation: An experimental study. *Eur. Econom. Rev.* 59:126–140.

Criscuolo P, Salter A, Ter Wal AL (2014) Going underground: Bootlegging and individual innovative performance. *Organ. Sci.* 25(5):1287–1305.

Dahlander L, O'Mahony S (2011) Progressing to the center: Coordinating project work. *Organ. Sci.* 22(4):961–979.

Danziger S, Levav J, Avnaim-Pesso L (2011) Extraneous factors in judicial decisions. *Proc. Natl. Acad Sci. USA* 108(17): 6889–6892.

Dean DL, Hender JM, Rodgers TL, Santanen EL (2006) Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation. *J. Assoc. Inform. Systems* 7(10):646–699.

Fama EF, Jensen MC (1983) Separation of ownership and control. *J. Law Econom.* 26(2):301–325.

Foss NJ (2003) Selective intervention and internal hybrids: Interpreting and learning from the rise and decline of the Oticon spaghetti organization. *Organ. Sci.* 14(3):331–349.
Fréchette GR (2011) Laboratory experiments: Professionals vs. students. Preprint, submitted XX, https://dx.doi.org/10.2139/ssrn.1939219.
Gallus J (2017) Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia. *Management Sci.* 63(12):3999–4015.
Gambardella A, Khashabi P, Panico C (2020) Managing autonomy in industrial research and development: A project-level investigation. *Organ. Sci.* 31(1):165–181.
Gist ME (1987) Self-efficacy: Implications for organizational behavior and human resource management. *Acad. Management Rev.* 12(3):472–485.
Goncalo JA, Polman E, Maslach C (2010) Can confidence come too soon? Collective efficacy, conflict and group performance over time. *Organ. Behav. Human Decision Processes* 113(1):13–24.
Gray SM, Knight AP, Baer M (2020) On the emergence of collective psychological ownership in new creative teams. *Organ. Sci.* 31 (1):141–164.
Gulati R, Gargiulo M (1999) Where do interorganizational networks come from? *Amer. J. Sociol.* 104(5):1439–1493.
Gutierrez C, Åstebro T, Obloj T (2020) The impact of overconfidence and ambiguity attitude on market entry. *Organ. Sci.* 31(2):308–329.
Hackmann JR (2002) *Leading Teams: Setting the Stage for Great Performances* (Harvard Business Press, Cambridge, MA).
Hackman JR, Oldham GR (1975) Development of the job diagnostic survey. *J. Appl. Psych.* 60(2):159.
Hamilton BH (2000) Does entrepreneurship pay? An empirical analysis of the returns of self-employment. *J. Political Econom.* 108 (3):604–631.
Hampel C, Perkmann M, Phillips N (2020) Beyond the lean start-up: Experimentation in corporate entrepreneurship and innovation. *Innovation* 22(1):1–11.
Hansen Z, Owan H, Pan J (2015) The impact of group diversity on class performance: Evidence from college classrooms. *Edu. Econom.* 23(2):238–258.
Harrison GW, List JA (2004) Field experiments. *J. Econom. Literature* 42(4):1009–1055.
Hasan S, Koning R (2019) Prior ties and the limits of peer effects on startup team performance. *Strategic Management J.* January: 1394–1416.
Heckman JJ, Pinto R (2015) Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Rev.* 34(1–2):6–31.
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 Sci.* 59(7):1514–1528.
Hoogendoorn S, Parker SC, van Praag M (2017) Smart or diverse start-up teams? Evidence from a field experiment. *Organ. Sci.* 28(6):1010–1028.
Horwitz SK, Horwitz IB (2007) The effects of team diversity on team outcomes: A meta-analytic review of team demography. *J. Management* 33(6):987–1015.
Huang L, Pearce JL (2015) Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions. *Admin. Sci. Quart.* 60(4):634–670.
Huckman RS, Staats BR, Upton DM (2009) Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Sci.* 55(1):85–100.
Imai K, Keele L, Tingley D (2010) A general approach to causal mediation analysis. *Psych. Methods* 15(4):309.
Ingram P, Morris MW (2007) Do people mix at mixers? Structures, homophily, and the life of the party. *Admin. Sci. Quart.* 52(4):14–26.
Kanter R (1985) Supporting innovation and venture development in established companies. *J. Bus. Venturing* 1(1):47–60.
Klapper H, Reitzig M (2018) On the effects of authority on peer motivation: Learning from Wikipedia. *Strategic Management J.* 39 (8):2178–2203.
Kleinbaum AM, Stuart TE, Tushman ML (2013) Discretion within constraint: Homophily and structure in a formal organization. *Organ. Sci.* 24(5):1316–1336.
Kling JR, Liebman JB, Katz LF (2007) Experimental analysis of neighborhood effects. *Econometrica* 75(1):83–119.
Knight D, Durham CC, Locke EA (2001) The relationship of team goals, incentives, and efficacy to strategic risk, tactical implementation, and performance. *Acad. Management J.* 44(2): 326–338.
Kuhn HW (1955) The Hungarian method for the assignment problem. *Naval Res. Logist. Quart.* 2(1–2):83–97.
Lamaze A, Van Knippenberg D (2014) Teams in pursuit of radical innovation: A goal orientation perspective. *Acad. Management Rev.* 39(4):423–438.
Lee GK, Cole RE (2003) From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organ. Sci.* 14(6):633–649.
Levine SS, Prietula MJ (2014) Open collaboration for innovation: Principles and performance. *Organ. Sci.* 25(5):1414–1433.
Lind EA, Tyler TR (1988) *The Social Psychology of Procedural Justice* (Springer Science & Business Media, New York).
Lindsley DH, Brass DJ, Thomas JB (1995) Efficacy-performing spirals: A multilevel perspective. *Acad. Management Rev.* 20(3):645–678.
Liu CC, Srivastava SB, Stuart TE (2016) An intraorganizational ecology of individual attainment. *Organ. Sci.* 27(1):90–105.
Lovas B, Ghoshal S (2000) Strategy as guided evolution. *Strategic Management J.* 21(9):875–896.
Lumpkin GT, Cogliser CC, Schneider DR (2009) Understanding and measuring autonomy: An entrepreneurial orientation perspective. *Entrepreneurial Theory Practice* 33(1):47–69.
MacCormack A, Baldwin C, Rusnak J (2012) Exploring the duality between product and organizational architectures: A test of the "mirroring" hypothesis. *Res. Policy* 41(8):1309–1324.
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.
Maxwell A (2011) Business angel decision making. *J. Bus. Venturing* 26(2):212–225.
McKenzie D (2017) Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition. *Amer. Econom. Rev.* 107(8):2278–2307.
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* 27:415–444.
Miner AS (1994) Seeking adaptive advantage: Evolutionary theory and managerial action. Baum J, Singh JV, eds. *Evolutionary Dynamics Organizations.* 76:89.
Moore DA, Healy PJ (2008) The trouble with overconfidence. *Psychol. Rev.* 115(2):502.
Moreland RL, Argote L (2003) Transactive memory in dynamic organizations. Day RD, ed. *Leading Managing People Dynamic Organizations.* 135:162.
Moskowitz TJ, Vissing-Jørgensen A (2002) The returns to entrepreneurial investment: A private equity premium puzzle? *Amer. Econom. Rev.* 92(4):745–748.
Munkres J (1957) Algorithms for the assignment and transportation problems. *J. Soc. Industry Appl. Math.* 5(1):32–38.
Pelz DC, Andrews FM (1966) Autonomy, coordination, and stimulation, in relation to scientific achievement. *Behav. Sci.* 11(2):89–97.
Pierce JL, Kostova T, Dirks KT (2001) Toward a theory of psychological ownership in organizations. *Acad. Management Rev.* 26 (2):298–310.

Puranam P, Håkonsson DD (2015) Valve's way. *J. Organ. Design* 4(2):2–4.

Puranam P, Alexy O, Reitzig M (2014) What's "new" about new forms of organizing? *Acad. Management Rev.* 39(2):162–180.

Rapp TL, Bachrach DG, Rapp AA, Mullins R (2014) The role of team goal monitoring in the curvilinear relationship between team efficacy and team performance. *J. Appl. Psych.* 99(5):976.

Reagans R, Zuckermann E, McEvily B (2004) How to make the team: Social networks vs. demography as criteria for designing effective teams. *Admin. Sci. Quart.* 49(1):101–133.

Ruef M, Aldrich HE, Carter NM (2003) The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *Amer. Sociol. Rev.* 68(2):195–222.

Sauermann H, Cohen WM (2010) What makes them tick? Employee motives and firm innovation. *Management Sci.* 56(12):2134–2153.

Seabright MA, Levinthal DA, Fichman M (1992) Role of individual attachments in the dissolution of interorganizational relationships. *Acad. Management J.* 35(1):122–160.

Shimizu K (2012) Risks of corporate entrepreneurship: Autonomy and agency issues. *Organ. Sci.* 23(1):194–206.

Shin SJ, Kim TY, Lee JY, Bian L (2012) Cognitive team diversity and individual team member creativity: A cross-level interaction. *Acad. Management J.* 55(1):197–212.

Simon M, Houghton SM, Gurney J (1999) Succeeding at internal corporate venturing: Roles needed to balance autonomy and control. *J. Appl. Management Stud.* 8(2):145.

Sitkin SB (1992) Learning through failure: The strategy of small losses. *Res. Organ. Behav.* 14:231–266.

Steffens P, Terjesen S, Davidsson P (2012) Birds of a feather get lost together: New venture team composition and performance. *Small Bus. Econom.* 39(3):727–743.

Thompson LL (2000) *Making the Team: A Guide for Managers* (Prentice Hall, Upper Saddle River, NJ).

Van de Ven AH (1986) Central problems in the management of innovation. *Management Sci.* 32(5):590–607.

Viktoria Boss is a PhD candidate at the Institute of Entrepreneurship, Hamburg University of Technology. Her research focuses on team formation and composition in entrepreneurship and sports. She received an MSc degree from RWTH Aachen University.

Linus Dahlander is a professor at ESMT Berlin and the holder of the Lufthansa Group Chair in Innovation. His research focuses on networks, communities, and innovation. He received a PhD from Chalmers University of Technology and was a postdoctoral fellow at Stanford University.

Christoph Ihl is a professor at Hamburg University of Technology and head of the Institute of Entrepreneurship. His research focuses on entrepreneurship, networks, and culture. He received a PhD from Technical University of Munich and was an assistant professor at RWTH Aachen University.

Rajshri Jayaraman is an associate professor at ESMT Berlin and the University of Toronto. Her research focuses on labor economics, networks, and how people respond to incentives. She received her PhD from Cornell University.