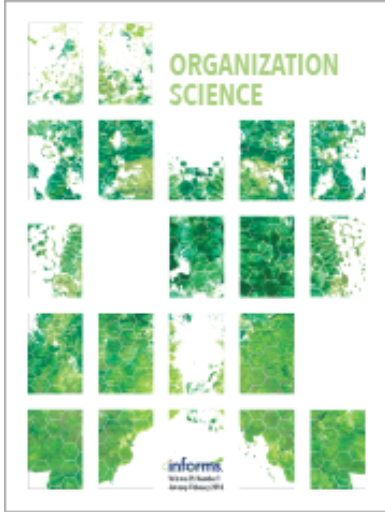




Organization Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Selection Regimes and Selection Errors

Dmitry Sharapov, Linus Dahlander

To cite this article:

Dmitry Sharapov, Linus Dahlander (2025) Selection Regimes and Selection Errors. *Organization Science* 36(6):2324–2348. <https://doi.org/10.1287/orsc.2023.17482>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Organization Science*. Copyright © 2025 The Author(s). <https://doi.org/10.1287/orsc.2023.17482>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2025 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Selection Regimes and Selection Errors

Dmitry Sharapov,^{a,*} Linus Dahlander^b

^aImperial College Business School, Imperial College London, London SW7 2AZ, United Kingdom; ^bESMT Berlin, 10178 Berlin, Germany

*Corresponding author

Contact: dmitry.sharapov@imperial.ac.uk,  <https://orcid.org/0000-0002-9330-776X> (DS); linus.dahlander@esmt.org,

 <https://orcid.org/0000-0003-3527-7440> (LD)

Received: March 13, 2023

Revised: January 2, 2024; September 5, 2024;

January 20, 2025; March 15, 2025

Accepted: March 26, 2025

Published Online in Articles in Advance:

May 14, 2025

<https://doi.org/10.1287/orsc.2023.17482>

Copyright: © 2025 The Author(s)

Abstract. How can the selection of innovation projects be designed to reduce false positives and false negatives? Prior research has provided theoretical insights into organizing to reduce errors, yet we know little about how organizations adapt selection over time and the effects of this on selection outcomes. Drawing from qualitative data from 126 interviews conducted over several years, we explore how an accelerator evolved through three selection regimes for high-stakes funding decisions, focusing on the organizational changes and their underlying reasons. We then analyze quantitative data from all 3,580 submissions they received, assessing false positives and false negatives across these regimes. Our findings reveal a persistent occurrence of both types of errors, with relatively small differences across the regimes despite deliberate efforts to enhance the process. In the final regime, which increased submission quality by emphasizing applicant track record and adding additional layers of screening, evaluators surprisingly became more prone to making selection errors. This finding stands net of accounting for (1) differences in the pool of submissions, (2) differences in treatment effects through training and resources provided, (3) learning, and (4) market evolution. By combining qualitative and quantitative data, we explain this through two mechanisms: (1) mean reversion in combination with increased emphasis on applicant track record and (2) within-type adverse selection enabled by a more stringent selection process. The study reveals that evolving an organization's selection regime may require adjustments across multiple aspects, resulting in unintended consequences.



Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Organization Science. Copyright © 2025 The Author(s). <https://doi.org/10.1287/orsc.2023.17482>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/orsc.2023.17482>.

Keywords: organization design • ideas • innovation • evaluation and selection of innovation projects • screening • selection error • false positives and false negatives • mixed methods • longitudinal research design • accelerator • app

Introduction

Choosing innovation projects comes with inherent difficulty. Organizations face not only substantial uncertainty about market and technological outcomes (Criscuolo et al. 2017, Joseph and Gaba 2020) but also the risk of making errors of commission—selecting unsuccessful projects (false positives)—and errors of omission—rejecting eventually successful ones (false negatives) (Csaszar 2012, Joseph and Gaba 2020). Decades of research on organizational structure and innovation have, accordingly, wrestled with one core question: How should firms organize selection processes to minimize these costly errors (Knudsen and Levinthal 2007)?

Building on foundational work by Simon (1955), one prominent research stream uses models to identify the effects of decision-making structures ranging from hierarchies to polyarchies on both types of errors (Sah and Stiglitz 1986, Christensen and Knudsen 2010). In

parallel, observational and laboratory studies examine how real organizations structure their selection processes (e.g., Reitzig and Sorenson 2013, Klingebiel and Adner 2015, Levinthal 2017, Criscuolo et al. 2021), revealing that few firms operate at the extremes of purely hierarchical or polyarchival decision making (Csaszar 2012, Eklund and Kapoor 2022). Scholars increasingly recognize that who selects and how they aggregate information can shape which errors predominate (e.g., false positives versus false negatives).

Yet, selection regime outcomes depend not only on how effectively a firm selects from a pool of alternatives but also on the composition of that pool. Recent studies have begun to disentangle how design choices affect both the generation and selection of submitted ideas (Keum and See 2017, Eklund 2022, Park et al. 2024). This challenge is heightened when external actors, such as entrepreneurs, can strategically choose where and when

to submit their ideas (Pahnke et al. 2015, Dahlander et al. 2021, Park et al. 2024). If a selection process becomes more stringent, some entrepreneurs with promising ideas may opt out, whereas others might be strategic in choosing which ideas to submit to this organization versus others.

Complicating matters further, most existing research treats adaptation to environmental shifts—such as adding or removing decision layers—as a single-dimension change (Tushman and Anderson 1986). However, organizations often alter multiple factors simultaneously, including both selection regime structure and the evaluation criteria used, which are not always reliable signals of true potential (Csaszar et al. 2023) and whose application may be cognitively taxing for evaluators dealing with different kinds of innovations (Polidoro 2020). Consequently, the effects of bundled adjustments may be difficult to predict and thus produce unintended consequences.

This paper offers a longitudinal lens on how selection regimes evolve inside organizations. We track one accelerator's shift through three distinct selection regimes—spanning 3,580 project applications with funding from €20,000 to €70,000—and integrate unique qualitative and quantitative data to capture their effects on the set of ideas submitted for evaluation, selection effectiveness, and the selection errors that emerge. We combine this with 126 interviews with selectors and entrepreneurs to gain a deep contextual understanding of the rationale behind both decisions to change selection regimes and evaluation criteria and the selection or rejection of projects for funding.

Our findings demonstrate that even though additional decision layers and stricter track record criteria improved average submission quality, they also increased false-positive decisions. This finding stands net of accounting for (1) differences in the pool of submissions, (2) differences in treatment effects through training and resources provided, (3) learning, and (4) market evolution. By combining qualitative and quantitative data, we explain this finding through two mechanisms: (1) mean reversion in combination with increased emphasis on applicant track record and (2) within-type adverse selection enabled by a more stringent selection process. Specifically, evaluator decisions became overly reliant on entrepreneur past achievements in a setting more random than they assumed, while some entrepreneurs with prior success submitted marginal projects, devoting less effort to them, and newcomers who may have had good ideas were passed over. By pinpointing these mechanisms, we illuminate why structural changes that seem beneficial can inadvertently undermine selection performance. Our study reveals that changing a selection regime is rarely a simple lever pull; instead, it triggers interconnected shifts in structure, criteria, and entrepreneur behavior that jointly affect who applies and which projects get selected.

Theoretical Background: Structure and Bundled Changes

A long line of research has sought to understand the search for and evaluation of alternatives in organizations, particularly under uncertainty. Simon's (1955) early work on bounded rationality identified two core challenges: knowing the set of alternatives (the search problem) and figuring out their consequences (the evaluation problem). These issues loom large in innovation selection, where organizations must assess potential products, services, or processes and decide which to fund despite uncertain outcomes. Building on Simon's insights, the behavioral theory of the firm (Cyert and March 1963) introduced the notion that organizational structures shape how alternatives are generated and ultimately chosen (Knudsen and Levinthal 2007). Subsequent models have emphasized decision makers' fallibility, stemming from bounded rationality and heuristics (March 1991, Levinthal 1997). Accordingly, organizational choices often fail to produce the desired benefits, in part because of time pressure and attention constraints (Cyert and March 1963, March and Olsen 1975).

Decision making under uncertainty is not just about cost-benefit calculations: biases can creep in even when organizations aim for meritocratic ideals (Boudreau et al. 2016, Criscuolo et al. 2017, Li 2017, Christensen et al. 2023, Csaszar et al. 2023, Piezunka and Schilke 2023). Selection decisions often deviate from a project's inherent merit (Mueller et al. 2012, 2014; Klingebiel and Adner 2015; Levinthal 2017; Sengul et al. 2018; Criscuolo et al. 2021; Bian et al. 2022). Innovator characteristics (Kanze et al. 2018) and situational factors such as timing (Bian et al. 2022) and sequence (Criscuolo et al. 2021) matter. Research has documented how the allocation of decisions across actors within an organization and the degree to which actors make decisions based on the parochial concerns of their local subunit, or the payoff to the broader organization, affect the adaptive capabilities of an organization (e.g., Dosi et al. 2003, Rivkin and Siggelkow 2003, Ethiraj and Levinthal 2004). For instance, Reitzig and Sorenson (2013) find that managers tend to favor projects from their own divisions. Individual-level biases thus aggregate into an organizational problem, highlighting how structure interacts with evaluation criteria to shape selection outcomes.

Sah and Stiglitz (1986) compared two stylized decision-making regimes: a polyarchy, in which an idea is selected if any one of the managers independently evaluating it supports it, and a hierarchy, where only those ideas sequentially approved by multiple managers are selected. This model shows that hierarchical dependence among decision makers increases the likelihood that a project will eventually not be supported, even though its objective quality recommends it, leading to a smaller share of false positives but a greater

share of false negatives. By contrast, a polyarchy's likelihood of funding a low-quality project is greater, leading to a smaller share of false negatives but a greater share of false positives. Several scholars have expanded these insights by using simulations to understand selection decisions within companies (Levinthal 1997, Rivkin and Siggelkow 2003, Ethiraj and Levinthal 2004, Knudsen and Levinthal 2007, Csaszar 2013, Csaszar and Eggers 2013, Böttcher and Klingebiel 2025). Knudsen and Levinthal (2007, p. 41) extend the work on hierarchies and polyarchies by also considering their hybrids. They argue that "organizations that are hierarchical in structure, even if composed of imperfect evaluators, tend to replicate the conservatism of perfect evaluators and become trapped by local peaks. Hybrid forms, comprising a mixture of polyarchy and hierarchy, effectively balance the dual imperatives of exploration and exploitation." This observation is essential in that few companies use the extreme forms of hierarchy or polyarchy, with most employing hybrid selection regimes.

Conceptual and modeling work has provided valuable insights into how to structure internally to avoid false positives and negatives (Sah and Stiglitz 1986, Knudsen and Levinthal 2007). But real organizations often do more than tweak one variable in isolation; bundled changes can inadvertently shift both the submission pool and the selection lens. This interplay remains under-studied, particularly over time, where organizations continually adjust their selection processes in pursuit of improvement yet may trigger new forms of error.

Large-scale empirical research on false positives and negatives remains rare because tracking accepted and rejected projects and their outcomes is difficult. When studying false positives, researchers typically analyze the performance of selected projects, which are often subject to survivorship bias. False negatives are even more challenging to observe, requiring knowledge of rejected ideas and their subsequent fate. Some progress has been made by examining discontinuities just below and above a threshold between accepted and "just rejected" ventures, revealing performance gains from selection (Kerr et al. 2014, Hallen et al. 2020), or by using matching methods to compare funded and unfunded firms (Pahnke et al. 2015).

Recent studies exemplify how structure influences selection errors. Csaszar (2012) examines stock-picking among mutual funds, finding that raising the consensus threshold reduces commission errors but increases omission errors, consistent with Sah and Stiglitz (1986). Keum and See (2017) complement this perspective by showing that hierarchy of authority can dampen idea generation but may also reduce selection errors. Eklund (2022) and Eklund and Kapoor (2022) similarly reveal how centralized versus decentralized R&D structures shape project creation, development, and the use of external innovation.

Overall, recent empirical work on selection processes and outcomes has demonstrated consequential differences between organizational practices and common modeling assumptions and has begun to investigate the influence of organizational structure on both the generation and selection of innovation. However, studies primarily focus on comparing static selection regimes or on a single dimension of change, and we know little about how evaluators reason about changes in how they organize and how entrepreneurs submitting ideas for selection respond. If organizations adapt their selection regimes over time and attract different ideas, then it is plausible that selection regime outcomes depend on both selection regime effectiveness and the set of alternatives available for selection. Accounting for differences in the available pool of ideas and treatment effects, we seek to understand how organizations organize to reduce selection errors. Rather than proposing formal hypotheses, we document what caused the selection regimes to change, using both the evaluators' and entrepreneurs' perspectives. We then turn to a quantitative analysis comparing selection errors in each regime, accounting for differences in the pool and treatment effects and evaluating alternative explanations. Lastly, we return to our qualitative and quantitative data to investigate the mechanisms at play.

Data and Empirical Approach

Research Context

AppCampus was a three-year accelerator created in May 2012 through a collaboration among Microsoft, Nokia, and Aalto University in Finland. Microsoft and Nokia each initially committed to providing €9 million in funding,¹ and Aalto University agreed to host the accelerator and cover its operating costs. The program was managed by a team of 8–12 people and was supported by several employees at Nokia and Microsoft. Control was exerted via a steering board of two representatives from Microsoft and Nokia, one from Aalto University, and one Finnish entrepreneur who served as an independent board member.

AppCampus initially had an open application process and offered grants of €20,000 to €70,000 and support to developers of applications for the Windows Phone platform. Submissions underwent a stringent selection process emphasizing application novelty and quality, yielding an overall acceptance rate of approximately 10%. Although the grants did not involve any exchange of intellectual property or equity, applications developed through AppCampus would, once released, be available exclusively on the Windows Phone for initially six and later three months.

By the end of the program in May 2015, AppCampus had received more than 4,300 submissions from more than 100 countries and had invested approximately €10 million in more than 400 app ideas.

Data

To capture changes in the accelerator’s selection regime, we use the detailed qualitative data collection performed by one of the coauthors following a grounded theory approach (Glaser and Strauss 1967). This data collection spanned more than two and a half years (December 2012 to June 2015) and comprised 10 waves (taking place roughly every quarter) of semistructured interviews and informal conversations with key informants at AppCampus, Nokia, Microsoft, and Aalto University; application developers funded by AppCampus; and attendees at key events. Overall, we conducted 126 semistructured interviews, which we complemented and triangulated using archival material from both online and physical sources, such as strategy and review documents, memos, and operational reports. The sources of evidence used to guide our insights are summarized in Table 1.

An interview protocol was developed, used, and adapted as interviews progressed. This enhanced data collection flexibility allowed us to probe emergent themes and take advantage of unexpected opportunities (Eisenhardt 1989). We initially identified the people involved in the creation of AppCampus (Head of Aalto Centre for Entrepreneurship (ACE) and AppCampus’s CEO, COO, Head of Screening, and Head of Quality Assurance) and interviewed them about how AppCampus came to be established and the initial organizational design and processes. On subsequent site visits, we extended our interviews to all AppCampus staff, as well as to the employees of Aalto University, Nokia, and Microsoft who became actively involved in AppCampus operations. The interviews began with us asking about the interviewee’s background and how they became involved with AppCampus, before proceeding to cover their roles in the organization, the work that they did, and any changes in the organization and its interactions with developers and other stakeholders since our last interview with them. Once AppCampus began to host periodical two- to four-week in-person camps for invited funded developers (called “AppCademy”), we strove to interview all of the developer teams taking part² about their background, the app ideas that they were working on, the reasons for them applying to

AppCampus, and their experience of the program. All interviews were transcribed verbatim, and the transcriptions were validated by both the coauthor who conducted the interviews and a research assistant, ensuring completeness and clarity of the data. One of the coauthors also had informal conversations with every employee of AppCampus and with a substantial number of the application developers regarding relevant events surrounding the Windows Phone platform and personal sentiments about the value, structure, and operation of AppCampus itself. Further, the coauthor attended several key strategic, planning, and operational meetings and a handful of open days and celebrations, including AppCampus’s first birthday event, a “demo day” organized to display applications of AppCampus participants, and four AppCademy events.

We analyzed the perspectives of AppCampus employees and the funded developers and how these evolved over time as selection regimes changed. We focused on selection (why the entrepreneur chose AppCampus, and how AppCampus chose the entrepreneurs) and treatment (what the entrepreneur hoped to get out of it, and what AppCampus hoped to provide to the entrepreneur). Both coauthors reviewed the interview transcripts and grouped quotes relevant to these broad themes into subthemes. Where a subtheme was represented by two or more quotes within a given selection regime, we compiled them into tables of illustrative quotes, both from the perspective of AppCampus (Online Appendix A) and from the perspective of the developers (Online Appendix B).

We employed a full-cycle research approach (Mortensen and Cialdini 2010) by integrating qualitative and quantitative data to comprehensively understand the mechanisms at play. Our qualitative insights guided the formulation of our quantitative analysis, whereas the quantitative results prompted us to revisit the qualitative data, thereby deepening our understanding. This iterative process allowed us to refine and better explain our quantitative findings (see, e.g., Di Stefano and Micheli 2023).

We gathered archival material from online and physical sources to triangulate primary data, including strategy and review documents, memos, and operational

Table 1. Overview of Data Sources

Type	Source	Number	Pages (single spaced)	Time
Primary sources	Interviews	126	1,146	Dec. 2012–May 2015
	AppCampus	55	549	Dec. 2012–May 2015
	Developers	50	422	Sept. 2013–Jan. 2015
	Aalto University	14	121	Dec. 2012–May 2015
	Microsoft	4	33	Sept. 2013–Nov. 2014
	Nokia	3	21	June 2013
Secondary sources	Internal company documents	11	142	Apr. 2012–Jul. 2015
	Press releases	104	172	Mar. 2012–Aug. 2016

reports. These were examined both during and after the fieldwork period. The archival data were used primarily to complement and verify our primary data. Archival data from industry sources were also used to cross-validate the emerging insights.

To quantify how AppCampus's changing selection regimes affected their decisions, we use data on all application idea submissions received by AppCampus. After removing duplicates and ineligible submissions, we had data on 3,580 investment decisions. Alongside the full text of the submission document describing the app idea and development team, we have a record of the decision made by AppCampus on every submission. For submissions selected for funding, we have data on their performance directly from AppCampus, which we supplement with hand-collected data following the approach described below to ensure that we account for their performance across different app markets in the years after AppCampus ceased to operate. For rejected submissions, we hand-collected performance data from mobile application stores with the help of four research assistants. The coauthors first took a small sample of submissions and attempted to find information about the release and performance of their proposed applications. Based on this experience, we developed a "search protocol" for the research assistants on how to use multiple different sources, including developer team websites; LinkedIn, Twitter, Facebook, and Instagram pages; mobile application stores; and the Wayback Machine internet archive to find whether that particular developer team eventually released apps corresponding to the app description submitted by the developer to AppCampus³ and, if so, how many times the app was downloaded.⁴ The research assistants first received a file of 100 randomly selected submissions to code. When they finished, they submitted their work to the coauthors, one of whom checked a sample of each file for errors in coding. Two research assistants were found to have a high error rate and were replaced with two others. To further guard against coding mistakes, 242 submissions were coded by pairs of different research assistants to triangulate their work and identify disagreements. A disagreement rate below 6% suggested that the research assistants consistently followed the guidance the coauthors gave; one of the coauthors resolved any disagreements in coding.

Qualitative Findings: The Three Selection Regimes

We use insights from our qualitative data collection to examine how AppCampus changed its selection processes over time. To provide a contextual understanding, we begin by describing AppCampus's prehistory and first year of operation. After describing its origins, we use our data to explain AppCampus's and the

entrepreneurs' perspectives during the three selection regimes.

How AppCampus Came To Be

For Nokia and Microsoft, AppCampus was, first and foremost, an investment in the creation of high-quality products and services in the form of applications that would complement the Windows Phone platform. The founding of AppCampus also reflected Microsoft's and Nokia's understanding that although their brands, in-house processes, and capabilities were well-suited to attracting experienced developers, attracting younger, less experienced application developers to the ecosystem would be challenging because of the legitimacy deficit these incumbents faced as a result of not having already established successful smartphone ecosystems, unlike their rivals Apple and Google. AppCampus was set up at Aalto University as it was considered to be well-positioned to address a target population of developers who had been largely overlooked by competing ecosystems and whose cost-benefit calculations could be most significantly affected by the incentives offered by AppCampus:

I think Finland's reach into emerging markets was critical to Aalto's selection as the university partner to AppCampus. While the grants we are giving may be relatively small to a developer in Silicon Valley, even a €20,000 grant can be very significant to a developer from, say, Russia, Romania, or Poland, covering salary equivalent of over 1 man-year. And these are proving to be among the strongest regions both in terms of deal flow and performance once in the program. —Head of Aalto Centre for Entrepreneurship (ACE)

With many submissions received in the first month of operation, processes were created for evaluating them and selecting the most promising for funding. A quality assurance process was developed to ensure that teams selected for funding met all formal requirements for program participation and would be guided through the development of a design document for their application. Once the design document met the AppCampus standard, 30% of the grant funding was released to the team, with the remaining 70% released once the application was approved for launch in the Windows Phone store.

Alongside building developer capabilities during the award and quality assurance processes, AppCampus also periodically hosted two- to four-week-long training camps for developer teams, paying the costs of developer training, travel, and accommodation. These consisted of sessions on application design and development, and also entrepreneurial skills such as pitching application ideas to investors, marketing, and communications, and concluded with pitches to potential investors at demo days or participation in the pitching competition at the Slush tech, design, and start-up conference.

By the end of 2012, AppCampus had received 1,647 submissions, 80 of which were approved for funding. The first funded apps were released to the public via the Windows Phone marketplace at the end of December 2012. The first released application, a game called “Haunted,” performed well, becoming one of the top 10 most downloaded applications in several countries by March 2013, and reaching hundreds of thousands of downloads by June 2013. It also produced substantial revenue for its developers and received good user ratings. Although other funded apps also succeeded, many did not.

We next discuss the three selection regimes used by AppCampus over time. Table 2 summarizes the key differences among these selection regimes (Online Appendix C gives a more detailed overview of how each selection regime was structured).

Selection Regime 1: The Investment Board Between May 2012 and April 4, 2013

In the first stage, the organization used an initial screening performed by AppCampus employees to reject submissions that did not meet the scope criteria (e.g., already released on iOS or Android; app not compatible with Windows Phone) and provide some comments on the eligible cases. These were then sent to an investment board of four people—the AppCampus CEO, AppCampus COO, Head of ACE, and an Entrepreneur in Residence at ACE—who discussed each case together and made the decision by voting, with the CEO casting the deciding vote in the case of a tie. The investment board met once a week and made decisions on 10 to 25 eligible submissions in a given meeting. Novelty of the app idea was the key evaluation criterion at this stage, as stated by the Head of Screening: “The first criterion we have is this innovative side, so just to be something new, if it’s done many many times, we won’t select it, so basically,

when I see something that is coming out of nowhere, I’m very happy.” The decision makers were concerned about finding hit apps but also realized the difficulty of doing this, and the trade-offs between false positives and false negatives involved, as the Head of ACE stated: “I’m ... worried that maybe, through this process we’ll miss a potential winner. Because that’s the big thing, it’s shown over and over again how easy it is to scoff at something that actually has huge potential. [...] You know, I think that is how we have to set the screen at a level that we’re able to get a quality, crazy idea in that might just be huge. And you’re willing to take a certain number of things that won’t fly in order to get that one.”

The investment board comprised different people from different areas of expertise. The reason was that by aggregating information from different people, the board would be more likely to capture novel projects that do not fit the mold and stand out. In practice, decisions taken by the investment board members were not independent. Even in situations where one decision maker had a strong view, this often changed as a result of the discussion, an aspect of the process that was missed if investment board members had to conduct evaluations by email because of travel commitments. As ACE’s Entrepreneur in Residence said, “I ... voted through email. And, well, it’s unfortunate because then you are looking at it yourself, you lack the discussion part ... there are a couple of cases I ... voted totally the opposite. But when we had the discussion, I understood something that I had totally missed.” Group decision-making processes of this kind allow more deliberate decision making, where stated evaluations may change the views of other members before the voting is finalized. As such, the hope was that the group could correct errors made by individual decision makers.

From the perspective of developers, those who applied to AppCampus during this selection regime did

Table 2. Overview of Differences Between the Selection Regimes

Dimension	Selection regime 1	Selection regime 2	Selection regime 3
Decision-making structure	Investment board votes on selections with group deliberation, with the CEO casting the deciding vote in case of a tie	Decisions delegated to AppCampus screening employees making an individual recommendation, reviewed by the CEO	Sourcing and prescreening using global camps with Microsoft partners, final decisions by AppCampus staff, reviewed by the CEO
Selection: screening and evaluation criteria	Emphasis on novelty of the app idea. Open to all applicants with a focus on innovative ideas	Emphasis on mass market potential, later focusing on applicant track record (timing used in the analysis)	Strong focus on developer track record, prioritizing teams with proven success. Invitation-only through partner organizations and prescreening events
Treatment: access to resources and support	General support available for all applicants, with a focus on enhancing app quality and developer skills	Increased emphasis on marketing and technical support postselection, targeted toward market success	Postselection support, emphasizing marketing, compliance, and delivery on proposed app features

Note. Additional information about the differences in the selection regimes can be found in the Online Appendix.

so primarily because of the funding enabling them to work on developing their idea full time: “I used to work in a software company and I did these apps as a side project. And then AppCampus was like an opportunity for me to switch from part time to work full time on this” (Developer 8). The developers selected for funding in this regime were often relatively inexperienced, as illustrated by Developer 7: “I didn’t have any of the skills I needed to be able to build like the back-end for it and the database and the website and stuff. So when AppCampus came along, later on, I got those skills, and decided to apply.” They were also often focused on developing for Windows Phone first, with releases on other platforms to follow if the Windows Phone launch was successful: “The plan is to release on Windows phone and then if it’s good, if it’s successful, if the feedback is good and we can fix everything in time then we will most probably release the game as early as possible on other platforms” (Developer 1).

For funded teams, the focus of AppCampus staff during this period was primarily on helping them to improve the quality of their applications: “One of the coolest things is that after you have done applications with AppCampus, you know the quality. I mean, you are very good at quality at that point, and you normally don’t reach that level of quality if you are an independent developer because you don’t need to” (Technical team member). This focus was recognized in our interviews with developers, with Developer 3 stating: “[AppCampus team member] really had some good things to say about the game and he came up with a lot of suggestions for improvements.” The developers who came to the in-person AppCademy training camps also found the networking and pitching training valuable (see Online Appendix B).

Shortly after the program began, issues around scaling began to surface. Because of the number of submissions received exceeding expectations, both investment board members and the screening team who presented cases to them were pressed for time and had to manage competing priorities. As a result, it was difficult to focus attention on all projects on which they had to make decisions. The head of the screening team stated: “We need to keep a fresh mind and this is really difficult, especially when people come and disturb you all the time.” As selection regime 1 (SR1) began to take its toll on the attention of the screening team and the investment board members, they decided to change the process.

Selection Regime 2: Delegation Between April 5, 2013, and November 2013

As the number of submissions for funding grew, the investment board decision making could not keep up. It was becoming increasingly difficult for the investment board to meet. As the COO explained: “It was becoming a bottleneck, so [the regime change] was a work-around

to ... avoid having these bottlenecks strangling the whole ... selection process.” To streamline and speed up the process, evaluation decisions were delegated to AppCampus screening employees. The screening employee responsible for evaluating a given submission had decision authority, and their results were presented to the CEO for review and acceptance. Although most submissions were evaluated one after another in the order in which they had been submitted, in cases where the evaluator thought that additional input from other screening team members could be useful, the app submission and the preliminary evaluation could be discussed at weekly “peer review” meetings. As the Entrepreneur in Residence at ACE said, “The screening team now makes a proposal list, which basically [the CEO] reviews, and stamps that, ‘Okay, this is good.’” The evaluation criteria used also evolved to include an explicit consideration of the app idea’s mass market potential: “[Our scope has changed] to Windows Phone innovation but then combined with potential for mass download or mass appeal” (Head of Screening).

In this selection regime, in addition to the funding offered, potentially privileged access to Windows Phone technology was an important reason for applying for some of the developers, for example, Developer 15: “I think technically it’s quite impressive where Windows Phone is in so little time, and another reason for us to be super stoked about working with them and getting on the platform early.” The cohort selected for funding in this selection regime still included many inexperienced developers, with Developer 20’s story being quite typical: “We were just a bunch of students and we worked on the game for five months before we had a playable prototype.” However, almost all of the teams were actively pursuing or considering a release for their apps across multiple platforms.

Funded teams in this selection regime continued to receive AppCampus’s assistance with the app’s design, although, from the perspective of AppCampus technical staff, the general level of technological understanding had increased because of the increasing availability of online materials, leading them to be contacted less but for more difficult problems: “Most [teams] understand Windows Phone well and currently there is enough research and information on the Internet and everywhere so that they don’t really need help on those basic features. Now when they have a problem they are seriously hard problems, they are not something that you can just Google and find the answer for. So I would say that the technical knowledge has risen a lot and the level of the problems also because of that has risen a lot” (Technical team member). A further aspect of postfunding support that became more of a focus in this regime was assistance with marketing the app postrelease, which was appreciated by several interviewed developers, alongside AppCampus’s input when it came to the

business aspects of being an app developer. Those taking part in AppCademy continued to point to the networking aspects as being highly valuable.

At the end of June 2013, the evaluation criteria used by the screening team changed to put additional emphasis on applicant track record. Two months later, this was followed by a substantial change in the way that AppCampus presented itself to potential applicants. In a blog post on the AppCampus website on August 28, 2013, it was announced that “it is time to evolve the selection criteria for funding great mobile applications for Windows phone,” with the main change being that “we are putting more focus on the track record of your team or company. If you have already proven that you can create successful mobile apps (on any platform) you are more likely to get our attention. We’ll be very interested in seeing evidence of your achievements (i.e. download numbers, revenue, rankings and ratings, awards). If you are a new company or have little experience in mobile development the bar will be much higher than before. You would really have to blow us away with an exceptionally innovative idea in order to qualify for an award.” The rationale behind this change was stated as follows by the CEO:

What we found out was that the lead time for getting a quality app out in the store for a start-up that is first time round or nearly first time round, is actually surprisingly long [...] We are not a volume program. The latest number that has been announced on number of apps in Windows Phone Store is 170,000, I think. So if we push out 300, 400 teams through the three-year program, it doesn’t really nudge the meter, so that’s not why we are here. We are here to create an environment where the developers who have potential can get successful sooner rather than later; that’s the idea. [...] we also need to get stuff out of the system [instead of] just, you know, have a lot of people that get helped along the way, but we also need to have real, hard results.

The emphasis on submissions being “Windows Phone first” was also relaxed for apps that were already successful on other platforms, as long as they would incorporate novel content or features exclusive to Windows Phone. Although not affecting the structure of the selection process at the time, this change demonstrated a shifting of AppCampus priorities from its initial focus on app idea innovativeness as a key evaluation criterion toward a much greater focus on developer track record as a key criterion, in the hope that this would help with delivering high-performing mobile apps.

Selection Regime 3: Preselection Between December 2013 and January 2015

Despite the changes in evaluation criteria, toward the end of SR2, there were growing concerns that the

selection process also needed to be reorganized to attract better projects. By having an open submission portal on the AppCampus website flooding the selection process with many app ideas, many lower-quality projects usurped attention. As a result, the selection process was reorganized to match the stated focus on top developers. As the screening team head explained: “It’s not anymore about the ... young team that starts and we help the young team getting into the platform. It’s more about the team which is successful and we help them, or we push them to ... the next success [...] So basically you need to have a successful app to be able to [make] money.” The reorganizations meant that open submissions through the AppCampus website were closed (apart from limited periods coinciding with industry conferences). Instead, submissions were primarily acquired through globally distributed two-day camps each involving approximately 10 developer teams invited by local Microsoft partners, or through direct referrals from Microsoft and other partner organizations. As Microsoft’s liaison to AppCampus explained:

We thought, why don’t we have the local teams ... help drive similar activities but in a little bit lighter fashion and do a two-day recruiting event; and they will provide coaching and support. And try to essentially get the teams in ... better shape, so that they can submit their ideas to AppCampus and increase the success rate and help us pre-screen those teams.

The top one to three of these teams were put forward for AppCampus funding but still needed to complete an application idea submission, which was evaluated by the screening team in the same manner used in selection regime 2. In using distributed expert opinions (of local Microsoft teams) in this manner, in combination with the emphasis on developer track record as a key selection criterion, the aim was to have a prescreened set of projects for consideration with a tighter quality distribution and a higher mean than the general population. This was echoed in interviews with a screening and tech team member:

Our main channel was this MAAC concept, Mobile Application Acceleration Camp, and that worked extremely well. We had like one per week as an average in the whole spring and we get dozens of new great teams from that channel. So for us it wasn’t—didn’t make that much sense to keep the online open. We would rather try to guide quality teams to the local MAAC events that actually worked pretty okay.

The developers applying to AppCampus in this selection regime tended to focus primarily on the money on offer in explaining their reasons for applying, often planning to use the Windows Phone release as a “soft launch” that enabled them to identify issues to be fixed before launching on iOS or Android: “Because of the

funding thing, it's sort of a chance to get two extra months of development time and to polish the game more, and to see if all our changes work and then get ready for our iOS launch, really" (Developer 24). Compared with the previous selection regime, the resulting cohort of funded developers tended to be more experienced, as illustrated by this quote from Developer 29—"I've been through a bunch of programmes similar in nature, not how they're programmed but in the nature of the things, the subjects that you listen to. I've been through four incubators with these and other projects"—and to produce cross-platform apps, with Windows Phone not being their first priority: "[AppCampus] understand the realities that people can't focus on a platform that is the third most popular, so we have to also think about our business as a whole and also the other platforms" (Developer 45).

In terms of postfunding support, many of the interviewed developers who were selected in this selection regime highlighted the funding as the main benefit. Those who took part in an AppCademy continued to appreciate the networking, pitching, and marketing support. From AppCampus's perspective, in addition to the continuing marketing support provided to some teams postfunding, members of the Quality Assurance and technical teams became increasingly involved with monitoring the compliance of funded teams in terms of delivering on the features proposed in their applications: "So it might be basically in any stage [of the Quality Assurance process] - me noticing that [developers] try to cut radically their scope, or they cut some key features they have promised to do, for which we have selected them to the programme. So we might say: Hey, we're not going to pay for that. So either you do what you promised, or you're out" (Head of Quality Assurance).

Selection Errors by Different Selection Regimes

The qualitative data explain how and why the selection regimes evolved and the entrepreneurs' perspective of what they hoped to gain from AppCampus. However, they do not give us a comprehensive picture of the selection errors and selection outcomes that these selection regimes produced. To evaluate selection errors in AppCampus's three selection regimes, we use data on whether submitted app ideas were released in an app store and on the performance of those released, measured in terms of their number of downloads.⁵ We define false positives as app ideas selected for funding that either were not released or were released but generated fewer than 10,000 downloads. False negatives are app ideas rejected by AppCampus that went on to be released and generated either more than 500,000 or more than 1 million downloads, corresponding to AppCampus's criteria for "successful" and "hit" apps, respectively.

Table 3 shows the descriptive statistics of the number of submissions, projects funded, and selection outcomes in the three different selection regimes. The table provides evidence of the prevalence of both false positives and false negatives across all three selection regimes. There are the fewest false positives in SR2 when considering the percentages of funded unreleased and funded decisions with fewer than 10,000 downloads. We also see that the share of true positives is greatest in SR3, although the numbers are small overall. SR3 also has the lowest share of true negatives. This is to be expected given that this selection was designed to increase the average quality of the pool of submitted ideas. The data on false negatives capture whether rejected projects turned into a success when the team continued to work on them without AppCampus funding. We find that

Table 3. Descriptive Statistics—Investment Decisions and Outcomes Across Regimes

Measure	SR1: investment board	SR2: delegation	SR3: preselection	Overall
Submissions	1,774	1,137	669	3,580
Funded	143	100	184	427
False positives				
Funded unreleased	30.8%	18.0%	19.6%	22.7%
Funded <10K DL	34.9%	31.7%	39.6%	36.8%
Funded failed ^a	65.7%	49.7%	59.2%	59.5%
True positives				
Funded 500K+ DL	5.6%	7.0%	8.2%	7.0%
Funded 1M+ DL	1.4%	2.0%	6.5%	3.8%
True negatives				
Unfunded unreleased	91.6%	86.0%	76.9%	87.5%
Unfunded <10K DL	6.4%	11.5%	15.9%	9.6%
Unfunded failed ^a	98%	97.5%	92.8%	97.1%
False negatives				
Unfunded 500K+ DL	0.4%	0.8%	2.5%	0.8%
Unfunded 1M+ DL	0.2%	0.4%	1.9%	0.5%

Note. DL, downloads.

^aFailed refers to the sum of unreleased apps and those released but generating fewer than 10,000 downloads.

0.8% of unfunded projects got more than 500,000 downloads. There are differences across selection regimes, with SR3 resulting in the most false negatives (2.5%).⁶ However, the summary statistics above ignore potential differences in the pools of submissions received in the three regimes and other factors that could influence whether an app idea fails or succeeds. Thus, we next turn to using a regression approach to estimate the effect of the different selection regimes on the likelihood of false positives and negatives.

Dependent Variables

As discussed above, we consider three dependent variables building on the same categories used by AppCampus in assessing the performance of funded apps. *Failure* is equal to one if the app idea was not released or was released but generated fewer than 10,000 downloads, and zero otherwise.⁷ *Success* is equal to one if the app was released and generated more than 500,000 downloads and zero otherwise. *Hit* equals one if the app was released and generated over one million downloads, and zero otherwise.

Independent Variables

Our key explanatory variables are dummies for each of the three *selection regimes*, a dummy for whether the app idea was selected and *funded* by AppCampus, and the interactions between *funded* and the *selection regime* dummies. The differences in the pool of app ideas received across the three selection regimes are reflected in the estimated coefficients on the dummies for selection regimes 2 and 3 when their interactions with *funded* are not included in the models (the dummy for the first selection regime is omitted as the base category).

With the interactions between the *selection regime* dummies and *funded* included in the models, the estimated coefficients of the *selection regime* dummies become estimates of how likely nonfunded app ideas were to fail (true negatives) or become successes or hits (false negatives) in SR2 and SR3, relative to SR1. The estimated coefficient on *funded* can be interpreted as the association between an app idea being selected for funding using the first selection regime and the likelihood of this idea becoming a failure, success, or hit, whereas the interaction terms between *funded* and SR2 and SR3 reflect the differences in the likelihood of these regimes producing false positives or true positives, relative to SR1.

Control Variables

We control for *team size* by counting the number of team members mentioned in the submission form, and for the % *female* team members by using the genderizeR R package on the first names of team members. We control for gender as previous research has suggested that female entrepreneurs may face biases against them when seeking funding (Brooks et al. 2014, Botelho and

Abraham 2017). There are also potential differences in developer experience that accumulate over time and may be correlated with both the likelihood of being selected for funding and producing a hit app. We thus control for whether an app is a *sequel* to an existing one. We also add a control for whether the app has been *ported* from a different platform as this may affect uncertainty about the app's likely performance and could vary among the regimes. Credible signals received about an app idea could also affect the probability of making correct decisions. We thus control for whether the app has a *referral* code from a different accelerator or laboratory. We include *app category* fixed effects to account for the possibility that some of the 17 categories of apps that developers classified their submissions into in the AppCampus application form may be more likely to be hits and failures than others.

We also use *evaluator* fixed effects to account for some evaluators potentially being inherently stricter than others and *evaluator experience*, measured as the number of submissions that the focal submission's evaluator has previously reviewed, to control for changes in evaluator judgment that may occur with accumulated experience. At the organizational level, experience with the three different selection regimes may also affect the likelihood of selection errors, leading us to control for *regime experience*, measured as the number of decisions taken prior to the focal submission within the same selection regime. AppCampus's portfolio of already-selected app ideas may also have affected decision making and app prospects, so we control for the number of app ideas that AppCampus had already *funded in the category* of the focal app idea at the time of its submission.

To account for differences in funding among selected ideas, we use a *highly funded* dummy equal to one if the focal app idea was selected for funding at the €50,000 or €70,000 level (as opposed to the €20,000 granted to over 75% of funded applications), and zero otherwise. Similarly, to account for differences in feedback for rejected ideas, we use a *limited feedback* dummy equal to one if the focal app idea was rejected and the submitter was informed of this using a generic email containing no feedback specific to the app idea.

Finally, in line with other work that has studied idea evaluation (Piezunka and Dahlander 2019), we also include further controls for textual characteristics of the app idea description submitted to AppCampus in one set of models, as this description was the main means by which app ideas were communicated to evaluators. Specifically, we first included all the dimensions of the text as measured by the Linguistic Inquiry and Word Count (LIWC) text analysis application (Pennebaker et al. 2015). We then proceeded to remove dimensions that appeared to have no association with any of our three outcome variables in a step-by-step process, starting with removing the variable with the highest *p*-value

and proceeding until we were left with only those LIWC dimensions with a p -value of 0.10 or lower in at least one of the models. Our LIWC control variables thus capture the *word count* of the app description, and the *clout*, *positive emotion*, *anger*, *sadness*, *swear words*, and *nonfluencies* LIWC dimensions.⁸

Estimation Technique

For estimation, as our three dependent variables are binary, we use logit models with robust standard errors, implemented using the *logit* command in Stata.⁹

Quantitative Results

Table 4 presents the summary statistics of the variables used in the quantitative analysis. Their correlations are presented in Online Appendix G. The estimation results from these models are presented in Tables 5, 6, and 7. Each of these tables has two sets of three columns: the first set of three columns presents results for models excluding interactions between *selection regime* dummies and *funded*, whereas the second set of three columns presents results with these interactions included. Within each set of three columns, the first one shows results from models without any control variables or fixed effects for evaluators and app categories, and the second column adds these fixed effects and all non-LIWC control variables, whereas the third column adds the LIWC control variables described above.

In Tables 5 to 7, it is important to note the main effects for SR2 and SR3, presented in the first three columns of each table. The results show that SR2 and SR3 resulted in submissions with a lower likelihood of failure

(average marginal effects of SR2 and SR3 estimated from a specification including all control variables suggest 10.1% ($p = 0.024$) and 16.5% ($p = 0.019$) lower likelihoods of failure than SR1, respectively) and higher likelihood of becoming a hit compared with the first selection regime (5.8% ($p = 0.110$) and 10.8% ($p = 0.059$) higher than SR1, respectively). There is also some evidence for submissions received in SR3 being more likely to become successes (although the p -value for the estimated marginal effect of 7.3% is above the conventional cutoff at 0.214). This suggests that the pool of submissions received in the first regime substantially differed from those received in the last two. This is to be expected, given the intentional changes made between the first and third selection regimes to increase the quality of the submission pool through an increasing focus on applicant track record during SR2 and the move to prescreening applications in SR3. The *funded* variable also works as expected. On average, funded projects are less likely to fail and more likely to result in successes, with limited influence on the likelihood of hits. Specifically, being funded is associated with a 12.4% reduction in the probability of failure (average marginal effect estimated from a specification including all control variables, $p < 0.001$) and a 2.2% increase in the probability of success ($p = 0.005$).

To investigate how the selection processes in SR2 and SR3 performed compared with SR1 in terms of the likelihood of selection for funding leading to false positives and false negatives, the variables of interest are the estimates on SR2 and SR3, and their interactions with *funded*. The estimated coefficients on the *funded* dummy

Table 4. Regression Analysis Summary Statistics

Variable	N	Mean	St. Dev.	Minimum	Maximum
Failure	3,580	0.926	0.262	0	1
Success	3,580	0.016	0.124	0	1
Hit	3,580	0.009	0.094	0	1
SR2: delegation	3,580	0.318	0.466	0	1
SR3: preselection	3,580	0.187	0.390	0	1
Funded	3,580	0.119	0.324	0	1
Team size	3,580	1.994	1.316	1	10
% Female	3,580	0.098	0.235	0	1
Port	3,580	0.016	0.127	0	1
Sequel	3,580	0.006	0.078	0	1
Referral	3,580	0.331	0.471	0	1
Evaluator experience	3,580	480.127	387.286	0	1,408
Regime experience	3,580	682.099	467.636	0	1,773
Funded in category	3,580	31.103	53.687	0	225
Highly funded	3,580	0.028	0.166	0	1
Limited feedback	3,580	0.451	0.498	0	1
Word count	3,562	107.042	114.968	2	1,229
Clout	3,562	78.000	15.017	2.31	99
Positive emotion	3,562	5.725	4.069	0	31.03
Anger	3,562	0.440	1.419	0	17.78
Sadness	3,562	0.164	0.578	0	7.69
Swear words	3,562	0.021	0.244	0	6.25
Nonfluencies	3,562	0.124	0.495	0	14.29

Table 5. Results—Failures

Variable	DV: failure			DV: failure		
SR2: delegation	−0.408*	−1.785*	−2.101*	−0.251	−1.537+	−1.788*
	(0.173)	(0.833)	(0.829)	(0.267)	(0.861)	(0.862)
SR3: preselection	−0.716***	−2.863*	−3.421*	−1.357***	−3.114*	−3.611**
	(0.187)	(1.332)	(1.329)	(0.250)	(1.317)	(1.322)
Funded	−2.976***	−2.480***	−2.575***	−3.260***	−2.691***	−2.774***
	(0.157)	(0.226)	(0.234)	(0.251)	(0.319)	(0.324)
SR2 × Funded				−0.361	−0.430	−0.489
				(0.377)	(0.438)	(0.447)
SR3 × Funded				1.080**	1.065*	1.144*
				(0.341)	(0.461)	(0.475)
Control variables						
Team size		−0.127*	−0.132*		−0.122*	−0.128*
		(0.056)	(0.057)		(0.056)	(0.057)
% Female		0.148	0.235		0.187	0.299
		(0.318)	(0.328)		(0.326)	(0.336)
Port		−1.586***	−1.605***		−1.652***	−1.694***
		(0.359)	(0.375)		(0.369)	(0.380)
Sequel		−1.394**	−1.385**		−1.370**	−1.381**
		(0.498)	(0.493)		(0.493)	(0.481)
Referral		−0.362*	−0.353+		−0.353+	−0.351+
		(0.181)	(0.184)		(0.181)	(0.182)
Evaluator experience		0.002+	0.002*		0.002*	0.002*
		(0.001)	(0.001)		(0.001)	(0.001)
Regime experience		−0.001*	−0.001**		−0.001*	−0.001*
		(0.001)	(0.001)		(0.001)	(0.001)
Funded in category		0.002	0.002		0.002	0.002
		(0.002)	(0.003)		(0.002)	(0.002)
Highly funded		0.387	0.531*		0.420	0.574*
		(0.262)	(0.267)		(0.266)	(0.273)
Limited feedback		0.487+	0.484+		0.376	0.370
		(0.254)	(0.256)		(0.265)	(0.267)
Word count			−0.000			−0.000
			(0.001)			(0.001)
Clout			0.005			0.006
			(0.006)			(0.006)
Positive emotion			−0.061**			−0.060**
			(0.018)			(0.017)
Anger			−0.093*			−0.096**
			(0.037)			(0.037)
Sadness			−0.215*			−0.238*
			(0.105)			(0.105)
Swear words			0.454			0.525+
			(0.300)			(0.307)
Nonfluencies			0.232			0.258
			(0.176)			(0.183)
App category FEs	No	Yes	Yes	No	Yes	Yes
Evaluator FEs	No	Yes	Yes	No	Yes	Yes
Constant	3.774***	19.399***	20.063***	3.911***	20.020***	21.729***
	(0.137)	(1.415)	(1.484)	(0.179)	(1.587)	(1.638)
N	3,580	3,552	3,534	3,580	3,552	3,534

Notes. Logit regression; robust standard errors in parentheses. The baseline for comparison is SR1. FEs, fixed effects; DV, dependent variable.
+*p* < 0.1; **p* < 0.05; ***p* < 0.01; ****p* < 0.001.

and its interactions with SR2 and SR3 in the last three columns of Table 5 suggest that SR1 and SR2 were best at avoiding false positives. Specifically, whereas selection for funding in the first two selection regimes is associated with a 13.3% reduction in the probability of failure (*p* < 0.001), being selected for funding in SR3 was associated with a lower, 7.8%, decrease in the likelihood of failure (sum of the average marginal effects of *funded* and its interaction with SR3). Table 6 suggests that selection regime 3 was also least effective at identifying true positives. Whereas ideas selected for funding in SR1 and SR2 were 3.8% more likely to succeed than non-funded ideas (*p* = 0.002), those selected for funding in SR3 were only 0.4% more likely to succeed.

Table 6. Results—Successes

Variable	DV: success			DV: success		
SR2: delegation	0.497 (0.374)	2.218 (1.529)	2.205 (1.678)	0.745 (0.542)	2.835 (1.809)	3.102 (1.999)
SR3: preselection	1.123** (0.377)	4.204+ (2.502)	4.234 (2.734)	1.927*** (0.503)	5.257+ (2.794)	5.654+ (3.105)
Funded	1.931*** (0.317)	1.184* (0.499)	1.273* (0.528)	2.776*** (0.548)	2.005** (0.694)	2.224** (0.734)
SR2 × Funded				−0.505 (0.761)	−0.732 (0.984)	−0.983 (1.035)
SR3 × Funded				−1.523* (0.677)	−1.789+ (0.912)	−2.019* (0.986)
Control variables						
Team size		−0.038 (0.113)	−0.026 (0.110)		−0.046 (0.115)	−0.036 (0.112)
% Female		−0.795 (0.728)	−0.631 (0.679)		−0.801 (0.722)	−0.618 (0.685)
Port		1.977*** (0.450)	1.860*** (0.461)		2.172*** (0.483)	2.099*** (0.502)
Sequel		1.406* (0.711)	1.099 (0.803)		1.433* (0.711)	1.099 (0.810)
Referral		0.113 (0.365)	0.126 (0.383)		0.193 (0.355)	0.245 (0.380)
Evaluator experience		−0.002 (0.002)	−0.002 (0.002)		−0.003 (0.002)	−0.003 (0.002)
Regime experience		0.002 (0.001)	0.002 (0.001)		0.002 (0.001)	0.002 (0.001)
Funded in category		−0.007 (0.004)	−0.007 (0.004)		−0.006 (0.004)	−0.006 (0.004)
Highly funded		−0.830 (0.698)	−0.798 (0.708)		−0.882 (0.712)	−0.888 (0.740)
Limited feedback		−0.971+ (0.521)	−1.044+ (0.538)		−0.860 (0.526)	−0.958+ (0.545)
Word count			−0.004 (0.002)			−0.004 (0.002)
Clout			−0.022* (0.011)			−0.024* (0.011)
Positive emotion			0.078* (0.032)			0.076* (0.032)
Anger			−0.051 (0.079)			−0.052 (0.081)
Sadness			0.345* (0.164)			0.363* (0.166)
Swear words			0.226 (0.261)			0.164 (0.270)
Nonfluencies			−0.311 (0.280)			−0.380 (0.297)
App category FEs	No	Yes	Yes	No	Yes	Yes
Evaluator FEs	No	Yes	Yes	No	Yes	Yes
Constant	−5.213*** (0.270)	−20.920*** (2.371)	−19.060*** (2.705)	−5.602*** (0.409)	−22.019*** (2.754)	−20.470*** (3.028)
N	3,580	2,816	2,800	3,580	2,816	2,800

Notes. Logit regression; robust standard errors in parentheses. The baseline for comparison is SR1.
+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The estimated coefficients on SR2 and SR3 in Tables 6 and 7 suggest that SR1 and SR2 were best at avoiding false negatives. Specifically, compared with unfunded ideas submitted in SR1 and SR2, unfunded ideas submitted in SR3 were 11.6% more likely to become hits ($p = 0.060$). Finally, compared with SR1, unfunded ideas submitted in SR2 and SR3 were less likely to fail (by

8.6% ($p = 0.058$) and 17.3% ($p = 0.012$), respectively), suggesting that SR1 was most effective at selecting out true negatives.¹⁰
There are, of course, other factors affecting the likelihood of app idea failure or success that our controls seek to account for. As expected, an app ported from a different platform is less likely to fail and more likely to

Table 7. Results—Hits

Variable		DV: hit		DV: hit	
SR2: delegation	0.612 (0.607)	3.864* (1.789)	5.073* (2.061)	0.743 (0.765)	3.922* (1.846)
SR3: preselection	2.040*** (0.531)	7.522** (2.795)	9.484** (3.264)	2.328*** (0.669)	7.649** (2.916)
Funded	1.480*** (0.392)	0.337 (0.617)	0.259 (0.656)	2.041* (0.917)	0.471 (1.006)
SR2 × Funded				−0.379 (1.266)	−0.050 (1.388)
SR3 × Funded				−0.735 (1.022)	−0.240 (1.170)
Control variables					
Team size		−0.133 (0.164)	−0.122 (0.167)		−0.132 (0.164)
% Female		−0.525 (0.822)	−0.242 (0.788)		−0.529 (0.828)
Port		1.851** (0.577)	1.811** (0.617)		1.877** (0.573)
Sequel		1.696* (0.698)	1.516* (0.763)		1.697* (0.702)
Referral		0.346 (0.498)	0.406 (0.514)		0.351 (0.485)
Evaluator experience		−0.003 (0.002)	−0.005+ (0.003)		−0.003 (0.003)
Regime experience		0.002+ (0.001)	0.003* (0.001)		0.002+ (0.001)
Funded in category		−0.011+ (0.006)	−0.011+ (0.006)		−0.011+ (0.006)
Highly funded		−0.395 (0.820)	−0.422 (0.884)		−0.409 (0.839)
Limited feedback		−0.980 (0.682)	−1.069 (0.724)		−0.963 (0.690)
Word count			0.001 (0.001)		0.001 (0.002)
Clout			−0.034* (0.014)		−0.035* (0.014)
Positive emotion			0.082+ (0.044)		0.082+ (0.044)
Anger			0.011 (0.089)		−0.014 (0.091)
Sadness			0.442* (0.196)		0.451* (0.193)
Swear words			−0.061 (0.313)		−0.082 (0.335)
Nonfluencies			−0.701* (0.355)		−0.731* (0.362)
App category FEs	No	Yes	Yes	No	Yes
Evaluator FEs	No	Yes	Yes	No	Yes
Constant	−6.109*** (0.450)	−22.644*** (2.881)	−22.782*** (3.399)	−6.296*** (0.578)	−22.619*** (3.019)
N	3,580	2,387	2,372	3,580	2,387

Notes. Logit regression; robust standard errors in parentheses. The baseline for comparison is SR1.
+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

succeed or become a hit. Similarly, sequels are less likely to fail and more likely to succeed or become hits. However, our findings on the selection regimes are robust when controlling for these factors. Another plausible explanation is competition for attention among consumers or different resources available in the category (see, e.g., Ketkar and Workiewicz 2022); that is, there

could be more competition from other apps in the same category, both within the accelerator and in the market, in different selection regimes that could explain our results. The selection regime results hold even when including app category fixed effects and the previous number of apps funded by AppCampus in the focal app idea’s category (with the latter having a marginally

significant negative effect on the likelihood of the focal app idea becoming a hit). Of the remaining control variables, team size and referral are associated with a lower likelihood of failure but not with becoming successful or a hit, whereas greater regime experience is associated with a lower likelihood of failure and a higher likelihood of hits. Evaluator experience, on the other hand, is associated with a higher likelihood of app idea failure, as is an app being highly funded. Higher measures of the positive emotion and sadness LIWC dimensions seem correlated with a lower likelihood of failure and a higher likelihood of successes and hits. Higher scores on the clout dimension are associated with a lower likelihood of success or the app idea becoming a hit, whereas higher scores on the anger dimension are associated with a lower likelihood of failure. Greater use of swear words is marginally associated with a greater likelihood of failure, whereas greater resort to nonfluencies in the app idea description is associated with a lower likelihood of the idea becoming a hit.

Alternative Explanations

We address four plausible alternative explanations besides changes in selection regime that may explain our findings: (1) changes in treatment effects, (2) changes in the submission pool, (3) learning, and (4) market evolution. These are discussed in detail in Online Appendix H and summarized in Table 8.

Underlying Mechanisms

Now that we go more for quality than quantity, it's so obvious that the long tail is there still, you can't escape the long tail. If I were to write a paper on that specific item, the heading would be: You can't escape the long tail. —AppCampus CEO

The results so far show that SR3 resulted in a greater likelihood of selection errors being made, despite the

submission pool in this regime being better than in the first two selection regimes. These results are robust when controlling for alternative explanations, innate differences between evaluators and app categories, and language characteristics of the description of the project as submitted by its developers. Two plausible mechanisms can account for why SR3 had worse results, given the pool of available submissions: (1) mean reversion in combination with an increased emphasis on applicant track record and (2) adverse selection enabled by a more stringent selection process. We elaborate on each of these below and consider the extent to which they resulted from changes in evaluation criteria during SR2 versus the additional screening steps introduced in SR3.¹¹

Mean Reversion in Combination with Increased Emphasis on Applicant Track Record.

The change in evaluation criteria during SR2, which increased the emphasis on applicant track record, in combination with the pre-screening approach introduced in SR3 set the bar high for submissions to be introduced into the consideration set with the aim of tightening the quality distribution. Several months into SR2, the AppCampus screening team began to rely on developer team track record as a key evaluation criterion and weeded out many teams without one, attributing the prior success to teams and their capabilities: “The bar is higher. We are choosing, in a way, safer bets ... If somebody has already proven that they can make a successful mobile app, I think it's easier for us to believe that they would make another one. And that's most of the cases” (Screening team member). This rationale was echoed by many other members of AppCampus, including the CEO: “[Developers with track records] have already done something, and they might have found the recipe how to become successful.” Using the previous record of accomplishment can lead to benefits if it is informative about the

Table 8. Overview of Alternative Explanations and How They Were Addressed

Alternative explanation	Concern addressed	Approach taken
Changes in treatment effect	Whether differences in outcomes arise from changes in treatment received rather than selection process differences.	Used coarsened exact matching (CEM) to match treated and untreated observations on all non-LIWC idea-level variables, and analyzed treatment effects across regimes. Results showed no evidence of changing treatment effects driving outcomes.
Changes in submission pool	Whether differences in outcomes are driven by variations in the quality of submissions across regimes.	Compared AppCampus performance to counterfactual portfolios using a bootstrapping approach. Results indicated submission pool quality impacted unfunded app performance, but regime-specific effectiveness influenced funded app failure rates.
Learning	Whether individual or organizational learning influenced selection effectiveness over time.	Included learning-related controls in regression models. Results showed significant differences in selection regime outcomes persisted even after accounting for learning effects.
Market evolution	Whether changes in the broader mobile app market influenced submission quality or outcomes.	Included month fixed effects and controls for market share and app population. Results were robust to these additions, with no evidence of market evolution driving the findings.

Note. Online Appendix H addresses these alternative explanations in detail.

probability of a follow-on success. But teams that were successful once may have difficulty following that up, and when this escalates to over-reliance on previous records of accomplishment, problems can occur.

To evaluate the extent to which our results are driven by the changes in evaluation criteria during SR2 versus the change in the selection process from SR2 to SR3, we looked through the evaluation forms completed by the AppCampus screening team and identified how these changed over time. Specifically, the initial evaluation criteria used throughout SR1 and the first part of SR2 were whether the app was “innovative/first to market,” “differentiated/not available on other platforms,” “supported key Windows Phone features,” “demonstrated design elegance and technical quality,” and “had potential to drive ecosystem momentum through device sales or high user numbers.” On June 27, 2013, these criteria changed to the extent to which the submission showed “differentiation/innovation,” “exclusive features,” “mass-market appeal,” “developer track record,” and a “convincing proposal for app development and user acquisition.” As this change may have influenced the selection of ideas and the resulting selection errors both in the latter part of SR2 and in SR3, regardless of any changes in the structure of the selection process, the first three columns of Table 9 present the results of fully specified models but with the variables capturing changes in selection regimes and their interactions with *funded* replaced by a *new evaluation template* dummy variable equal to one for ideas submitted on or after June 27, 2013, and zero otherwise, and its interaction with *funded*. As can be seen in columns 1–3 of Table 9, making these changes produces results suggesting no difference in the likelihood of nonfunded submissions before or after them becoming failures, successes, or hits. For funded submissions, there is a marginally significant difference ($p = 0.058$) in their likelihood of failure after the change to the new evaluation template, with no differences in the likelihood of funded app idea success or becoming a hit. Thus, although the change in evaluation criteria does seem to have somewhat contributed to the greater likelihood of false positives in SR3, it does not appear, by itself, to explain the greater likelihood of false negatives nor the lower likelihood of true positives in the results in Tables 5–7.¹² Although the strengthening of evaluation criteria would by itself be expected to result in a greater share of false negatives, even if the selection regime was otherwise unchanged, the results in Tables 5–7 and 9 show it only seemed to make a substantial difference once it was reinforced by the change to SR3 extending the number of times that a submission had to clear the “track record” bar to be funded.

Our empirical work highlights Harrison and March’s (1984) observation, where evaluators may select the project with the highest estimate. However, because each estimate has a noise component, selecting the best

tends to be associated with more postdecision surprises. Prior success may be over-attributed to the team’s capabilities rather than the luck and good fortune that helped them in the process (see also Denrell and Liu 2012, 2021). With regression toward the mean, this then leads to lower-than-expected outcomes. By focusing too much on past track record, the screening team may have failed to appreciate that the developer’s next project may not be as successful as their earlier success. Furthermore, ignoring submissions by teams lacking such a track record, regardless of the app idea’s other qualities, risked missing future hits. Looking through the comments on false negatives—rejected opportunities that later had more than one million downloads—we do find some support for this argument. Of the twelve false negatives in SR3, a common reason given for rejection was “No track record given.”

To further investigate whether there is evidence of this mechanism in our data, we examined the reasons for rejecting each of the 26 false-negative cases in our data set. These reasons across the three selection regimes are summarized in Table 10. In SR1, two of the six false-negative cases were proposed for funding by the investment committee as exceptions to the “no porting cases” policy in force at the time but were rejected by the steering board, with “lack of innovation” and “execution barriers” being the next two most common reasons. In SR2, “lack of innovation” was the most common reason for rejection, followed by “lack of mobile development skills.” In SR3, the most common reason for rejection was “lack of track record,” followed by “borderline cases” in which the evaluation comments, though positive on the whole, stated that they were not fully convinced by the submission, leading to eventual rejection.

The patterns of rejection reasons for submissions that became successful apps are consistent with the qualitative evidence and mean reversion arguments: false-negative cases shifted from a nearly exclusive focus on evaluating the strengths and weaknesses of the proposed app idea in SR1, to most of these cases being rejected because of the team’s lack of track record in SR3.

Within-Type Adverse Selection Enabled by a More Stringent Selection Process. SR3 set the bar high for submissions to tighten the quality distribution because of a combination of higher emphasis on developer track record and greater hierarchy in the selection process. However, developers with established track records are likely to have developed their craft on competing Android and iOS app stores, which had much larger user install bases because of their earlier entry. Top developers targeted by AppCampus in this regime

Table 9. The Effects of Changing Evaluation Criteria

Variable	New evaluation template			Post-blog post		
	DV: Failure	DV: Success	DV: Hit	DV: Failure	DV: Success	DV: Hit
Funded	−2.967*** (0.318)	1.561* (0.656)	0.220 (0.980)	−2.976*** (0.309)	1.493* (0.618)	0.026 (0.948)
New eval. template	−0.549 (0.344)	0.591 (0.813)	0.346 (0.824)			
New eval. template × Funded	0.881+ (0.464)	−0.625 (1.030)	0.243 (1.307)			
Post-blog post				−0.735+ (0.386)	0.187 (0.779)	0.650 (0.702)
Post-blog post × Funded				1.009* (0.466)	−0.496 (0.990)	0.482 (1.271)
Control variables						
Team size	−0.125* (0.057)	−0.028 (0.108)	−0.116 (0.158)	−0.126* (0.056)	−0.023 (0.110)	−0.116 (0.157)
% Female	0.236 (0.327)	−0.679 (0.664)	−0.281 (0.734)	0.247 (0.329)	−0.681 (0.666)	−0.265 (0.733)
Port	−1.716*** (0.377)	1.909*** (0.482)	1.767** (0.597)	−1.695*** (0.373)	1.911*** (0.474)	1.729** (0.586)
Sequel	−1.423** (0.491)	1.199 (0.797)	1.668* (0.777)	−1.377** (0.483)	1.182 (0.795)	1.649* (0.771)
Referral	−0.455* (0.180)	0.264 (0.387)	0.543 (0.472)	−0.457* (0.178)	0.241 (0.366)	0.538 (0.479)
Evaluator experience	0.000 (0.000)	0.000 (0.001)	0.002 (0.001)	0.000 (0.000)	0.001 (0.001)	0.002 (0.001)
Regime experience	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)
Funded in category	0.001 (0.002)	−0.004 (0.004)	−0.004 (0.005)	0.001 (0.002)	−0.004 (0.004)	−0.004 (0.005)
Highly funded	0.604* (0.271)	−0.879 (0.716)	−0.495 (0.905)	0.601* (0.269)	−0.878 (0.713)	−0.476 (0.894)
Limited feedback	0.423+ (0.257)	−1.057* (0.537)	−1.142+ (0.693)	0.377 (0.262)	−1.046* (0.527)	−1.167 (0.719)
Word count	−0.000 (0.001)	−0.004+ (0.002)	0.001 (0.001)	0.000 (0.001)	−0.004 (0.002)	0.001 (0.001)
Clout	0.005 (0.006)	−0.022* (0.011)	−0.032* (0.014)	0.005 (0.006)	−0.022* (0.011)	−0.032* (0.014)
Positive emotion	−0.060** (0.017)	0.075* (0.031)	0.085+ (0.045)	−0.060** (0.017)	0.075* (0.031)	0.084+ (0.044)
Anger	−0.090* (0.037)	−0.054 (0.081)	0.004 (0.092)	−0.088* (0.036)	−0.055 (0.081)	0.001 (0.091)
Sadness	−0.217* (0.104)	0.343* (0.162)	0.398 (0.194)	−0.215* (0.104)	0.342* (0.163)	0.400* (0.194)
Swear words	0.436 (0.307)	0.217 (0.273)	−0.093 (0.364)	0.405 (0.313)	0.234 (0.272)	−0.082 (0.374)
Nonfluencies	0.225 (0.170)	−0.308 (0.274)	−0.630+ (0.377)	0.235 (0.171)	−0.297 (0.280)	−0.628+ (0.376)
App category FEs	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes
Constant	17.544*** (1.007)	−16.189*** (1.611)	−16.921*** (2.068)	17.644*** (1.017)	−15.799*** (1.549)	−17.257*** (1.904)
N	3,534	2,800	2,372	3,534	2,800	2,372

Note. Logit regression; robust standard errors in parentheses.
+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

would also likely have access to alternative funding sources for their app ideas, whether from their earnings or other investors. Once the increased focus on track record as a key selection criterion was broadcast in the AppCampus blog post on August 28, 2013, if such developers had a better understanding of the quality of

their app ideas compared with AppCampus evaluators, they could potentially have exploited their information advantage to use AppCampus for funding app ideas they considered to be lower quality compared with other app projects that they had in development, and so less likely to be funded by other means. In other words,

Table 10. False-Negative Rejection Reasons

Measure	SR1: investment board	SR2: delegation	SR3: preselection	Overall
False-negative successes ^a	6	8	12	26
Rejection reason				
Lacks track record	0	1	5	6
Lacks innovation	1.5	3.5	0	5
Borderline case	0	0	4	4
Steering board decline	2	1	0	3
Lacks mobile skill	0	2.5	0	2.5
Execution barriers	1.5	0	0	1.5
Out of scope	1	0	0	1
Too niche	0	0	1	1
Accepted another app	0	0	1	1
Lack of response	0	0	1	1
False-negative hits ^b	3	4	9	16
Rejection reason				
Lacks track record	0	1	3	4
Lacks innovation	1	2	0	3
Borderline case	0	0	3	3
Steering board decline	1	0	0	1
Lacks mobile skill	0	1	0	1
Out of scope	1	0	0	1
Too niche	0	0	1	1
Accepted another app	0	0	1	1
Lack of response	0	0	1	1

^aApp ideas rejected by AppCampus that went on to achieve at least 500,000 downloads.
^bApp ideas rejected by AppCampus that went on to achieve at least one million downloads.

they may not have submitted their best projects to AppCampus. The potential for this would have further increased with the change to the SR3 selection process resulting in the track record criterion being evaluated multiple times by different evaluators. AppCampus’s reliance on the team’s track record as a key criterion in this regime could thus have backfired because of within-type adverse selection (see, e.g., Huangfu and Liu 2023, Nguyen and Tan 2023), further explaining the regime’s lower performance in avoiding false positives and selecting true positives. This concern was raised by a member of the AppCampus screening team: “The guys with the brightest ideas and access to more resources and with more entrepreneurial experience, they may already have the iPhone App or the Android App, so it’s a bit difficult to get these guys to apply for the program or then, if they apply, then maybe they are already in a stage where they don’t really need us anymore.”

As the potential mechanism sketched out above requires the submitters with multiple funding options to know that their submission is likely to be viewed more favorably by the focal organization, we first investigate the effects of the explicit announcement of the increased weight given to developer track record in the blog post on the AppCampus website on August 28, 2013. Columns 4–6 of Table 9 present the results of fully specified models but with the variables capturing changes in selection regimes and their interactions with *funded* replaced by a *post-blog post* dummy variable equal to one for ideas submitted on or after August 28, 2013, and zero

otherwise, and its interaction with *funded*. These results show that nonfunded submissions after this change were marginally less likely to fail ($p = 0.057$), and the effect of funding on the likelihood of funded app failure was similar, though weaker, to the results for SR3 and its interaction with funding presented in Table 5. However, this change had no effect on the likelihood of nonfunded submissions succeeding or becoming a hit, nor on the likelihood of funded submissions achieving these outcomes. Overall, similar to the results for the changes in evaluation criteria presented in columns 1–3 of Table 9, this announcement of the changed evaluation criteria does seem to have contributed to the increased likelihood of false positives observed in SR3 but does not seem to explain SR3’s higher likelihood of false negatives nor its lower likelihood of true positives.¹³

To further investigate evidence of the adverse selection mechanism in our data, we examined each of the 254 false-positive cases. Using the data on the progress of each idea from submission to either release or termination, we were able to classify false-positive cases as having failed for one of three reasons: achieving fewer than 10,000 downloads postrelease; being terminated after the app idea had passed the design milestone and the developer had been paid 30% of the total grant amount; and being terminated before the app idea had passed the design milestone, without the developer having been paid any of the grant amount. The latter two categories contain cases that were terminated because of the app idea not meeting the quality

standards to pass the next required milestone (release candidate and design milestones, respectively), those that were terminated at the developer's request, and those that were terminated because the developer had stopped responding to communications from AppCampus. These data allowed us to see the extent of interaction between the AppCampus teams and the application developers throughout the process, as well as the share of false-positive cases that received additional support from AppCampus, either through participation at the AppCademy residential training camps or through the provision of marketing support, mostly in the form of free impressions on an app advertising network.

The resulting summary statistics for each selection regime are presented in Table 11. The share of false positives because of low number of downloads increased from 48.9% to 59.6% between *SR1* and *SR3*, whereas *SR2* saw the highest share of false positives occurring without the app in question being released, but after the developer had received 30% of the grant amount (29.4%). *SR1* saw the highest share of false positives because of projects being terminated prior to passing the design milestone, without the developer receiving any of the grant amount (27.7%). Although these statistics are not proof of adverse selection becoming more of

an issue in *SR3*, they are consistent with that possibility as substantially more failed app projects selected for funding in this regime failed not because of their developer's inability to deliver the level of design and technical quality expected, but rather because of poor postrelease performance.¹⁴

Further examination of the comments from the AppCampus team on each false-positive case provided additional supporting evidence for the adverse selection mechanism. Although none of the comments on cases in *SR1* mentioned aspects of the team's performance consistent with adverse selection, there was one such case in *SR2* ("The title is too good to be doing so poorly! Part of the issue is price, but I also think they just haven't done much to promote this title either. Need to get them more engaged"), and six such cases in *SR3* (e.g., "They are working on the version 3 already and are not interested on the version 2 anymore"; "In terms of studio priorities [they] have moved on from the project under discussion. [They] are working on other games now"; "Team's commitment seems a bit low"; "The team has been approached by a number of publishers over the past couple of months and as a result of those conversations have decided that they cannot bring this title to Windows Phone first").

Table 11. False-Positive Failure Reasons and Summary Statistics

Measure	<i>SR1</i> : investment board ^a	<i>SR2</i> : delegation	<i>SR3</i> : preselection	Overall
Funded	143	100	184	427
False positives	94	51	109	254
False positives %	65.7%	51.0%	59.2%	59.5%
<10K DL ^b	46	28	65	139
% of false positives	48.9%	54.9%	59.6%	54.7%
Av. review rounds ^c	6.61	6.96	7.74	7.21
% AppCademy ^d	47.8%	21.4%	20.0%	29.5%
% Ad support ^e	15.2%	7.1%	0%	6.5%
Failed post-30% paid ^f	21	15	20	56
% of false positives	22.3%	29.4%	18.3%	22.0%
Av. review rounds	5.14	5.07	4.85	5.02
% AppCademy	42.9%	13.3%	15%	25.0%
Failed pre-30% paid ^g	26	8	24	58
% of false positives	27.7%	15.7%	22.0%	22.8%
Av. review rounds	0.54	0.13	0.38	0.42

^aIn this selection regime, one false-positive case is unclassified as the developer achieved all the milestones and produced an app ready for release, only for problems to arise with the complementary hardware that the app was designed to work with, making the release unviable.

^bFunded app projects that were released but generated fewer than 10,000 downloads.

^cReview rounds of developer team vetting documents, app design documents, and released candidate software by AppCampus staff.

^d% AppCademy refers to the percentage of apps in a given failure reason category whose developers attended a residential two- or four-week training event at AppCampus.

^e% Ad support refers to the percentage of apps that received marketing support postrelease, provided mostly through free app advertising network impressions.

^fThis failure reason category covers all cases in which the developer received 30% of the grant amount but did not proceed to release the app and consists of the following reasons for failure: failed release candidate milestone, developer requested termination after passing the design milestone and being paid 30% of the grant amount, and developer unresponsive after passing the design milestone and being paid 30% of the grant amount.

^gThis failure reason category covers all cases in which the developer did not make it past the design milestone and did not get paid any of the grant amount and consists of the following reasons for failure: failed vetting milestone, failed design milestone; developer requested termination before passing the design milestone and being paid 30% of the grant amount, developer unresponsive before passing the design milestone and being paid 30% of the grant amount, and developer released app without AppCampus approval before passing the design milestone and being paid 30% of the grant amount.

As in the mean reversion mechanism above, a combination of changing evaluation criteria and the additional layers of screening introduced in *SR3* seems to produce this mechanism. Empirically, the results presented in Tables 5–7 and 9 suggest that although the change in selection criteria drives the greater share of false positives in *SR3*, the lower performance of *SR3* in the selection of true positives is observed only once these changes are compounded by increased hierarchy in the selection process.

Discussion

Rich conceptual, modeling, and experimental work has considered how to organize for selecting innovation projects (Csaszar and Eggers 2013), yet we know little about how organizations adapt to improve their selection and the effect of this on selection errors (see Keum and See (2017) and Eklund (2022) for exceptions). Our qualitative work documents how selection regimes evolved and adapted in structure, the evaluation criteria used, and the selectors' and entrepreneurs' divergent perspectives of the selection regimes. Despite changes to enhance selection regime performance, our quantitative findings reveal similarity in overall selection errors across these regimes. However, this similarity conceals contrasting effects of the changes made on the pool of submitted ideas and the effectiveness of selection among these ideas. Although eliciting a higher-quality pool of submitted ideas, the third and most selective regime was less effective in avoiding false positives and negatives than its predecessors. By combining our qualitative and quantitative data, our findings suggest two mechanisms that explain this: (1) mean reversion combined with an increased emphasis on applicant track record and (2) adverse selection caused by a more stringent selection process, which led some experienced teams to use the accelerator for projects where they did not fully commit. We elaborate on the theoretical implications for evaluation and selection literatures below. Finally, we discuss how using thick qualitative and quantitative data from a single accelerator may limit the generalizability of our findings, which future research could address.

Theoretical Implications

Most empirical studies of selection study a single regime (see, e.g., Criscuolo et al. 2021) or compare regimes across organizations (e.g., Csaszar 2012). This study contributes to the literature by highlighting adaptation in selection processes and demonstrating that organizations respond to emerging challenges by making bundled changes to their selection regimes. This underscores flexible organizational design, where changes are made in response to real-time feedback and evolving circumstances rather than through a static, predetermined approach. We find adaptation within organizational selection processes

without evidence of effective learning, possibly because feedback from the market on what works can take a long time to manifest. It is important to remember that the selection regimes we examine do not precisely align with those in existing studies (see Sah and Stiglitz 1986, Knudsen and Levinthal 2007, Csaszar 2012). Specifically, AppCampus incorporated specialists with overlapping yet complementary expertise to leverage diverse experiences and viewpoints and used multiple criteria for evaluating which projects to fund. Additionally, well-intended changes to selection regimes to address emerging issues may be bundled, making it more challenging to foresee outcomes and leading to unintended consequences.

Our findings suggest selection regime changes can affect both the pool of ideas submitted to the organization and the regime's performance in selecting among them, which has implications for work on selection that has treated the pool of ideas as fixed. For instance, Sah and Stiglitz's (1986) predictions about greater hierarchy in the selection process being associated with more true positives hold in our data, as they do in Csaszar (2012), but not because the *SR3* selection process is better than the prior, less hierarchical ones, at finding true positives. Instead, this more hierarchical process resulted in a greater likelihood of true-positive outcomes because of a higher-quality set of ideas self-selecting into it despite being less effective at selecting among these ideas. Our results also complement and contrast with the findings of Keum and See (2017). Although that paper hypothesizes and finds increased hierarchy of authority having a detrimental effect on idea generation, primarily in terms of the number of alternatives generated by participants in a laboratory experiment, we find that increased selection process hierarchy in *SR3* has a positive effect on the pool of ideas submitted for evaluation in terms of their likelihood of avoiding failure and being successful. Our work thus provides a complementary perspective on the effects of the selection process hierarchy on the pool of alternatives when these come from outside the focal organization, where the mechanism of self-censoring to conform to supervisor preferences proposed in Keum and See (2017) is likely less relevant. Similarly, our findings of the effects of hierarchy on selection among alternatives contrast with those in Keum and See (2017), likely because the reduction of bias in the evaluation of own ideas that drives their findings is ruled out by design in the organization we study, as nobody involved in AppCampus operations could submit app ideas for evaluation. Our findings therefore contribute to prior work on selection by helping us to understand better the likely effects of selection regimes both on the pool of alternatives submitted to the organization and on errors in selecting between these alternatives in settings where innovation ideas may come from outside of the organization (e.g., Pahnke et al. 2015, Lifshitz-Assaf et al. 2021).

Considering both changes in evaluation criteria and the decision-making structure over time, we find evidence of two novel mechanisms that appear to make selection among the higher-quality set of alternatives submitted in SR3 more challenging. The first is a variant of mean reversion of performance in combination with an increased emphasis on applicant track record. By separating innate attributes of projects and the associated applicants rather than considering the quality of the project as a unidimensional measure (Knudsen and Levinthal 2007), we can distinguish how evaluation can have unintended consequences. When evaluators placed heavy emphasis on teams' prior track record, the quality of the idea itself lost some of its significance. This parallels how venture capitalists (VCs) often hail the team as more important than the idea. The flip side of this argument is that if the process is more random than assumed (Denrell and Liu 2012), then past team track records may be less informative about future performance and thus given too much weight in the decision.¹⁵ The second mechanism of within-type adverse selection from SR3 targeting teams who likely have multiple outside options for funding their *best* ideas and who potentially use AppCampus to fund their *inferior* ideas also has important implications. Recent work on within-type adverse selection (e.g., Huangfu and Liu 2023, Nguyen and Tan 2023) considers ways information availability across markets and bundled trades may reduce its detrimental effects. These potential solutions may be more or less viable based on how the organization soliciting and evaluating ideas is designed. In the case of AppCampus, the combination of their focus on selecting among particular app ideas and the grant-based support provided precluded using a bundled trade approach. However, accelerators that provide their services in exchange for an equity stake in the firms that they support effectively benefit from the success of all of the selected firms' current and future projects and so may be able to reduce the likelihood of such within-type adverse selection. Even if the context we study is one in which innovation projects are submitted from outside of the selecting organization, both of the above mechanisms affect within-firm selection regimes, as evaluation criteria may not always reflect the best indicators of desired outcomes and because employees may have options for finding greater support and rewards for their best ideas elsewhere. The extents to which the above mechanisms are affected by the openness of the organization's innovation process and the relative standing of the organization (and its ecosystem) are interesting questions for future research.

Our findings can be seen in the light of previous empirical work on the value of accelerators (see, e.g., Hallen et al. 2020, 2023) and how they organize (Cohen et al. 2019). For example, Hallen et al. (2020) use inverse probability treatment weights to estimate the effects of *top* accelerators and find *mixed* effects for their impacts

on venture development. Similarly, Yu (2020) examines how accelerator participation impacts venture performance by leveraging quasi-experimental methods, finding that accelerators provide feedback that helps founders resolve uncertainty about their ideas more quickly, leading to earlier and more frequent closures of ventures, as well as reduced funding amounts for those that do close. In our setting, by contrast, the effects of changing selection regimes come from a combination of changes in the effectiveness of the selection process and in the pool of alternatives available for selection rather than changes in the value that the accelerator brought to the funded teams. In a related paper, Cohen et al. (2019) find huge variation *across* accelerators in their organizational design. Using unique within-accelerator data, we show that there may also be major variation in design *within* an accelerator over time and that such changes are likely to affect both the entrepreneurs applying and the effectiveness of selecting among them. Although the changes that we observe generally proceeded in the direction of greater selectiveness and more hierarchy in the selection process, it is not obvious whether we would expect opposite findings if the progression between selection regimes had been reversed, as starting with a more hierarchical selection regime with strict evaluation criteria may establish a harder-to-change perception among potential applicants of what the accelerator is looking for, potentially reducing the variety that organizations later get exposed to (Park et al. 2024). Innovation results from considering many different options, meaning most ideas fail or are rejected on the journey toward adoption. Analyzing all projects shows that some rejected ideas succeeded. Of the 32 submitted app ideas that achieved more than one million downloads, 16 were funded by AppCampus. In every selection regime, the best-performing unfunded (false-negative) submission generated downloads equal to at least 40% of the total downloads generated by all funded submissions in that regime. This highlights the difficulty of separating "good" from "bad" projects and the magnitude of potential consequences. This challenge is consistent with Yu's (2020) emphasis on the conditional and context-dependent nature of accelerator outcomes, as accelerators must navigate inherent uncertainty when selecting ventures to support.

Limitations and Future Research

Our paper has several limitations, most notably that of studying a single accelerator. Little research exists on how rejected ideas lead to commercial success for many projects in other settings, so it becomes difficult to know how well the findings generalize. AppCampus used different ways to organize and changed multiple aspects of its selection regimes. Despite ruling out several alternative explanations in our post hoc tests, several limitations remain and need to be addressed in future work.

First, a limitation of our study is the presence of simultaneous events in the environment that could confound our results. Although we have attempted to mitigate these confounding influences by incorporating fixed effects for different app categories, controlling for the performance of the Windows Phone ecosystem over time, and examining the interplay between selection regimes and treatment effects, these measures may not fully capture the complex dynamics at play. The broader external environment, including market fluctuations, technological advancements, regulatory changes, and competitive actions, can significantly impact the availability and nature of projects and the effectiveness of selection regimes. Future research could leverage exogenous shocks in the external environment to address this limitation. Such shocks could include sudden regulatory changes, significant technological breakthroughs, economic crises, or shifts in consumer preferences, which are external to the organization and thus provide a source of variation independent of the internal processes under study. Analyzing how organizations adapt their selection regimes in response to these natural experiments could yield valuable insights into the robustness and adaptability of different selection strategies under varying conditions.

Second, a difficulty with considering false negatives is that rejected ideas may pivot because of feedback received, which makes it more difficult to trace such ideas (Tidhar and Eisenhardt 2020). We painstakingly coded the fate of all rejected projects, but we might have missed projects that transformed so much that we could no longer trace them. If these could still be considered the same ideas as those evaluated by AppCampus, we would underestimate the true number of false negatives. This is difficult to rule out completely, but we would not expect this to differ systematically across the selection regimes.

Third, an inherent limitation of our study is that the modifications to the selection regimes were not exogenous but rather emerged from the organization's learning and adaptation processes. This characteristic inherently restricts our ability to draw strong causal inferences about the direct impact of these changes on organizational outcomes. Recognizing this limitation, future research would benefit from adopting an experimental approach. Such an approach could more precisely disentangle the separate and joint effects of changing evaluation criteria and selection regime structure on both selection effectiveness and outcomes and test our proposed causal mechanisms. Furthermore, extending the analysis to include a comparative study of various organizations with differing selection regimes would provide valuable empirical evidence to support or challenge our findings. Such comparative analyses, potentially leveraging longitudinal data, would enrich our understanding of how different organizational cultures, structures, and strategies influence the efficacy of selection regimes. It is

likely, however, that convincing organizations to run field experiments with different selection regimes will be challenging.

Conclusion

Managing idea selection and preventing false positives and negatives is one of the most critical organizational processes affecting both performance and employee morale (Giarratana et al. 2018). Our study suggests that mistakes happen frequently and highlights the potentially contrasting effects that changing selection regimes can have on the pool of ideas submitted for evaluation and the challenges of selecting among these ideas. We hope our findings help managers design and implement more effective selection processes and inspire future research to further expand our understanding of organizational selection processes and their intended and unintended effects.

Acknowledgments

The authors thank three anonymous reviewers and Senior Editor Stefano Brusoni for raising great questions and making constructive suggestions that have helped to substantially improve the paper. The authors are grateful for feedback on earlier versions of this manuscript from Carliss Baldwin, Matt Bothner, Thorbjørn Knudsen, Reddi Kotha, Chengwei Liu, Samuel MacAulay, Ammon Salter, Maciej Workiewicz, and seminar participants at Bocconi University the China Innovation and Entrepreneurship Seminar, ESADE, ESSEC, INSEAD, the NIMES workshop, Rotterdam School of Management, the SIE seminar, the Strategic Management Society annual conference 2021, Singapore Management University, Stockholm School of Economics, University of Groningen, University of Liverpool, and Yale University. Raluca Presecan and Subleen Kaur provided excellent research assistance. All errors are those of the authors alone (sadly!).

Endnotes

¹ Microsoft became AppCampus's sole funder after it acquired Nokia's Devices and Services division in April 2014.

² We were not present at the first AppCademy so were only able to interview two developer teams from that initial group of 13, and two interviews with developer teams involved in the later AppCademies did not take place because of a lack of time in the AppCademy schedule.

³ Although application ideas may have developed and changed between the submission of the idea to AppCampus and their eventual release (if it occurred), such changes are likely to have taken place for both funded and nonfunded ideas during all three selection regimes. We have no a priori reason to believe that there were systematic differences in the likelihood of such changes between the three regimes, nor any empirical evidence suggesting that this was the case, as we discuss later in the paper.

⁴ This data collection took place between November 2018 and April 2019, at which point, if released, the apps in question would have been available to download for between three and six years. As this is substantially longer than the estimated average timespan for a game released in this period to reach over 90% of its total cumulative downloads (App Annie 2016, p. 30), we believe that the

download data collected in this manner capture total cumulative downloads regardless of the release date of the app in question.

⁵ Note that although we know the exact Windows Phone store performance of apps funded by AppCampus, we do not have data on the exact number of downloads for funded apps on the iOS or Google app stores, nor for app ideas rejected by AppCampus. In these cases, we use publicly available information from iOS and Google app stores that provides a range of downloads they have achieved, for example, 500,000 to 1 million downloads, and take the minimum value of that range. For apps released on both the Windows Phone and other app stores, we use their download performance in their best-performing app store. For app ideas that were already released on other platforms prior to submission to AppCampus (ported apps), we consider only their postsubmission download performance.

⁶ Online Appendix D presents an alternative way to compare the selection regimes using confusion matrices. Online Appendix E presents the Epanechnikov kernel density plots of $\log(\text{downloads} + 1)$ by selection regime and funding status, treating the download numbers we have as continuous rather than categorical (see Endnote 5 above). The patterns presented by these plots are consistent with the statistics presented in Table 3.

⁷ The results for models with failure as the dependent variable presented below are robust to alternative definitions of failure, taking either 50,000 or 100,000 downloads as the cutoff below which an app idea is considered to have failed. These results are presented in Online Appendix F.

⁸ We exclude 18 observations from the analysis when these LIWC control variables are included because the app description in these cases was written in Mandarin, which LIWC is not able to process.

⁹ As we have small numbers of successes and hits in each regime, the logit models estimating the effects of our explanatory variables on these rare outcomes may be biased. We therefore perform robustness checks by reestimating our models using the penalized maximum likelihood method (Firth 1993), implemented using the *firthlogit* command in Stata, and the rare events logit model (King and Zeng 2001), implemented using the *relogit* command in Stata. The results from both approaches are consistent with those reported below. As our three dependent variables arise from the same selection process, we also check whether our results are robust to using models that explicitly take potential nonindependence between outcomes into account. Using a three-way multivariate probit model implemented with the *mprobit* command in Stata (Cappellari and Jenkins 2003) produces estimates that converge only if no control variables are used, corresponding to columns 1 and 4 of Tables 5–7. These results are consistent with those in Tables 5–7, except for the interaction between funding and selection regime 3 losing statistical significance ($p = 0.15$). Running three two-way models for all pairwise combinations of the dependent variables allows us to include all control variables except app category fixed effects and produces estimates that are fully consistent with our findings, and the same is true if these models are estimated using the *biprobit* command instead of *mprobit*.

¹⁰ If AppCampus were constrained to funding only a set number of submissions per regime, these results would then reflect not only selection regime effectiveness at avoiding false positives and false negatives but also changes in the pool of submissions across regimes. Although no such constraint was in place, we cannot rule out overall budget considerations playing some role in how many apps were funded in each regime, and we investigate the evidence for these findings being driven by the changing pool of submissions rather than by selection regime effectiveness further as an alternative explanation below.

¹¹ Another plausible mechanism is that evaluator incentive to free-ride on the work of other evaluators (e.g., Gibbons and Roberts 2013) increased for the AppCampus internal evaluators in SR3 because the

submissions they received had already been approved by two prior levels of hierarchy during the prescreening phase. We find no qualitative evidence for this mechanism among our interviews with the evaluators, nor any quantitative evidence consistent with it. Free-riding of this kind would imply a lower level of effort or diligence of evaluators in SR3 compared to the prior two regimes. Although we do not have comprehensive data on the time taken to evaluate submissions across regimes, which would be one potential measure of evaluation effort, we do have the text written by evaluators stating their evaluation of every submission allocated to them and their recommendation for or against funding. If AppCampus evaluators in SR3 were more likely to free-ride on the work of the prescreening evaluators, we would expect evaluation texts in this selection regime to be shorter compared to the previous two regimes. This is not the case. The average word count of evaluation texts falls from 21.115 to 18.864 words between SRs 1 and 2 (two-sample *t*-test with unequal variances, $p = 0.04$) but rises to 44.384 words in SR3 (two-sample *t*-test with unequal variances, $p < 0.001$ when compared against either SR1 or SR2). This pattern is also consistent when comparing the average word counts of only funded or only nonfunded applications across the three regimes.

¹² In results available on request from the authors, we also included *new evaluation template* in the models without its interaction with *funded* to investigate whether this change was associated with a change in the pool of submissions received by AppCampus. The estimated coefficient on *new evaluation template* does not approach statistical significance in any of the three models (lowest $p = 0.462$), suggesting that the change to the pool of submissions observed in SR3 was not driven by a change in evaluation template alone. The results in Tables 5–7 are also consistent when adding this variable as a control (results available on request from the authors). In that case it does not approach statistical significance by itself. We do not include this control in the main specifications reported in Tables 5–7 because it is highly correlated with the dummy variable for selection regime 3 ($\rho = 0.62$).

¹³ In results available on request from the authors, we also included *post-blog post* in the models without its interaction with *funded* to investigate whether this change was associated with a change in the pool of submissions received by AppCampus. The estimated coefficient on *post-blog post* does not approach statistical significance in any of the three models (lowest $p = 0.166$), suggesting that the change to the pool of submissions observed in SR3 was not driven by the announcement blog post alone. Adding a control variable for this change to the fully specified models in Tables 5–7 does not change the results. The *post-blog post* dummy itself never approaches statistical significance. We do not include this control in the main specifications reported in Tables 5–7 because it is highly correlated with the dummy variable for selection regime 3 ($\rho = 0.75$).

¹⁴ The statistics on average review rounds performed by the AppCampus team show an increase of more than one additional round between SR1 and SR3, with this increase likely balancing out the reduction in support in the form of AppCademy participation and advertising from SR1 to SR3. These patterns triangulate our impressions from qualitative data collection and the results of quantitative tests: although different aspects of postselection support were emphasized in the three regimes by both AppCampus and developers, the overall extent and quality of treatment did not vary substantially between them.

¹⁵ Even app market success stories such as the company Supercell, with multiple hits such as *Clash of Clans*, *Boom Beach*, *Clash Royale*, and *Brawl Stars*, seem to achieve this success through intense and structured experimentation with many parallel projects, knowing that only a fraction will succeed (Takahashi 2018). It is plausible that greater success could be achieved by reducing the cost of experimentation with multiple ideas instead of limiting the selection chances of developers without strong track records, as in SR3.

References

- App Annie (2016) App Annie 2015 retrospective. Accessed December 12, 2023, https://media.hotnews.ro/media_server1/document-2016-01-22-20745861-0-app-annie-2015-retrospective.pdf.
- Bian J, Greenberg J, Li J, Wang Y (2022) Good to go first? Position effects in expert evaluation of early-stage ventures. *Management Sci.* 68(1):300–315.
- Botelho TL, Abraham M (2017) Pursuing quality: How search costs and uncertainty magnify gender-based double standards in a multistage evaluation process. *Admin. Sci. Quart.* 62(4):698–730.
- Böttcher L, Klingebiel R (2025) Organizational selection of innovation. *Organ. Sci.* 36(1):387–410.
- 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.
- 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.
- Cappellari L, Jenkins SP (2003) Multivariate probit regression using simulated maximum likelihood. *Stata J.* 3(3):278–294.
- Christensen M, Knudsen T (2010) Design of decision-making organizations. *Management Sci.* 56(1):71–89.
- Christensen M, Dahl CM, Knudsen T, Warglien M (2023) Context and aggregation: An experimental study of bias and discrimination in organizational decisions. *Organ. Sci.* 34(6):2163–2181.
- Cohen SL, Bingham CB, Hallen BL (2019) The role of accelerator designs in mitigating bounded rationality in new ventures. *Admin. Sci. Quart.* 64(4):810–854.
- Criscuolo P, Dahlander L, Grohsjean T, Salter A (2017) Evaluating novelty: The role of panels in the selection of R&D projects. *Acad. Management J.* 60(2):433–460.
- Criscuolo P, Dahlander L, Grohsjean T, Salter A (2021) The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. *Organ. Sci.* 32(4):987–1008.
- Csaszar FA (2012) Organizational structure as a determinant of performance: Evidence from mutual funds. *Strategic Management J.* 33(6):611–632.
- Csaszar FA (2013) An efficient frontier in organization design: Organizational structure as a determinant of exploration and exploitation. *Organ. Sci.* 24(4):1083–1101.
- Csaszar FA, Eggers JP (2013) Organizational decision making: An information aggregation view. *Management Sci.* 59(10):2257–2277.
- Csaszar FA, Jue-Rajasingh D, Jensen M (2023) When less is more: How statistical discrimination can decrease predictive accuracy. *Organ. Sci.* 34(4):1383–1399.
- Cyert RM, March JG (1963) *A Behavioral Theory of the Firm* (Prentice-Hall, Englewood Cliffs, NJ).
- Dahlander L, Gann DM, Wallin MW (2021) How open is innovation? A retrospective and ideas forward. *Res. Policy* 50(4):104218.
- Denrell J, Liu C (2012) Top performers are not the most impressive when extreme performance indicates unreliability. *Proc. Natl. Acad. Sci. USA* 109(24):9331–9336.
- Denrell J, Liu C (2021) When reinforcing processes generate an outcome-quality dip. *Organ. Sci.* 32(4):1079–1099.
- Di Stefano G, Micheli MR (2023) To stem the tide: Organizational climate and the locus of knowledge transfer. *Organ. Sci.* 34(6):2436–2463.
- Dosi G, Levinthal DA, Marengo L (2003) Bridging contested terrain: Linking incentive-based and learning perspectives on organizational evolution. *Indust. Corporate Change* 12(2):413–436.
- Eisenhardt KM (1989) Building theories from case study research. *Acad. Management Rev.* 14(4):532–550.
- Eklund JC (2022) The knowledge-incentive tradeoff: Understanding the relationship between research and development decentralization and innovation. *Strategic Management J.* 43(12):2478–2509.
- Eklund JC, Kapoor R (2022) Mind the gaps: How organization design shapes the sourcing of inventions. *Organ. Sci.* 33(4):1319–1339.
- Ethiraj SK, Levinthal D (2004) Modularity and innovation in complex systems. *Management Sci.* 50(2):159–173.
- Firth D (1993) Bias reduction of maximum likelihood estimates. *Biometrika* 80(1):27–38.
- Giarratana MS, Mariani M, Weller I (2018) Rewards for patents and inventor behaviors in industrial research and development. *Acad. Management J.* 61(1):264–292.
- Gibbons R, Roberts J (2013) *The Handbook of Organizational Economics* (Princeton University Press, Princeton, NJ).
- Glaser BG, Strauss AL (1967) *Discovery of Grounded Theory: Strategies for Qualitative Research* (Aldine, Chicago).
- Hallen BL, Cohen SL, Bingham CB (2020) Do accelerators work? If so, how? *Organ. Sci.* 31(2):378–414.
- Hallen BL, Cohen SL, Park SH (2023) Are seed accelerators status springboards for startups? Or sand traps? *Strategic Management J.* 44(8):2060–2096.
- Harrison JR, March JG (1984) Decision making and post-decision surprises. *Admin. Sci. Quart.* 29(1):26–42.
- Huangfu B, Liu H (2023) Information spillover in multi-good adverse selection. *Amer. Econom. J. Microeconom.* 15(3):118–165.
- Joseph J, Gaba V (2020) Organizational structure, information processing, and decision-making: A retrospective and road map for research. *Acad. Management Ann.* 14(1):267–302.
- Kanze D, Huang L, Conley M, Higgins E (2018) We ask men to win & women not to lose: Closing the gender gap in startup funding. *Acad. Management J.* 61(2):586–614.
- Kerr WR, Lerner J, Schoar A (2014) The consequences of entrepreneurial finance: Evidence from angel financings. *Rev. Financial Stud.* 27(1):20–55.
- Ketkar H, Workiewicz M (2022) Power to the people: The benefits and limits of employee self-selection in organizations. *Strategic Management J.* 43(5):935–963.
- Keum DD, See KE (2017) The influence of hierarchy on idea generation and selection in the innovation process. *Organ. Sci.* 28(4):653–669.
- King G, Zeng L (2001) Logistic regression in rare events data. *Political Anal.* 9(2):137–163.
- Klingebiel R, Adner R (2015) Real options logic revisited: The performance effects of alternative resource allocation regimes. *Acad. Management J.* 58(1):221–241.
- Knudsen T, Levinthal DA (2007) Two faces of search: Alternative generation and alternative evaluation. *Organ. Sci.* 18(1):39–54.
- Levinthal DA (1997) Adaptation on rugged landscapes. *Management Sci.* 43(7):934–950.
- Levinthal DA (2017) Resource allocation and firm boundaries. *J. Management* 43(8):2580–2587.
- Li D (2017) Expertise versus bias in evaluation: Evidence from the NIH. *Amer. Econom. J. Appl. Econom.* 9(2):60–92.
- Lifshitz-Assaf H, Lebovitz S, Zalmanson L (2021) Minimal and adaptive coordination: How hackathons' projects accelerate innovation without killing it. *Acad. Management J.* 64(3):684–715.
- March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.
- March JG, Olsen JP (1975) The uncertainty of the past: Organizational learning under ambiguity. *Eur. J. Political Res.* 3(2):147–171.
- Mortensen CR, Cialdini RB (2010) Full-cycle social psychology for theory and application. *Soc. Personality Psych. Compass* 4(1):53–63.
- Mueller JS, Melwani S, Goncalo JA (2012) The bias against creativity: Why people desire but reject creative ideas. *Psych. Sci.* 23(1):13–17.
- Mueller JS, Wakslak CJ, Krishnan V (2014) Construing creativity: The how and why of recognizing creative ideas. *J. Experiment. Soc. Psych.* 51(March):81–87.

- Nguyen A, Tan TY (2023) Markets with within-type adverse selection. *Amer. Econom. J. Microeconom.* 15(2):699–726.
- Pahnke EC, Katila R, Eisenhardt KM (2015) Who takes you to the dance? How partners' institutional logics influence innovation in young firms. *Admin. Sci. Quart.* 60(4):596–633.
- Park S, Piezunka H, Dahlander L (2024) Coevolutionary lock-in in external search. *Acad. Management J.* 67(1):262–288.
- Pennebaker JW, Boyd RL, Jordan K, Blackburn K (2015) *The Development and Psychometric Properties of LIWC2015* (University of Texas at Austin, Austin).
- Piezunka H, Dahlander L (2019) Idea rejected, tie formed: Organizations' feedback on crowdsourced ideas. *Acad. Management J.* 62(2):503–530.
- Piezunka H, Schilke O (2023) The dual function of organizational structure: Aggregating and shaping individuals' votes. *Organ. Sci.* 34(5):1914–1937.
- Polidoro F (2020) Knowledge, routines, and cognitive effects in non-market selection environments: An examination of the regulatory review of innovations. *Strategic Management J.* 41(13):2400–2435.
- Reitzig M, Sorenson O (2013) Biases in the selection stage of bottom-up strategy formulation. *Strategic Management J.* 34(7):782–799.
- Rivkin JW, Siggelkow N (2003) Balancing search and stability: Interdependencies among elements of organizational design. *Management Sci.* 49(3):290–311.
- Sah RK, Stiglitz JE (1986) The architecture of economic systems: Hierarchies and polyarchies. *Amer. Econom. Rev.* 76(4):716–727.
- Sengul M, Almeida Costa A, Gimeno J (2018) The allocation of capital within firms: A review and integration toward a research revival. *Acad. Management Ann.* 13(1):43–83.
- Simon HA (1955) A behavioral model of rational choice. *Quart. J. Econom.* 69(1):99–118.
- Takahashi D (2018) Supercell CEO thrives on trusting the instincts of game developers. Interview with Ilkka Paananen. *VentureBeat* (March 29), <https://venturebeat.com/games/supercell-ceo-thrives-on-trusting-the-instincts-of-game-developers/>.
- Tidhar R, Eisenhardt KM (2020) Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strategic Management J.* 41(7):1245–1273.
- Tushman ML, Anderson P (1986) Technological discontinuities and organizational environments. *Admin. Sci. Quart.* 31(3):439–465.
- Yu S (2020) How do accelerators impact the performance of high-technology ventures? *Management Sci.* 66(2):530–552.

Dmitry Sharapov is an associate professor at Imperial College Business School, Imperial College London. His research focuses on competitive strategy, innovation management, and decision making under uncertainty. He received his PhD from the University of Cambridge.

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.