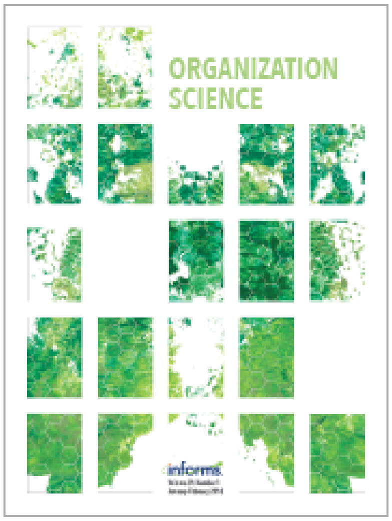


Organization Science

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

Learning from Mixed Signals in Online Innovation Communities

Christoph Riedl, Victor P. Seidel

ORGANIZATION SCIENCE
Vol. 29, No. 6, November–December 2018, pp. 1010–1032
ISSN 1047-7039 (print), ISSN 1526-5455 (online)
http://pubsonline.informs.org/journal/orsc/

Learning from Mixed Signals in Online Innovation Communities

Christoph Riedl,[a,b] **Victor P. Seidel**[c,d,e]

[a] D'Amore-McKim School of Business, and College of Computer and Information Science, Northeastern University, Boston, Massachusetts 02115; [b] Institute for Quantitative Social Science, Harvard University, Cambridge, Massachusetts 02138; [c] F.W. Olin Graduate School of Business, Babson College, Babson Park, Massachusetts 02457; [d] Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138; [e] Said Business School, University of Oxford, Oxford OX1 1HP, United Kingdom

Contact: c.riedl@neu.edu, https://orcid.org/0000-0002-3807-6364 (CR); vseidel@babson.edu, https://orcid.org/0000-0003-3066-7132 (VPS)

Received: June 9, 2016
Revised: June 18, 2017; November 19, 2017; February 22, 2018; April 11, 2018
Accepted: April 24, 2018
Published Online in Articles in Advance: October 3, 2018

https://doi.org/10.1287/orsc.2018.1219

Abstract. We study how contributors to innovation contests improve their performance through direct experience and by observing others as they synthesize learnable signals from different sources. Our research draws on a 10-year panel of more than 55,000 individuals participating in a firm-hosted online innovation community sponsoring creative t-shirt design contests. Our data set contains almost 180,000 submissions that reflect signals of direct performance evaluation from both the community and the firm. Our data set also contains almost 150 million ratings that reflect signals for learning from observing the completed work of others. We have three key findings. First, we find a period of initial investment with decreased performance. This is because individuals struggle to synthesize learnable signals from early performance evaluation. This finding is contrary to other studies that report faster learning from early direct experience when improvements are easiest to achieve. Second, we find that individuals consistently improve their performance from observing others' good examples. However, whether they improve from observing others' bad examples depends on their ability to correctly recognize that work as being of low quality. Third, we find that individuals can successfully integrate signals about what is valued by the firm hosting the community, not just about what is valued by the community. We thus provide important insights into the mechanisms of how individuals learn in crowd-sourced innovation and provide important qualifications for the often-heralded theme of "learning from failures."

Funding: This research was funded by the National Science Foundation [Grant IIS-1514283].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/orsc.2018.1219.

Keywords: innovation • crowdsourcing • online communities • contests • learning • knowledge

1. Introduction

Contest-based online communities are a form of crowd-sourced innovation in which individuals of online innovation communities (e.g., Jeppesen and Frederiksen 2006) repeatedly contribute to contests sponsored by firms (e.g., Jeppesen and Lakhani 2010, Terwiesch and Xu 2008) to generate ideas and solve research and development (R&D) problems. Contributors are attracted to join the community by the contests, and their membership subsequently fuels participation (Hutter et al. 2011, Majchrzak and Malhotra 2016). As individuals repeatedly participate, they receive a mix of evaluation signals about their own performance from both community members and the firm hosting the contest. They use these signals to learn how to improve their performance and ultimately increase their chances of winning a contest. Individuals can also observe and evaluate the openly disclosed work submitted by other community members. This provides them with additional signals for learning.

Learning has been found to be an important motivation for both joining online innovation communities (Lakhani and Wolf 2005, Roberts et al. 2006) and for continued participation in communities generally (Brown and Duguid 1991, Wenger 1998). Yet while the perception that one will learn has been widely documented as a motivation to join and contribute to online innovation communities, if and how individuals learn from such a mix of diverse information sources is unclear. Learning theories generally assume that signals are unambiguous and easy to interpret in guiding performance improvements (Argote 2012, Newell and Rosenbloom 1981). However, signals in contest-based online communities come from multiple sources and are less directive, making learning difficult. Individual learning in these contexts is an important topic because it is an important motivation to sustain their engagement in the community (Afuah and Tucci 2012, Brown and Duguid 1991, Dahlander and Piezunka 2014, Wenger 1998). Individual learning is also foundational for group and organizational learning (Argote and Epple 1990, March 1991). This study examines how individuals in online innovation communities improve their performance by synthesizing learnable signals from different sources.

Online innovation communities are becoming an increasingly relevant way for organizations to gain access to innovative ideas and solve R&D problems (Almirall et al. 2014, Boudreau et al. 2016, Majchrzak and Malhotra 2016). Examples include innovation in household electronics (Hutter et al. 2011), fashion (Brabham 2010, Füller et al. 2014), science (Riedl and Woolley 2017), motor vehicles (Seidel and Langner 2015), and software algorithms (Boudreau et al. 2016, Riedl et al. 2016). Though not formally organized as contests, similar contexts are also found on app development platforms. There, individual developers compete to have their work selected by the sponsoring firm, while other developers provide informative signals from which they can learn (Boudreau 2012).

Learning theory suggests that individuals learn from both direct experience with an activity (Arrow 1962, Epple et al. 1996, Foster and Rosenzweig 1995, Newell and Rosenbloom 1981, Rosen 1972) and indirect experience from observing others (Bandura 1977, Banerjee 1992, Bikhchandani et al. 1992, Hanna et al. 2014).[1] Although these learning models are diverse, a common assumption is that the key learning input is informative signals (Andrieux and Proteau 2016, Bandura 1977, Camerer 2011, Hanna et al. 2014).

The learning context determines which signals are (theoretically) available (Argote et al. 2003, Bandura 1977, Brown and Duguid 1991). However, the self-directed nature of human actions determines the signals to which an individual is exposed in the first place (Bandura 1977). Cognitive structures then determine which signals are noticed and what information can be extracted from them (Bandura 1977, Hanna et al. 2014).[2] Contest-based online communities offer two actions that determine the signals that individuals receive. Firstly, submitting an entry to a contest generates evaluation signals for one's own work (i.e., direct experience). Secondly, exploring and evaluating the work submitted by others generates signals for observational learning (i.e., indirect experience). What and how individuals learn from these signals critically depend on what they pay attention to (Hanna et al. 2014), as well as the cognitive structures they have in place to synthesize informative data from the signals (Andrieux and Proteau 2016, Van Knippenberg et al. 2015). As a result, learning rates may not be constant over time because individuals develop cognitive structures that help them synthesize learnable signals from these mixed signals from different information sources.

The context also dictates what it means to be successful. In contest-based online communities (as in many firm-hosted communities in general), individuals are faced with the challenge of learning both the preferences of the community and the preference of the firm (Bauer et al. 2016, Boudreau et al. 2011, Dahl et al. 2014, Dahlander and Piezunka 2014, Huang et al. 2014). To be successful in these communities, individuals need to integrate information from the performance evaluation they receive to learn about both community and firm preferences. This may be particularly challenging when individuals receive mixed signals, as information from different sources often diverges (Hildebrand et al. 2013).

We study a 10-year panel of more than 55,000 individuals participating in a firm-hosted online community that sponsors creative t-shirt design contests. Our data set contains almost 180,000 submissions that reflect signals of direct performance evaluation from the community and the firm. Our data set also contains almost 150 million ratings that reflect signals for learning from observing the work others. That is, we use the casting of a rating as a measure for a unit of indirect experience. A key characteristic of our setting (like many other innovation settings) is that when the work of others is being observed (i.e., at the moment the indirect experience is made), it is not yet known whether that observed work will be successful. Using this longitudinal data, we show how individuals in contest-based online communities synthesize learnable signals from different information sources for self-guided learning and performance improvements.

We have three key findings. First, we find that individuals participating in contest-based online communities need to overcome a period of initial investment with decreased performance, prior to seeing positive returns. This finding is contrary to other studies that report faster early learning when improvements are easiest to achieve, followed by diminishing returns later (Newell and Rosenbloom 1981). Second, we find that individuals consistently improve their performance from observing others' good examples. However, whether they improve from observing others' bad examples depends on their ability to correctly recognize that work as being of low quality and correctly identify the sources of error in these bad examples. We thus provide important qualifications for the often-heralded theme of "learning from failures." Third, we find that individuals can successfully integrate signals not only about what is valued by the community, but also about what is valued by the firm hosting the community.

We make two primary contributions to research on crowdsourced product innovation and online innovation communities in general. First, a large body of community research has focused on how individuals learn from direct interaction and knowledge sharing with other community members (Brown and Duguid 1991, Wenger 1998). Posing and answering questions as the work is performed is considered essential to fostering learning (Bechky 2003). Engagement in conversation or discussion threads has been documented as the primary way in which individuals share knowledge and learn from each other (Dahlander and Piezunka 2014, Hutter et al. 2011, Orlikowski and Yates 1994). We highlight another complementary pathway by which individuals learn from observing the completed work of others, without direct interaction and knowledge exchange (Nonaka 1994). This pathway allows individuals to learn in a competitive rather than a cooperative environment, in which they are exposed only to sparse and hard-to-interpret signals. This suggests that the rich interactions thought necessary for learning in online communities may only be one pathway of knowledge creation (Barrett et al. 2016). We provide evidence of individuals' ability to integrate the signals of both community and firm preferences into a consistent learning curve. As such, we contribute to a broader stream of community research that has looked at the interrelationship between firm and community goals (Dahl et al. 2014, Dahlander and Piezunka 2014, Hildebrand et al. 2013, Huang et al. 2014). We develop a richer conceptual discussion and explore the role of different preference signals in the same data.

Our data also include unique measures of the key construct of observational learning that is typically difficult to quantify. This allows for a richer and more nuanced analysis. Our results provide further evidence of the value of studying the effects of making large stocks of completed work available for inspection by an entire community. However, our results also highlight the need for a deeper understanding of transparent disclosure, beyond the more direct effect it has as source material or as inspiration to aid in the development of further innovations (Boudreau and Lakhani 2015, Stanko 2016, Terwiesch and Xu 2008, Yu and Nickerson 2011).

Second, a large body of research has examined learning as a key reason to join communities but has not studied learning that occurs within those communities (Lakhani and Wolf 2005, von Krogh et al. 2012). By exploring individuals' learning curves in competitive contest-based online communities, we add to an emerging stream of research that studies learning in these communities (Huang et al. 2014, Singh et al. 2011), as well as the processes by which individuals join those communities (Preece and Shneiderman 2009, von Krogh et al. 2003). Insights into how individuals rapidly get "up to speed" has the potential to make whole communities more productive. It may also help firms select which individuals to engage further. Theoretically, exploring learning in a competitive environment may help in developing new theories of community-based learning. By focusing on individuals' learning curves and the initial investments they need to make to see performance improvements, we also complement and extend work by von Krogh et al. (2003) and Preece and Shneiderman (2009). They focus on the role of different "joining scripts" of community members and their qualitative contributions as the drivers of sustainable online communities. Our results show that peripheral participation—learning by observing the work practices of others—is only one way to improve performance in a community and it is severely limited when individuals observe bad work.

2. Background

Learning is never observed directly, but is inferred from performance improvements (Bandura 1971). Thus, an increase in productivity or performance is depicted as a learning curve (Epple et al. 1996, Newell and Rosenbloom 1981, Rosen 1972). Improvement is informed by an individual's past success and failures and is not constant over time, demonstrating a learning curve effect with the rate of improvement varying depending on the information content of the signal from which one tries to learn. We theorize that individuals' learning curves in the context of contest-based online communities are influenced by the specific characteristics of signals from different sources of information.

In the following sections, we first theorize how individuals learn from coarse and conflicting evaluation signals they receive on the work they submit to contests (Section 2.1). We then conjecture that individuals' learning from observing the work of others may depend on whether the work they observe is of high or low quality, and may depend on their ability to correctly recognize the quality of that work (Section 2.2). Lastly, given that activities in online communities are self-directed and different actions generate different signals, we argue that individuals face a trade-off in allocating their effort. They can either submit their own work to increase their direct experience or observe completed work of others to increase their indirect experience (Section 2.3).

2.1. Learning from Evaluation Signals

Individuals learn from direct experience through the informative feedback signals they receive (Arrow 1962, Epple et al. 1996, Foster and Rosenzweig 1995, Newell and Rosenbloom 1981). Learning from direct experience is characterized by initial rapid advances in performance with diminishing returns. This results in the common "learning curve" that is steep initially but slows down as improvements get harder to make (Newell and Rosenbloom 1981). Learning from direct experience has been established as a general phenomenon that applies to human behavior, ranging from manual tasks that require motor skills to cognitive tasks like problem solving. These tasks are informed by an individual's feedback on relative successes and failures (e.g., Bandura 1977, Harlow 1949, Newell and Rosenbloom 1981, Thorndike 1898). We discuss three specific aspects related to learning from evaluation signals relevant in contest-based online communities and speculate how the characteristics of these signals may affect learning from direct experience.

2.1.1. Coarse Evaluative Feedback. The rate of learning from direct experience depends on informative feedback (Bandura 1971). Learning theory assumes implicitly that the feedback signal is informative, relevant to outcomes, and objective (Argote 2012, Newell and Rosenbloom 1981). However, the nature of evaluation feedback in contest-based online communities may not satisfy these assumptions. In contest-based online communities, the evaluative feedback that individuals receive is a simple numeric rating of the average community score, as well as the binary firm choice of contest winners. While both positive and negative feedback can contain valuable information about an individuals' submission, these evaluation signals may be hard to interpret. Consequently, individuals participating in innovation contests may have difficulty discovering connections between their submissions and the evaluations they receive, making learning very difficult (Bandura 1971).

Individuals learn from direct experience as it allows them to generate hypotheses about outcomes, test these hypotheses through action, and use informative feedback to guide future action (Bandura 1971). However, if the evaluation signal that an individual receives has low information content and is hard to interpret (such as the binary win–lose feedback received in contest-based online communities), it may be difficult to effectively evaluate hypotheses. Learning has been shown to be surprisingly imperfect or even completely blocked when signals are observed through coarse discrete filters such as a win–lose decision (Bikhchandani et al. 1998).

Work by Howell (2017) shows that entrepreneurs participating in new venture contests learn more about the quality of their ventures if they are informed of their relative rank in the contest, rather than if they are only informed if they won or lost. This form of learning from evaluative feedback contrasts sharply with user innovation communities in which individuals share personal insights, ask questions, and receive help (Brown and Duguid 1991, Dahlander and Magnusson 2005, von Krogh et al. 2003). Posing and answering questions as the work is performed has been shown to be essential to fostering learning (Bechky 2003). Engagement in conversation or discussion threads is the primary way in which individuals share knowledge and learn from each other (Orlikowski and Yates 1994, Yates et al. 2003). That is, while members in user innovation communities contribute to mutual learning through task-oriented communication, individuals in contest-based online communities are faced with the challenge of improving their performance in a competitive environment with only coarse performance feedback.

Given the absence of rich feedback and communication in contest-based online communities, individuals may struggle to learn from the numeric or binary evaluative feedback signals available. How successful individuals are in learning from these coarse evaluation signals will depend on their ability to process and extract meaning from them. We turn now to discuss how prior experience may help individuals develop cognitive structures to process evaluation signals and that when participants examine a mix of different evaluation signals, they may find conflicting information.

2.1.2. Cognitive Structures Required to Interpret Evaluation Signals. Cognitive structures are the mental processes that individuals use to process and understand information, and are central to learning (Gioia and Manz 1985). Learning from evaluation signals is most effective when an individual is familiar with a task and has developed—through prior experience—the necessary cognitive structures to effectively interpret such signals (Bandura 1977). Through cognitive structures formed by past experience, individuals develop more abstract representations of their knowledge. This enables them to pay more attention to relevant, structural features across diverse concepts (Glaser 1989, Newell and Simon 1972).

In contrast, individuals lacking appropriate cognitive structures tend to pay more attention to less relevant, superficial features. This often distracts them from attending to more meaningful connections across diverse concepts. As a result, an individual's prior experience and knowledge affects their ability to recognize the value of new information and assimilate it. For example, research on memory development suggests that accumulated prior knowledge increases the ability to assimilate and gain insights from new information such as evaluation signals (Bower and Hilgard 1981).

In the context of contest-based online communities, we would expect cognitive structures formed by experience to allow individuals to process evaluation feedback better and to synthesize more meaning from it. For example, by finding patterns between the specific characteristics of contest submissions and how they have been evaluated. Hence, an important contingency that determines the rate of learning from evaluation signals is the ability with which learnable information can be extracted from them. This ability critically depends on an individual's prior knowledge (Singh et al. 2011). An individual may have to make initial investments to develop cognitive structures before they are able to extract meaningful information and benefit from it (Sweller 1988). That is, individuals may need to participate in several contests before their performance improves because of the feedback they have received. If individuals are required to make initial investments to develop necessary cognitive structures, the learning curve may not be as steep initially as it is in contexts where evaluation signals are easy to interpret (Newell and Rosenbloom 1981). Insights into the shape of the learning curve (and whether it is flat in the beginning of individuals' tenure in contest-based online communities) has important implications. In particular, it might lead individuals to drop out from frustration at their lack of performance improvement.

2.1.3. Incorporating Evaluative Signals on Competing Performance Dimensions. Learning theory generally assumes that individuals are learning to improve on a single performance dimension (Hanna et al. 2014). Individuals in contest-based online communities are faced with the challenge that they need to integrate two sources of evaluation to do well: the firms' choices of contest winners (signaling firm preferences) and community ratings (signaling community preferences).

While community ratings may reflect the general taste of peers and consumers (Hildebrand et al. 2013), the firms' choice of contest winners is a signal for the ultimate goal of participation: winning. However, the firm's choice of a winning submission offers little information about all of the other submissions that are not chosen. That is, a positive success signal is very rare. The majority of individuals are simply informed about their failure to win the contest. On the other hand, community evaluations are available for all submissions. Hence, learning from the rare firm signals and integrating them with the more abundantly available community signals is critical.

The challenge of having to cater to two different sets of preferences (or production functions) is not unique to contest-based online communities. Both online and offline communities are often embedded in firms, requiring individuals to learn about the preferences of both the firm and the community (Lave and Wenger 1991, Wenger 1998). Firm-hosted open-source software and user innovation communities create benefits for both the community and the firm (Levina and Arriaga 2014, O'Mahony and Lakhani 2011). Consequently, individuals need to learn how their contributions are valued by these two stakeholders (Dahlander and Magnusson 2005, Dahlander and Piezunka 2014, Krogh and Hippel 2006). For example, work by Huang et al. (2014) shows that individuals in the Dell IdeaStorm community can learn about their own ability to generate ideas based on abundant community evaluations. However, they are slow to learn about the firm's cost structure based on Dell's rare decisions to implement an idea. Similarly, entrepreneurs in crowdfunding communities need to integrate and learn from the signals they receive from both crowd investors as well as professional investment firms (Catalini and Hui 2017)

A common feature in all of these cases is that the signals of the firm are more valuable but rare, while the community signals are more readily available but only proxies for those of the firm. If individuals only rely on the more valuable but rare signals of the firm, performance improvements may be slow (Huang et al. 2014). Furthermore, since the signals of firm are so coarse, they contain greater noise, which can completely shut down learning as individuals stop attending to them (Bikhchandani et al. 1998). While other studies have hinted at the existence of two opposing preferences, none have explicitly investigated if the learning rates of the two preferences differ and if (and how) both signals can be integrated.

In contest-based online communities, we would expect that individuals pay attention to both signals: those from the firm (because winning is the goal of participation) and those from the community (because these signals are available for all submitted work). While attention to the signal is a necessary precondition for learning (Bandura 1971), the learning from firm preferences may be hampered by the sheer rarity of the signal and its low information (i.e., binary win–lose) nature (Howell 2017). As a result, individuals may predominantly learn from community evaluations but fail to successfully integrate the signals of the firm. It is not clear if (and how) individuals could learn from sparse binary firm signals over and above what they learn from the more readily available and more informative community evaluation. Whether individuals can create learning curves that somehow aggregate these two sources of evaluation would provide generalizable insights that are applicable to other learning contexts where there are two preferences (production functions).

This discussion leads to the first set of guiding questions for our empirical analysis on how individuals learn from direct evaluation signals of their work: *What is the shape of the learning curve of direct experience given that individuals receive only coarse and possibly conflicting evaluation signals about their performance? Are individuals able to aggregate the signals of community and firm preferences into a coherent learning curve (i.e., integrate the evaluation signals of two response functions)?*

2.2. Learning from Observing the Work of Others

Observational learning (also known as imitation or modeling) is based on the premise that individuals can improve their own work by observing the work of others (Bandura 1977). Specifically, by studying "model" examples submitted to contests by others, individuals might learn how to create winning submissions themselves. The capacity to learn by observation enables individuals to acquire large, integrated patterns of behavior without having to form them gradually by tedious trial and error (Bandura 1977). In fact, many complex behaviors can only be learned indirectly through observation as there are simply too many behavior patterns to try to learn exclusively from direct experience. Learning through observation occurs principally through its informative function. Observers acquire symbolic representations of the observed model, rather than simply mimicking a model (Bandura 1971).

Observational learning is facilitated when the observer pays attention to the model, retains the information, and has the capacity to enact the model's behaviors (Bandura and Jeffrey 1973). In an innovation context, observing the work of others enables individuals to better understand the space of possible solutions and thus revise their own experimentation (Thomke 1998). That is, observing the work of others gives individuals an opportunity to learn what to do, not just how to do things. Attentional processes determine what is selectively observed and what information is extracted in the profusion of modeling influences (Andrieux and Proteau 2016, Hanna et al. 2014). Simple exposure to models does not in itself ensure learning (Bandura 1971). This suggests that an individual's rate of learning may not be constant over time. Instead, it may in fact evolve with cognitive structures that guide what information is extracted from the observed models.

2.2.1. Need to Deconstruct Completed Solutions. Social learning theory generally assumes that the model is observed while performing a certain action, together with the finished product, and knowledge of the model's performance (Bandura 1971). In contest-based online communities, the unit of disclosure is the complete submission (Boudreau and Lakhani 2015). Individuals need to deconstruct the *opus operatum* (i.e., what is visible after the task is finished) to select and extract the model's most relevant characteristics (Bourdieu 1990, Brown and Duguid 1991). Furthermore, the absolute level of quality in many innovation contexts is not known until future market success is realized.

Deconstructing complex completed work such as submissions to innovation contests is extremely challenging and contrasts with the positive effects on learning of observing someone perform a task (Nadler et al. 2003, Nonaka 1994). This is one of the foundational ideas behind communities of practice (Brown and Duguid 1991). Specifically, individuals need to decide what attributes of a piece of work to pay attention to and then accurately perceive those characteristics they happen to notice (Van Knippenberg et al. 2015). So, while the vast body of openly disclosed completed work provides rich opportunities for observational learning, individuals may find it difficult to learn from these completed works because they need to be deconstructed before meaningful information can be gleaned from them. As a result, individuals may need to form appropriate cognitive structures to select and correctly perceive relevant characteristics of the completed work they observe. Given the complexity of many interconnected problem dimensions found in many innovation settings (Kornish and Ulrich 2011), this deconstruction process may be particularly difficult. As a result, it is not clear if simply disclosing large stocks of complete submissions that need to be deconstructed will lead to learning from observation.

2.2.2. Learning from Good and Bad Examples. Individuals cannot learn much by observation if they do not attend to (or recognize) the essential features of the model to which they are exposed (Bandura 1971). Simple exposure to a model does not in itself ensure that an individual will select the most relevant characteristics from the model or even that they will perceive them accurately (Andrieux and Proteau 2016, Bandura 1971). So, while individuals can generally learn from observing work produced by others, which types of work should they observe to improve the quality of their own work? For example, observing the work of experts ("good examples") will help them to develop a reference of what to do and how to do it correctly. Or should they also observe the work of novices ("bad examples") to give them a better chance to detect and learn from errors?

Research has shown that significant learning can occur from observing the work of both experts who have mastered a skill (Al-Abood et al. 2001, Heyes and Foster 2002) and novices who commit errors (Buchanan and Dean 2010). Observing errors facilitates learning by helping to build cognitive structures for error detection and correction strategies (Jones and Endsley 2000). Errors have been shown to contain particularly valuable learnable signals that are not present in observing the work of experts (Andrieux and Proteau 2016). Specifically, observing errors can foster the development of mechanisms to detect deviations from ideal behaviors (Blandin and Proteau 2000). However, learning from errors is hard if the cause of the error is unknown or difficult to discern (Haunschild and Sullivan 2002). Determining the cause of an error is particularly difficult if performance is determined by the interplay of many attributes on different dimensions (Haunschild and Sullivan 2002, Kornish and Ulrich 2011). This may make learning from errors particularly difficult in an innovation context that has many complex, interacting dimensions.

Individuals learn from observing errors by developing efficient processes for error detection and correction (Carroll and Bandura 1990). To effectively learn from errors, the observer needs to be informed of the model's performance (Andrieux and Proteau 2016, Blandin and Proteau 2000). In fact, the knowledge of results provided to observers about the model's performance is considered to be the most important variable affecting observational learning (Badets and Blandin 2004). However, in many online community settings, only the work itself is observed without knowledge of its actual performance. For example, a programmer in an open-source community may observe the code written by another without explicit knowledge of the performance of the code (such as whether it contains any errors or how much a specific section of the code contributes to its overall performance). That is, while there is an absolute level of quality (or success/failure) associated with a submission, individuals may vary in their ability to correctly recognize this absolute level of quality (i.e., their perception of the true quality may be erroneous). As a result, signals from the work of others are observed with noise, and how much individuals can learn from such signals will depend on the level of noise (Bikhchandani et al. 1998).

Consequently, individuals are faced with the challenge that they first need to determine if they are exposed to a good example that informs ideal behavior, or if they are exposed to a bad example from which they may learn error detection and correction strategies. Since the absolute level of quality is unknown to the individual at the moment they observe the work, individuals may develop an erroneous reference of what is good technique, and their performance will suffer (Blandin and Proteau 2000). That is, exposure to bad examples and subsequent indiscriminate imitation may even have the potential to *reduce* performance, rather than increase it. Whether learning from errors is effective in contest-based online communities may critically depend on an individual's ability to detect and discern causes of the error in bad examples, and then to glean error detection and correction strategies from them (rather than imitating them). That is, better detection and quantification of the model's performance may favor learning from bad examples (Andrieux and Proteau 2016). However, observing many good examples may help individuals develop the necessary cognitive structures to better detect errors in bad examples and allow them to subsequently learn effective error detection strategies from them.

In summary, while it is likely that individuals learn from exposure to both good and bad examples, it is an empirical question how the learning rates in this context differ for learning from signals of good and bad examples. The learning rates may depend on an individual's ability to correctly perceive the quality of a submission. This discussion leads to the second set of guiding questions for our empirical analysis on how individuals learn from observing the work of others: *What is the shape of the learning curve of indirect experience given that individuals need to deconstruct complete submission instead of being able to observe others perform a task? How do the learning rates differ for learning from signals of good and bad examples and how do the learning rates depend on individuals' ability to correctly perceive the quality of examples?*

2.3. Facing Effort Allocation: Trading-off Direct and Indirect Experience

Both direct and indirect experience have been shown to enhance performance independently. However, there is also strong evidence for a positive effect of combining both modes of learning (Bandura and Jeffrey 1973, Blandin et al. 1999, Darr et al. 1995, Levitt and March 1988). Despite this insight, it is unclear how strong the complementarity is (or if it exists at all) in a context in which observed examples cannot be modeled directly because derivative work is frowned upon (Bauer et al. 2016). Synergies between direct and indirect experience reported in the learning literature (Bandura 1971, Blandin and Proteau 2000) may be nonexistent (or much weaker) in the context of contest-based online communities. Individuals in these communities may not model behavior immediately after observing it, nor may they model what they observe directly. Instead, individuals are required to form cognitive structures and form mental representations to store the modeling stimulus for later and modified use. That is, individuals are required to form a generative model rather than a one-to-one mapping between cognitive representation and action that is enhanced through the combination of observation and direct experience (Bandura 1971).

Given that it is unclear how strong a possible synergy between practice and observation is in contest-based online communities, how should individuals allocate their effort between entering the contests themselves and observing the work of others? In online communities, effort is typically self-guided, and members face a trade-off in allocating their effort because different actions will generate different learning signals. Understanding how experience from direct and indirect signals interacts is critical to guiding this trade-off. However, individuals cannot gain more direct experience and observe the work of others simultaneously. They therefore must decide how to allocate their time between these activities. The decision is important because they may learn more effectively if they allocate their effort in a way that generates the most efficient set of learning signals. While it is likely that there is a synergy between direct and indirect experience, it is an empirical question of how individuals should best allocate their effort between the two.

This discussion leads to the third set of guiding questions for our empirical analysis on facing effort allocation: *What is the shape of the learning curve resulting from combining direct and indirect experience in contest-based communities? How should individuals allocate their effort between submitting their own work and observing the work of others?*

Table 1 summarizes key aspects of our discussion and anticipates the structure of our empirical analysis. To summarize, we explore the relationships between the signals available to individuals in contest-based

Table 1. Summary of Signals Available to Guide Performance Improvement and Theoretical Learning Perspectives

Signal observable by individual	Characteristics of signal	Theoretical perspective	Context-specific complication
Firm choice (win/lose) on own submission	Valuable: winning is goal Infrequent: small proportion chosen Coarse: binary win/lose	Response learning from direct experience (e.g., Newell and Rosenbloom 1981) Signal informativeness affects learning rate (e.g., Bandura 1971) (Section 2.1: *Learning from Evaluation Signals*)	Low information content; not developmental Cognitive structures necessary to interpret signal Integrate signals of firm vs. community preferences
Community rating of own submission	Less valuable: proxy of firm signal Frequent: available for all own submissions More granular: numeric average rating		
Completed submissions of others	Value uncertain: no quality judgment available Frequent: many submissions available Complex: many characteristics	Observational learning from indirect experience (e.g., Bandura 1977) Error detection and correction strategies (e.g., Carroll and Bandura 1990) Exposure to expert and novice examples (e.g., Blandin and Proteau 2000) (Section 2.2: *Learning from Observing the Work of Others*) Integration of observational learning and direct experience (e.g., Bandura and Jeffrey 1973) (Section 2.3: *Facing Effort Allocation*)	Need to deconstruct completed work to extract relevant characteristics Not clear what errors are and what caused them Cognitive structures necessary to correctly perceive if example is good or bad Trade-off allocation of effort between direct and indirect experience

online communities and their performance improvement. In doing so, we consider (a) evaluation signals from the community and the firm on the work submitted, and (b) signals from observing work completed by others.

3. Methods and Measures

3.1. Research Setting

Data for our study comes from a longitudinal panel of t-shirt designs submitted to the Threadless website (https://www.threadless.com), a platform for t-shirt design contests and an e-commerce site. Our data contain all design submissions made by more than 55,000 individuals during a 10-year period from 2001 through 2011, consisting of almost 180,000 submissions and nearly 150 million evaluations of these designs. Since the notion of performance improvement that we can interpret as learning is only meaningful for individuals who make at least two design submissions, our panel regression includes the 33,813 individuals who made two or more design submissions. Obvious spam, blank, or duplicate submissions were filtered out by Threadless, are not put up for voting on the website, and are not part of our data set. Our data set averages about 340 design submissions per week (i.e., per contest).

Every week, submitted designs were posted on the website for a peer assessment rating of other community members (fellow community members as well as interested consumers), using a scale from zero to five. Ratings were anonymous, and no average score was shown at the time of rating—only a counter of how many ratings had been submitted is shown. This community rating serves as a sign of market potential (Huang et al. 2014) and is our dependent variable to assess performance improvement, which we interpret as a learning curve. It is important to note that, like in many innovation contexts, the goal of the community rating is to predict the absolute quality of the contest submissions (i.e., new product ideas), which would only be revealed at some point in the future through other measures such as sales revenue (Blohm et al. 2016).

Among the submissions that received the highest ratings, Threadless would typically pick three to six designs per week to be printed and subsequently sold through the Threadless e-commerce site. Decisions on which t-shirts were printed rested with the management team at Threadless. The management team would consider not only community evaluation of designs but also how to balance a portfolio of t-shirt themes, and other strategic factors—thus effectively curating a “house style” idiosyncratic to Threadless. Winning designers would receive a cash prize and store credit if their design were selected for printing. In effect, the community ratings represented a *judgment-based* evaluation of design performance, based on a continuous variable of score from zero to five; the firm further used a *choice-based* evaluation of design performance, choosing specific designs as contest winners that were not necessarily the absolute top-rated score.

3.2. Dependent and Independent Variables

3.2.1. Dependent Variable. Our primary dependent variable $Score_{it}$ measures the performance of an individual i at time t. Performance is assessed on a scale from zero to five with higher scores representing better submissions. $Score_{it}$ is the mean of all community ratings a submission received. The online appendix provides several robustness tests to support the claim that *Score* captures actual performance rather than an individual's popularity within the community. We do this by demonstrating that submissions that were ultimately selected by the firm to be printed ranked in the highest percentile among competing submissions. Second, we show that the community rating is a robust measure for market potential. In the context of our empirical model that builds on a standard model to describe change in performance (see next section), the dependent variable $Score_{it}$ is then interpreted as "rate of learning," which specifies a learning curve. In summary, in our empirical setting, the "learning curve" is reflected in a rate of increasing score over time.

3.2.2. Direct and Indirect Experience. We measure prior direct experience as the cumulative number of submissions made by individual i prior to making the current submission (Exp_{it-1}). This prior direct experience contains any submission made, independent of how well or poorly it fared in a contest. Indirect experience is measured by the cumulative number of all prior ratings cast by individual i before submitting the current submission ($Votes_{it-1}$). The number of ratings an individual has cast serves as a proxy variable for the knowledge acquired indirectly by observing and evaluating others' submissions. To test the differential effect of exposure to either a submission of high performance or a submission of low performance, we create separate measures that count the respective numbers of ratings cast for high- and low-performance submissions ($Votes_{it-1\,(Good)}$ and $Votes_{it-1\,(Bad)}$). We consider a submission to be of high performance if—ex post considering all ratings that a submission received—the community consensus (average score) of that submission was higher than the mean score of all submissions (i.e., $Score > 1.91$) and of low performance otherwise.[3] The two separate counts added together yield the total indirect experience ($Votes_{it-1(Good)} + Votes_{it-1(Bad)} = Votes_{it-1}$). It is important to note that our separate measures of prior indirect experience of evaluating good and bad examples of others do not depend on the valence of the rating made by the individual, but rather on the aggregate (ex post) score of the submission being rated.

3.2.3. Error in Quality Perception. To evaluate if an individual's ability to learn from exposure to bad examples depends on their ability to correctly detect them, we compute a measure of perception error. That is, we compute the root-mean-squared rating error for all past ratings cast by an individual ($RatingError_{it-1}$). The rating error is the difference between individual i's rating of a submission and the average (ex post) community rating for that submission. Thus, the deviation between i's own rating and the community consensus gives us a measure for i's error in correctly perceiving the quality of a submission; that is, i's cognitive structures for correctly perceiving the work of others. Smaller values indicate more accurate quality perception (lower error), while larger values indicate more erroneous quality perception (higher error).

3.2.4. Control Variables. Our dependent variable *Score* derives from individual consumer ratings, and the number of scores submitted varies across submissions. To account for the fact that submissions with fewer ratings may provide a less robust measure of performance and may be systematically higher as they are more easily influenced, we control for the count of ratings a submission received ($ScoreCount_{it}$).

We include a control variable $Tenure_{it}$, which captures the number of years an individual i has been active on the Threadless platform at time t to distinguish between learning from participation in our setting and changes in performance scores that may occur because of other factors (Argote and Epple 1990, Thornton and Thompson 2001). $Tenure_{it}$ effectively also captures the general passage of time and thus incorporates possible experience accumulated outside the organization through training, education, or other activities outside the online community. It is also important to control for other variables that could impact performance ratings, such as the size of prize money offered in a contest (which may affect individuals' effort), the level of contest in a given week, and general changes in trend. We control for both observed and unobserved contest characteristics through contest fixed effects.

3.3. Empirical Model

Following common approaches to study learning (e.g., Argote and Epple 1990, Wiersma 2007), we estimate an empirical model to explore performance improvements (i.e., learning rate) from direct and indirect experience from feedback on relative success and failure as given in Equation (1):

$$\ln Score_{it} = \beta_1 \ln Exp_{it-1} + \beta_2 \ln Votes_{it-1} + \alpha_i + \epsilon_{it}. \quad (1)$$

$Score_{it}$ is the performance of the t-th submission of individual i. Exp_{it-1} is the cumulative measure of all prior direct experience, and $Votes_{it-1}$ is the indirect experience. β_1 in Equation 1 is the learning rate from direct experience, while β_2 is the learning rate from indirect experience. α_i are individual-level random effects and ϵ_{it} are error terms. To test for complementarities and after accounting for control variables, contest effects, and individual effects, we estimate the following full model:

$$\begin{aligned}\ln Score_{it} = {} & \beta_1 \ln Exp_{it-1} + \beta_2 \ln Votes_{it-1} \\ & + \beta_3 \ln Exp_{it-1} \times \ln Votes_{it-1} + \beta_{4-5} Controls_{it} \\ & + \beta_{6-610} Contest_t + \alpha_i + \epsilon_{it}.\end{aligned} \tag{2}$$

Learning from direct experience is supported if β_1 is positive and significant, and learning from indirect experience is supported if β_2 is positive and significant. A complementary effect of direct and indirect experience is supported if the coefficient β_3 of the interaction term is positive. To test if indirect experience from evaluating high- or low-performance submissions is more strongly associated with performance improvements, we compare the coefficients $\beta_{2(Good)}$ and $\beta_{2(Bad)}$, which we test in a separate model. If $\beta_{2(Good)}$ is larger than $\beta_{2(Bad)}$, this would indicate that indirect experience from evaluating high-performance submissions leads to higher performance improvement than indirect experience from evaluating low-performance submissions.

Our data have a crossed multilevel structure: submissions are nested within individuals, and individuals submit to one or more contests. We specify contests as fixed effects and individuals as random effects. Fixed effects are appropriate if the levels in the study represent all possible levels of the factor about which inferences are made, which is the case in our data, as it includes all contests that took place during the 10-year period. A fixed-effects formulation assumes that estimating different intercept terms for each unit can best capture differences across units. This approach is appropriate if units differ in their average level of the outcome measure, which is likely in our case as contests vary in popularity, difficulty, and prize money, which probably affects the value of a submission. A random-effects specification, on the other hand, assumes that levels of the factor used in the study represent a random sample of a larger set of potential levels (Greene 2003). Since we wish to generalize the learning among individuals in our sample to other individuals, we define the individual level as random.

We address three main concerns with regard to the identification of causal learning effects: (1) dropout (panel attrition) may be correlated with the dependent variable (performance); (2) decision to rate good or bad submissions is endogenous and may be correlated with the dependent variable (performance); and (3) possible concerns regarding the dependent variable, in particular whether community scores reflect submission quality, and whether community scores of submission quality are affected by positive reciprocity. To address the first point, we apply a variable addition test for attrition bias (Verbeek and Nijman 1992) and inverse probability weighting to adjust for dropout in estimation of our models (Fitzgerald et al. 1998, Moffitt et al. 1999, Robins et al. 1995, Wooldridge 2002). We include the analysis in the online appendix and provide additional details on the estimation procedure. Second, high-performing individuals may not bother to evaluate bad ideas, which could explain differences in effects from learning from good examples versus bad examples. The multilevel specification outlined above accounts for unobserved individual-specific characteristics that are fixed over time. This includes latent levels of talent, taste (for certain submissions), and raw intelligence. As a result, learning effects are identified by observing within-individual changes in behavior over time. To address point three, we performed additional analyses to address concerns that the community rating is driven by reciprocity or otherwise manipulated for which we find no support (see the online appendix).

3.4. Estimation

Given the crossed panel nature of our data, we adopt a multilevel estimation approach (Gelman and Hill 2006) that allows us to model fixed and random effects, accommodate the nesting of repeated measures within individuals and contests, and account for autocorrelation and unequal spacing of observations in time. We estimate the model using linear mixed model fit by maximum likelihood. We used the lme4 (Bates et al. 2015) and nlme (Pinheiro et al. 2015) packages available in R (R Core Team 2015) and full maximum likelihood that yields log-likelihood numbers that can be used to evaluate the incremental fit across nested models. We obtain similar results when using restricted maximum likelihood instead.

Multilevel modeling is the recommended approach for hierarchical data with multiple nested and crossed levels (Gelman and Hill 2006). Furthermore, it affords considerable flexibility to deal with varying numbers of observations per individual and contest. We conduct several specification checks for the models, including specifying the contest level as random instead of fixed, and the individual level as fixed instead of random. Results are consistent with our original specifications, although our original specifications yield the best model fit. To control for serial correlation among submissions made close together in terms of time, we estimate a model including an AR(1) covariance structure as an additional robustness test. Since submissions are not equally spaced in time, we use a generalized AR(1) structure that allows unequal spacing between observations. Such a spatial power function models the covariance between two observations at t_1 and t_2 as

$\text{cov}(y_{t1}, y_{t2}) = \sigma^2 \rho^{|t1-t2|}$, where ρ is an autoregressive parameter assumed to satisfy, and $\sigma^2\rho$ is the overall variance.[4]

We apply variable addition tests for attrition bias (Verbeek and Nijman 1992) and inverse probability weighting to adjust for dropout in estimation of pooled models (Fitzgerald et al. 1998, Moffitt et al. 1999, Robins et al. 1995, Wooldridge 2002). The online appendix provides additional details on the estimation procedure.

4. Results

4.1. Descriptive Statistics

Our unbalanced, crossed panel data set comprises 33,813 individuals with varying panel length ranging from two to 359 submissions for a total of 136,989.[5] The panel covers a period of 10 years from 2001 to 2011. Prior experience is the cumulative sum of all submissions made by an individual. The dependent variable *Score* is the aggregate of a total of 149 million ratings with an average of 851 ratings per submission (SD = 657.5). Rating is anonymous, and each individual is allowed only one rating per submission. Table 2 shows descriptive statistics and correlations among key variables. The average tenure is 10 months (0.83 years) and a maximum of almost nine years. Online Appendix Figure S1 shows the distribution of prior direct and indirect experience.

Next, we investigate if individuals' rating behavior is stable over time or whether there is any discernable behavior change. Figure 1 shows individuals' number of submissions against their number of ratings on a log-log scale. The plot shows how many ratings individuals submit on average between making submissions over the full spectrum of individual careers. If the pattern in the data changes, that would imply that the rating pattern between submissions changes. A linear pattern would imply that the number of ratings that an individual provides between submissions remains stable. From the plot, we see that early in their careers (up to the fifth submission), individuals cast on average about 100 ratings between each submission. Then, their voting increases to about 135 ratings between each submission on average. This pattern remains stable between the fifth and the 50th submission. Among the very experienced individuals—those making their 50th or later submission—the number of ratings cast between submissions decreases slightly to about 127. Generally, at this high level, the pattern becomes less reliable and noisier as it is based on fewer numbers of observations. In summary, we conclude that the pattern between submissions and voting does not change significantly over the career of individuals and stays relatively stable at around 114 ratings between submissions across the whole data set.

4.2. Learning from Evaluation Signals and from Observing the Work of Others

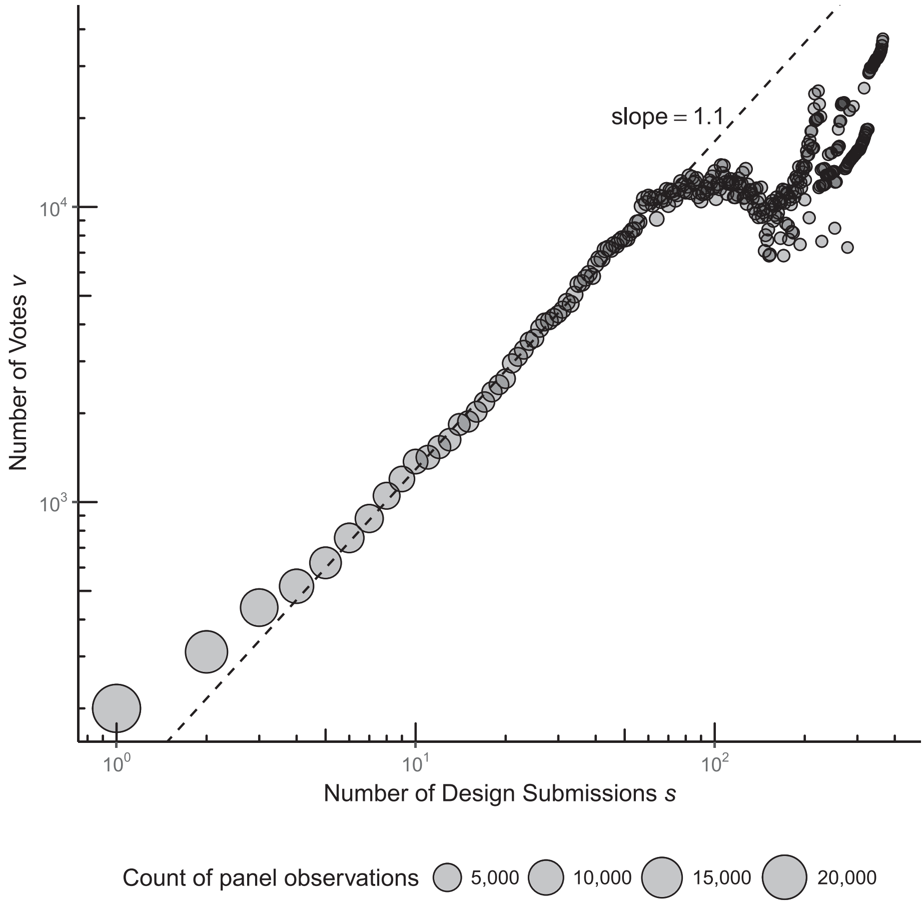
Table 3 presents the estimation results for learning from evaluation feedback (direct experience) and observing the work of others (indirect experience). We first establish that there is significant variation in submission scores, reflecting a wide range of relative success and failure, by estimating only an intercept term (Model 1). There is more variation in scores between submissions than there is variation between individuals. Model 2 adds controls and the direct and indirect experience variables, Model 3 adds squared terms of the experience variables, Model 4 adds an interaction term between direct and indirect experience, Model 5 replaces measure of overall indirect experience with separate measures for indirect experience from evaluating high- and low-performance submissions, Model 6 adds an interaction term between the two measures of indirect experience, and Model 7 adds the measure of rating error. Model 8 (shown in the online appendix) allows for autocorrelation to control for unequal spacing between submissions (i.e., it uses an AR(1) correlation matrix), and Model 9 (shown in the online appendix) uses inverse probability weighting to control for panel attrition.The results indicate that an individual's current performance increases with increasing prior direct experience ($\beta = 0.004$; $p < 0.001$) and indirect experience ($\beta = 0.010$; $p < 0.001$). We provide additional analyses to better interpret the effect size in a simulation study below. To test whether direct or indirect experience is more effective, we use a Wald chi-square test to compare the two regression coefficients. We find that the coefficient for indirect experience is significantly larger than the coefficient for direct experience

Table 2. Descriptive Statistics

	Mean	Standard deviation	Minimum	Maximum	(1)	(2)	(3)	(4)	(5)
$Score_{it}$ (1)	1.91	0.51	0.25	5.00					
$ScoreCount_{it}$ (2)	850.61	657.46	1.00	3,791.00	0.30				
Exp_{it-1} (3)	16.48	30.84	1.00	365.00	0.31	0.06			
$Votes_{it-1}$ (4)	2,080.83	6,498.96	0.00	128,864.00	0.24	0.05	0.44		
$RatingError_{it-1}$ (5)	1.63	0.55	0.00	4.41	−0.16	−0.05	−0.05	−0.20	
$Tenure_{it}$ (6)	0.83	1.13	0.00	8.84	0.30	0.02	0.44	0.37	−0.08

Note. All pairwise correlations are statistically significant at $p < 0.001$.

Figure 1. Ratio of Submissions to Voting on Log-Log Scale Indicates That Individuals Submit About 110 Evaluations for Every Submission



($\chi^2 = 39.82$; $p < 0.001$). Given the logged values in our model, the coefficients can be interpreted as percentage changes: doubling an individual's indirect experience from voting results in a larger expected increase in that individual's current performance than doubling that individual's direct experience. However, it is important to bear in mind that submissions and cast ratings operate on different orders of magnitude with roughly a 1:100 ratio (see Figure 1). For example, an average individual with a prior direct experience of five submissions will usually have cast about 500 ratings. To double their direct experience, an individual would need to submit another five submissions and cast another 500 ratings to double their indirect experience.

We find statistically significant squared terms of both direct and indirect experience, suggesting that the learning curve for both types of experience is nonlinear (we investigate the exact shape of the learning curves in more detail below). To test whether direct experience and indirect experience have a complementary effect on improving an individual's current performance, we include an interaction term between the two. We find a positive and statistically significant effect ($\beta = 0.003$; $p < 0.001$) suggesting a complementarity between direct and indirect experience.

4.2.1. Learning from Good and Bad Examples. With regard to learning from the indirect experience from observing others good and bad examples, we find a statistically significant and positive coefficient for indirect experience from evaluating high-performance submissions ($\beta = 0.017$; $p < 0.001$), and a statistically significant but negative coefficient for indirect experience from evaluating low-performance submissions ($\beta = -0.007$; $p < 0.001$). This indicates that while individuals improve their performance when evaluating good examples, their performance decreases when evaluating bad examples. Next, we evaluate the complementarity between evaluating high- and low-performing submissions by introducing an interaction term of the two measures. We see a statistically significant and positive coefficient ($\beta = 0.002$; $p < 0.001$), indicating that prior experience in evaluating others' good examples can help individuals overcome their inability to improve their performance when evaluating others' bad examples. Next, we investigate if an individual's ability to improve her performance would

Table 3. Effects of Prior Direct and Indirect Experience on Performance (Dependent Variable: ln*Score*)

	ln*Score*						
Dependent variable	(1) Baseline	(2) Experience	(3) Squared terms	(4) Direct–indirect interaction	(5) Good vs. bad	(6) Good–bad interaction	(7) Rating error
Intercept	0.868***	1.119***	1.112***	1.154***	1.120***	1.153***	1.108***
	(0.163)	(0.133)	(0.133)	(0.133)	(0.133)	(0.133)	(0.083)
ln*Exp*		0.004***	−0.013***	0.002**	0.004***	0.001*	0.006***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln*Votes*		0.010***	0.005***	0.013***			
		(0.000)	(0.001)	(0.000)			
$\ln Exp^2$			0.005***				
			(0.000)				
$\ln Votes^2$			0.001***				
			(0.000)				
ln*Exp* × ln*Votes*				0.003***			
				(0.000)			
$\ln Votes_{Good}$					0.017***	0.024***	0.024***
					(0.001)	(0.001)	(0.001)
$\ln Votes_{Bad}$					−0.007***	−0.010***	−0.014***
					(0.001)	(0.001)	(0.001)
$\ln Votes_{Good} \times \ln Votes_{Bad}$						0.002***	
						(0.000)	
RatingError							−0.024***
							(0.002)
RatingError × $\ln Votes_{Good}$							0.001
							(0.001)
RatingError × $\ln Votes_{Bad}$							−0.010***
							(0.001)
Tenure	No	Yes	Yes	Yes	Yes	Yes	Yes
ScoreCount	No	Yes	Yes	Yes	Yes	Yes	Yes
Contest	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed
Individual	Random	Random	Random	Random	Random	Random	Random
N	136,989	136,989	136,989	136,989	136,989	136,989	122,070
N groups: individual	33,813	33,813	33,813	33,813	33,813	33,813	28,382
log likelihood	13,137.699	40,248.207	40,435.414	40,416.161	40,320.697	40,449.323	37,295.627
Variance							
Individual (intercept)	0.027	0.013	0.013	0.013	0.013	0.013	0.012
Residual	0.038	0.026	0.026	0.026	0.026	0.026	0.026

$^*p < 0.05$; $^{***}p < 0.001$.

depend on her cognitive structures that allow her to correctly recognize bad examples. To investigate this role of cognitive structures, we include the measure of rating error and interaction terms with indirect experience from evaluating high- and low-performing work of others. We find a statistically significant negative main effect for the rating error ($\beta = -0.024$; $p < 0.001$), indicating that the more erroneous an individual's perception of the quality of submissions, the lower their performance improvement. We find no significant interaction between rating error and indirect experience from evaluating good examples ($\beta = 0.001$; *ns*), indicating that performance improvements from exposure to good templates is independent of whether those examples are perceived correctly. However, we find a statistically significant negative interaction between rating error and indirect experience from evaluating bad examples ($\beta = -0.010$; $p < 0.001$), indicating that an individual's performance decrease from evaluating bad examples is amplified by their inability to correctly recognize that these examples are indeed bad. That is, the worse an individual is at correctly judging the quality of submissions (the higher their rating error), the lower their performance improvement from observing bad examples. Together, this indicates that the cognitive structures that allow individuals to correctly perceive the quality of work produced by others matter for performance improvements from exposure to bad examples but not for performance improvements from exposure to good examples. Put another way, individuals consistently improve their performance from observing others good examples independently of correctly recognizing the quality of these examples, but whether they improve from observing others bad examples depends on their ability to correctly recognize that work as being of low quality.

4.2.2. Incorporating Evaluative Signals on Competing Performance Dimensions. We now examine if individuals can aggregate signals of community and firm preferences into a coherent learning curve. We perform time-series–cross-section analyses using a binary dependent variable indicating whether a submission has been selected for printing in a given week. Specifically, we estimate the probability of winning a contest conditional on an individual i's time-varying measures x_i, which in our case are measures of direct and indirect experience (see the online appendix for details on the estimation procedure). This set of analyses allows us to address the question of what role experience plays in winning a contest as opposed to receiving higher scores from community evaluations. Repeating the main analyses using a different measure of success provides additional support for our main findings. This analysis uses two separate evaluations. The evaluation score used in the main analysis is a signal of community preferences, and the second measure we use in this analysis is a signal of firm preferences by the winner it chooses from among the set of submissions made to the same contest.[6]

To test whether individuals learn from their prior direct and indirect experience about firm preferences *above and beyond* what they learned about community preferences, we estimate a model in which we include a time-varying measure of the highest submission score based on the community judgment that an individual has been able to achieve in the past (we simply take the maximum score that an individual received up to that point). The intuition behind the analysis is as follows: if the coefficient for either direct or indirect experience is still significant when controlling for the community-based score, then the individual must have learned something from this experience about Threadless's decision-making process of selecting contest winners.

Table 4 shows the regression results for the time-series–cross-section analysis. Model 1 includes our two key measures for direct and indirect experience along with an intercept. We begin by interpreting the intercept and the baseline hazard rate of winning. As expected, the intercept is negative and quite large ($\beta = -8.823$; $p < 0.001$), indicating a low overall likelihood of winning (Online Appendix Figure S3 shows the baseline hazard rate of winning a contest as a function of time). Next, we investigate the role of direct and indirect experience in winning a contest. The coefficients for direct and indirect experience are both positive and statistically significant (both $p < 0.001$),

Table 4. Effects of Prior Direct and Indirect Experience on Winning (Dependent Variable: Pr(*Won*=1, *t*))

	Pr(*Won=1, t*)	
Dependent variable	(1) Experience	(2) Past performance
Intercept	−8.823*** (0.072)	−15.403*** (0.154)
Exp	0.039*** (0.001)	0.022*** (0.001)
Votes	0.000*** (0.000)	0.000*** (0.000)
Max score		3.360*** (0.049)
t	−0.019*** (0.003)	−0.036*** (0.003)
t^2	−0.000 (0.000)	0.000** (0.000)
t^3	0.000 (0.000)	−0.000 (0.000)
log likelihood	−9,142.513	−6,819.416
Deviance	18,285.026	13,638.833
N	21,799,462	21,799,462

$^{**}p < 0.01$; $^{***}p < 0.001$.

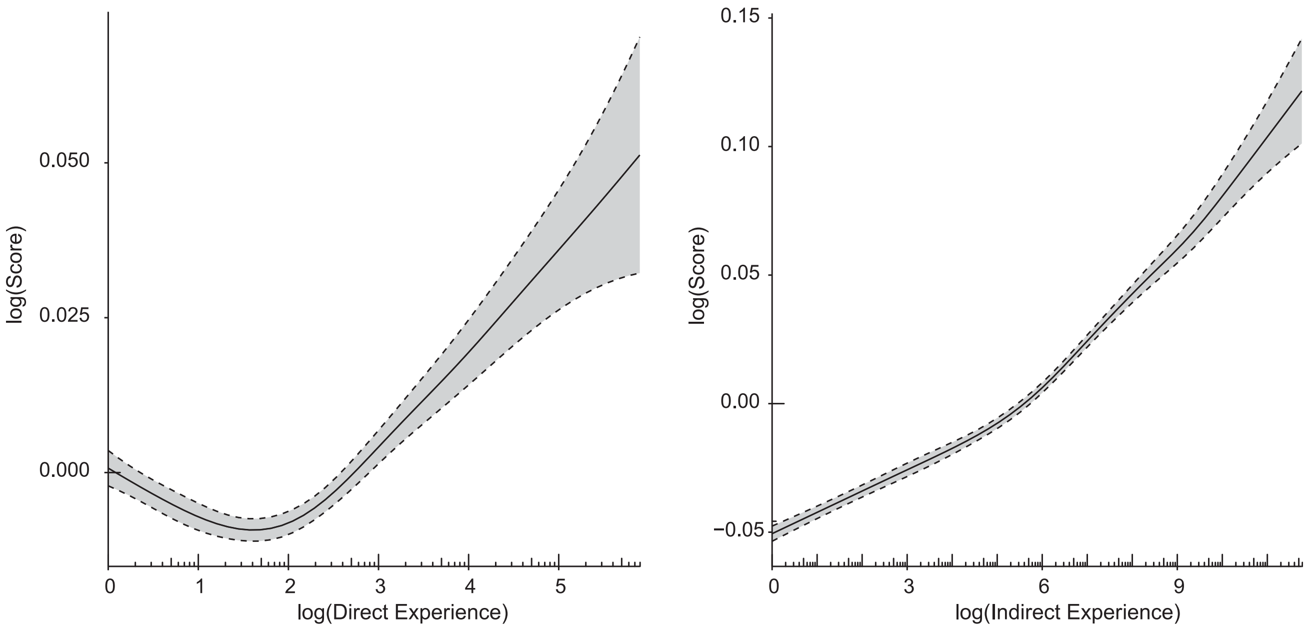
indicating that experience improves the odds of winning a contest over the (low) baseline hazard rate.

Model 2 introduces the additional covariate capturing the community score of the best submission an individual has been able to produce so far. The coefficient is quite large and statistically significant ($\beta = -3.36$; $p < 0.001$), indicating that having demonstrated the ability to produce high-scoring submissions in the past significantly increases the chance of winning a contest. More important for our analysis is, however, that the coefficients for direct and indirect experience both remain statistically significant, albeit decreasing in size. This indicates that—controlling for past community scores—individuals still benefit from both direct and indirect experience to increase their chance of winning a contest. That is, individuals can aggregate signals of both firm and community preferences into a coherent learning curve.

4.2.3. Shape of the Learning Curve. Model 3 in Table 3 includes squared terms for both direct and indirect experience. The coefficients for both squared terms are statistically significant, thus establishing that the relationship is nonlinear. But what exactly is the shape of the learning curve and is there evidence of initially low learning rates as individuals develop cognitive structures that help them interpret the signals they receive? Leveraging our extremely rich data set of more than 130,000 submissions and long panels, we directly investigate the shape of learning curves using generalized additive models (GAM) that we fit via maximum-likelihood (James et al. 2013).

The analysis confirms that the nonlinear smooth terms are a significantly better fit for the data than a linear model (ln*Exp* has 4.56 degrees of freedom, $p < 0.001$; ln*Votes* has 4.57 degrees of freedom, $p < 0.001$). The nonparametric GAM model explains 26.8% of deviance. Figure 2 shows the rate of learning on the log-log scale. We reestimate the GAM using the same inverse

Figure 2. Learning Curves of Direct and Indirect Experience



Notes. Shown are the nonlinear shapes of the learning curves for direct (left) and indirect (right) experience. Clearly visible is the period of initial investment for learning from direct experience up to about the seventh submission. Models estimated using flexible GAM shown with 95% confidence interval (dependent variable: ln*Score*). Model based on Model (3) from Table 2. Shape is on the scale of the response variable. Plot excludes competition and individual-specific intercepts.

probability weights as above to adjust for panel attrition. The shape of the curve is quite similar (not shown). For direct experience, the shape is downward sloping early in individuals' careers (from the first through about the seventh submission), indicating a period of decreasing performance. After about the eighth submission, the learning curve slopes upward and performance improvements increase. A possible explanation for the downward slope instead of simply a flat curve is as follows. When joining the platform, individuals may already possess a backlog of ideas that they sort by quality before submitting. We test this possibly by estimating the elapsed time between submissions. We find that the first seven submissions are submitted with 29% shorter interval times than between submissions following the seventh. Even very late in individuals' careers, after about 150 submissions, the learning curve is positive and of the same slope, indicating that even highly experienced individuals continue to learn from direct experience, although at diminishing returns. We address the dip of the learning curve under direct experience in our discussion section.

4.3. Facing Effort Allocation: Combining Direct and Indirect Experience

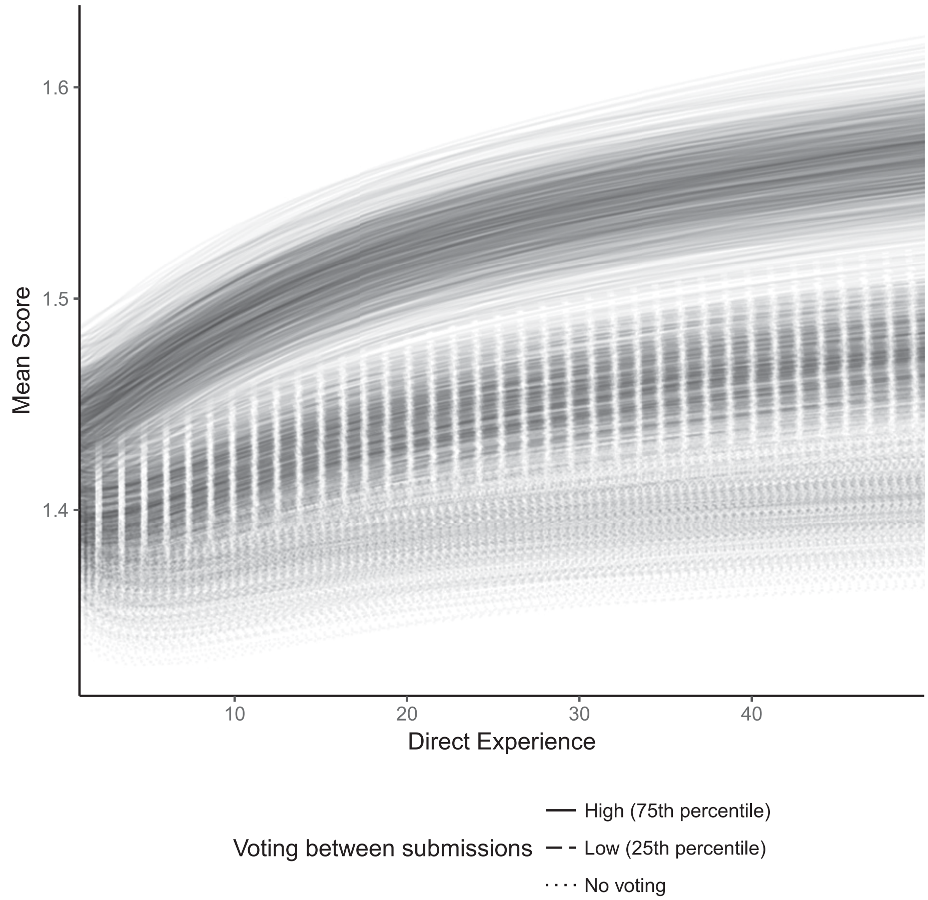
To further explore the complementary effect of direct and indirect experience, we simulate canonical behavior and plot the resulting learning curves. We simulate three canonical scenarios: an individual who does not engage in voting between submissions at all, an individual with a moderate amount of rating activity (four evaluations between submissions, which corresponds to the 25th percentile of rating volume), and an individual with high rating activity (140 ratings between submissions corresponding to the 75th percentile of rating volume). The simulation follows the approach suggested by King et al. (2000) and draws simulation parameters from parameter posteriors. We display the simulation as visually weighted regression (Hsiang 2013). We simulate 1,000 instances for each of the three canonical scenarios using the fitted nonparametric GAM model from above. Results are shown in Figure 3 using the measures on the natural scale relating experience directly to performance. Results show that the slope of the curve changes with increased numbers of ratings between submissions.

Without any rating between submissions, the performance decreases continuously. Individuals are not able to improve the performance of their subsequent submissions. Even with just a few evaluations between submissions, performance increases (on the log-scale this means that more and more improvement is derived from subsequent submissions). High voting activity substantially complements direct experience such that increasing benefits derive from subsequent submissions.

5. Discussion and Conclusion

Our research investigated how individuals in contest-based online innovation communities improve their

Figure 3. Simulation of Direct Experience Learning Curves Under Three Policies of Indirect Experience Behaviors



Notes. Bottom, dotted-line plot simulates individual who engages in no voting between submissions. Middle, dashed-line plot simulates individual who engages in low voting frequency (25th percentile; four ratings between submissions). Top, continuous-line plot simulates individual who engages in high voting frequency (75th percentile; 140 ratings between submissions).

performance by synthesizing learnable signals from different sources. We examined how more than 55,000 members of a firm-sponsored online community learned from the evaluation signals they received from making almost 180,000 t-shirt design submissions and from observing the submissions made by others while making almost 150 million quantitative evaluations.

We provide an overview of our research questions and empirical findings in Table 5, and we highlight three of the most significant results. First, we find that individuals participating in contest-based online communities need to overcome a period of initial investment in which their performance decreases. This finding is contrary to other studies that report faster early learning when improvements are easiest to achieve followed by diminishing returns (Newell and Rosenbloom 1981). Second, we find that individuals fail to improve their performance after observing others' bad examples (KC et al. 2013). However, they do improve after observing good examples. Whether individuals can learn from observing others' bad examples depends on their ability to correctly recognize the low quality of the work and then to learn error detection and correction strategies from them. Third, we find that individuals can successfully aggregate evaluation signals from both the community and the firm into a coherent learning curve (Hanna et al. 2014). This allows them to not only learn about community preferences, but also about firm preferences. This improves their likelihood of being chosen as a contest winner by the firm.

Our findings have implications for how to best structure and manage crowdsourced product innovation and for learning in both on-line and off-line communities more broadly. In the following sections, we discuss implications for crowdsourcing research, community-based learning research, and how our unique measure of the key construct of observational learning enables a richer and more nuanced analysis of learning from indirect experience.

5.1. Learning from Mixed Signals

Our study finds that individuals participating in contest-based online communities can improve their performance by learning from mixed and coarse signals

Table 5. Summary of Findings and Implications for Theory

Research questions	Empirical findings	Implications for theory
How do individuals learn from direct evaluation signals of their work?	There is a period of initial investment in which performance decreases followed by positive and rapid learning.	Extending Bower and Hilgard (1981), and Glaser (1989), individuals need to make initial investments to update cognitive structures before they are able to synthesize meaningful information from performance evaluations, especially in environments where actions are largely self-guided (cf. Brown and Duguid 1991, Wenger 1998).
	Individuals learn about preferences of the firm above-and-beyond learning to receive high scores from community.	Individuals can learn two separate preferences simultaneously (Hanna et al. 2014, Newell and Rosenbloom 1981, Rosen 1972), which is important in may contexts (Bauer et al. 2016, Dahlander and Piezunka 2014, Huang et al. 2014).
		Learning may form an additional "joining script" (von Krogh et al. 2003) in the socialization of new community members.
How do individuals learn from observing the work of others?	Observing good examples is valuable but individuals fail to improve after observing others' bad examples.	Individuals can learn merely from observing the completed work of others despite the competitive environment, the absence of developmental feedback, and the lack of rich direct interaction thought necessary for learning (Barrett et al. 2016, Brown and Duguid 1991, Wenger 1998).
	Ability to learn from bad examples depends on ability to correctly recognize the low quality of the work.	Extends work on how individuals learn to notice relevant aspects of the work they observe (Hanna et al. 2014) and quantifies observational learning that has been difficult to empirically characterize (Bandura 1977, Banerjee 1992, Bikhchandani et al. 1992).
		Failures are only valuable sources of learning if they are correctly recognized as such (e.g., Bandura 1977, Harlow 1949, Newell and Rosenbloom 1981, Thorndike 1898), and informational content of observed work is critical for learning (Bandura 1977, Firestein 2015).
How should individuals allocate effort between direct and indirect experience?	Performance decreases when individuals do not observe work of others, but there are substantial complements when level of observation is high.	Learning in environments without objective performance measures may require input from different sources, which is more of a challenge than in traditional contests where criteria are clear (Boudreau et al. 2011, Jeppesen and Lakhani 2010, Terwiesch and Xu 2008).
		Extends recent work on contests (Boudreau and Lakhani 2015, Stanko 2016, Yu and Nickerson 2011) that has only looked at the direct effects of making contest results transparently available for inspiration and reuse, but ignored learning effects.

from different information sources. We theorize that individuals' ability to synthesize learnable information from mixed signals depends critically on cognitive structures built up through prior experience in the same context.

Our finding of a dip in performance early during individual's tenure contrasts with the results reported in past empirical studies of individual learning curves. These studies generally found faster performance improvements early on when improvements are easiest to make (cf. Newell and Rosenbloom 1981). In our context, the individuals participating in online design contests (even if they are skilled designers) may be unfamiliar with the specific context of t-shirt design and what this specific community values. As a result, individuals need to make initial investments to gather the necessary experience to develop appropriate cognitive structures before they are able to synthesize the feedback they receive (Bower and Hilgard 1981, Glaser 1989). This gathering of experience takes time. Until their cognitive structures are updated, the individuals have pre-existing cognitive structures that reflect their past experience from a different context. These structures may not allow them to interpret the feedback they receive correctly. For example, an entrepreneur moving into a new market may attend to the wrong dimensions of a product innovation because their past experiences can create blinders.

The key result is that experience has a potential cost: Previous experience with a similar market may teach the entrepreneur to attend to the "wrong" things (Rogers 2003). As they begin to compete, individuals put forward their best initial efforts first, unable to synthesize useful information from the feedback they receive. Only when cognitive structures are updated by a sufficient accumulation of experience are they able to synthesize useful information from the feedback they receive. This then allows them to improve their performance. Our finding contrasts with prior learning research that generally assumes that feedback signals are objective and easy to interpret (Argote 2012, Newell and Rosenbloom 1981). This research ignores the important contingency of the updating of cognitive structures that is necessary to correctly interpret feedback signals. Consequently, research on offline learning should more closely attend to myopic processes in the early stages of innovation work.

Empirically, we exploit detailed data on the number and type of work (good versus bad examples) that individuals observe. This provides us with a key construct to measure learning from indirect experience (which is typically difficult to quantify, especially in offline settings). Measuring the rate of improvement from indirect experience allows us to complement prior work on observational learning (Bandura 1977, Banerjee 1992, Bikhchandani et al. 1992). This prior work has struggled to quantify and characterize such learning curves. In addition, prior research that has highlighted a key benefit to learning from failure builds on the assumption that the cause of failures is known and easily discernible (Bandura 1977, Banerjee 1992, Bikhchandani et al. 1992). Yet our results show that in contexts such as the one we study, this may not be the case. We theorize that the ability to recognize causes of errors is critical to learning from the failures of others. Our finding that individuals unconditionally improve their performance by observing others' good examples, but only improve by observing others' bad examples if they correctly recognize that work as low quality, is significant. It extends work on how individuals learn to notice relevant aspects of the work they observe (Hanna et al. 2014).

Our results highlight the contingency that learning from the bad examples of others critically depends on the ability of observers to recognize that they are in fact observing bad examples. Therefore, when investigating learning from other's failures, this additional contingency should be considered. While existing learning theory is built on the assumption that failures can easily be recognized (e.g., Bandura 1977, Harlow 1949, Newell and Rosenbloom 1981, Thorndike 1898), our theory suggests that failures are only valuable sources of learning if they are correctly recognized as such. Our study thus contrasts from prior work that often focused on psychological aspects of failure (e.g., overconfidence and attribution theory) in settings where failures are easy to identify, such as surgical procedures (KC et al. 2013).

We highlight the role of the informational content of observed work, especially for good versus bad examples. We therefore provide important qualifications for the often-heralded theme of "learning from failure" that applies to both online and offline contexts. An important implication of this insight in the context of product innovation is that since individuals are unlikely to improve their performance from observing bad examples, this may have unintended negative effects on the ideas being generated. This is because learning from good examples can lead to less disruptive ideas (Firestein 2015). This raises new questions regarding the design of contest-based online communities where it is common to make large stocks of past work available for inspection.

Our theorizing and empirical results show how one can induce learning without providing new data. This can be done by simply providing individuals with information about the quality of work they observe, thus highlighting relationships to which the individual has previously not attended. However, this points to a goal conflict between the learning of individuals and collecting valid group evaluations. Theories on collective intelligence suggest that individual evaluations should remain independent to avoid bias from social influence (Lorenz et al. 2011, Muchnik et al. 2013). This poses important questions about process and information system design to achieve both goals simultaneously.

Together the finding of the difficulty of synthesizing learnable signals from evaluation feedback and failure to learn from observing bad examples both point to the important role of cognitive structures, built up from past experience. This points to an important qualifier in the relationship between receiving ambiguous and hard-to-interpret signals from different sources and learning in environments where actions are largely self-guided with limited direction from mentors or teachers, and with little direct interaction and knowledge sharing with other community members (cf. Brown and Duguid 1991, Wenger 1998). We observe that different signals and the cognitive structures necessary to synthesize learnable data from these signals play an important role for individuals' performance improvement. This raises novel questions about learning in environments without objective performance measures. In the absence of a formal contest framework, developing cognitive structures to correctly recognize good and bad examples may be even more difficult. For example, individuals in an open-source software development community may observe the code written by another without explicit knowledge of the performance of that code. The ability to correctly recognize the low quality of software code (built from observing high-quality code) could then provide the basis to learn error detection and correction strategies.

These findings are relevant in the broader context of online innovation communities. They suggest important questions about learning from multiple sources of information and the role of cognitive structures in synthesizing learnable signals. These findings extend recent work on contests (Boudreau and Lakhani 2015, Stanko 2016, Yu and Nickerson 2011) that has only looked at the direct effects of making contest results transparently available for inspiration and reuse, but ignored learning effects. For example, this prior work has investigated the role of disclosure policies in terms of how they would affect reuse, but not how disclosure policies might affect learning. In this vein, our results showing learning as an integral aspect of practice raise new questions about how we should think about the parallel search facilitated by contests (Nelson 1961). Our results pose important questions regarding how individuals should weigh signals from different sources.

In summary, we advance the growing work on individuals' behavior in contest-based online communities (Boudreau et al. 2011, Jeppesen and Lakhani 2010, Terwiesch and Xu 2008) that has theorized learning as a motivation to join such communities, but never investigated it directly (Hutter et al. 2011, Majchrzak and Malhotra 2016).

5.2. Joining and Learning in Online and Offline Communities

Our study complements online community research by von Krogh et al. (2003) on the socialization of new members. Their study demonstrates that individuals follow a "joining script" to become productive members of an online community. However, our results suggest that the simple peripheral participation they describe may not be sufficient to become a successful member in contest-based online communities. Learning to notice and extract learnable signals among different mixed signals may be a critical aspect of the joining process. Our results thus extend our understanding of the critical process by which individuals successfully join online communities, allowing these communities to be sustainable over time.

Our results also inform the broader literature on community-based learning (both online and offline) by providing novel insights using a large, longitudinal data set from one of the most successful crowdsourcing firms. A key insight is that despite the competitive environment, the absence of developmental feedback, and the lack of rich direct interaction thought necessary for learning (Barrett et al. 2016, Brown and Duguid 1991, Wenger 1998), individuals can learn merely from observing the completed work of others. This highlights a complementary pathway by which individuals learn from observing the completed work of others, one that is not fully considered in current learning theories (Nonaka 1994). Thus, our work highlights an important pathway for learning that should be considered in research on both online (Barrett et al. 2016, Dahlander and Piezunka 2014) and offline user communities. This research generally assumes that the strongest community benefits accrue through the direct interactions that these communities facilitate.

The ability to learn (even in an environment that fosters direct competition and provides individuals with only rare and coarse evaluation signals) and the disclosure of completed work shows that the rich interactions thought necessary for successful knowledge creation in online communities (Barrett et al. 2016) may not be the only pathway for these communities to generate value. However, our analyses also illuminate how learning from these sparse and hard-to-interpret signals critically depends on individuals having the cognitive structures necessary to synthesize learnable data from them. By investigating different signals in the same study, we not only examine the relationships between these signals and individual performance improvement. We also provide insights into the relative importance of each signal and their synergies. This suggests the need to more clearly distinguish different pathways of learning in future work.

Our finding that individuals can learn from both the rare signals from the firm and more regular signals from the community shows that it is possible (even though difficult) to learn two production functions simultaneously. Prior research on learning has not considered learning of two separate preferences simultaneously (Hanna et al. 2014, Newell and Rosenbloom 1981, Rosen 1972). However, multiple sets of preferences need to be learned in many contexts (Bauer et al. 2016, Dahlander and Piezunka 2014, Huang et al. 2014). Our results complement these studies by demonstrating that individuals' learning curves can reflect not only what is valued by the community, but also what is valued by the firm—thus integrating information about two production functions simultaneously. This is an important insight because the ability to cater to the diverging preferences of multiple stakeholders is also common in other settings with combined community and firm evaluations (Catalini and Hui 2017). Similar dynamics also exist in offline innovation contexts. For example, industrial designers will receive feedback from peers within their community on works-in-progress, while also seeking rare signals of winning a design competition or a new client.

Taken together, our results suggest that the processes by which members join communities play an important role. This raises new questions about how new members can be successfully brought up to speed to make contest-based online communities more productive and sustainable. Thus, our study of learning in a competitive environment may help in developing new theories of community-based learning that should more directly consider learning from observing completed work (in addition to direct interaction and knowledge exchange).

5.3. Managerial Implications

Our results also suggest several implications for managers. First, managers of online innovation communities can mitigate the negative effects of the performance dip by simply increasing member awareness that there is a period of initial investment in which their rate of improvement might be slow. Second, managers can encourage individuals to observe the work produced by others to bolster their learning from indirect experience. Furthermore, individuals could be encouraged (through various platform features) to specifically observe good examples early on in their careers to facilitate the development of references of both what to do and how to do it correctly. Conversely, individuals should be informed when they are observing a bad example so that they can more effectively learn error detection and correction strategies from them. This may be particularly critical as there are typically many low-performance contest submissions available for individuals to engage with. Automated recommender algorithms could be developed that expose individuals with limited experience (who likely lack sophisticated cognitive structures to correctly recognize the quality of work) to good examples early in their career, rather than bad ones. Third, managers may actively try to provide more frequent and more informative signals about firm choices. Rather than simply announcing one winner per contest, managers could improve the learning of individuals by providing additional information about who was a runner-up (or even a full ranking of entrants). They could also provide explanations for why one submission was chosen as the winner and what aspects were lacking in the other entries. This would create additional, rich learning signals to help individuals get up to speed quickly and thus produce even more valuable innovation ideas.

5.4. Conclusion

Our study context provides a rare opportunity for insight on how individuals learn in a complex environment where they are presented with a mix of conflicting and coarse signals from different sources. While much past theorizing on learning assumes relatively unambiguous signals that individuals can use to improve their performance, we show how they can improve by becoming better at synthesizing learnable signals from a variety of information sources. We empirically measure and characterize the shape of learning curves for both direct and indirect experience. Individual learning is a cornerstone of group and organizational learning processes. Understanding how individuals make sense of myriad and at times conflicting signals is fundamental to understanding how aggregate levels of group and organizational learning might be impacted by how information is made available to individuals. While our research context is contest-based online innovation communities, many of the signals that we describe can be found in other environments where an individual can observe the completed work of others and gain direct feedback from peers and organizations. We provide new insight to the commonly held view that learning curves always lead to improvement by examining how individuals synthesize learnable signals from different information sources and the role of a period of initial investment to update cognitive structures. Our study shows how innovation and individual performance improvement unfolds over time in a complex context of information flows as they are increasingly a part of our innovation landscape.

Acknowledgments

The authors thank David Lazer, Ella Miron-Spektor, Siobhan O'Mahony, Davide Ravasi, Imke Reimers, and Ray Reagans for helpful comments on earlier versions of the manuscript. The project benefitted from feedback from seminar participants at Harvard Business School, Boston College, Mines Paris Tech, and the University of Massachusetts at Amherst.

Endnotes

[1] Learning from direct experience is also called learning from personal experience or trial-and-error learning. Learning from indirect experience is also called vicarious learning, observational learning, or learning from the experience of others.

[2] See Lave and Wenger (1991) for a similar conception of learning at the synthesis of the agent, the activity, and the environment.

[3] Using different cutoff values to specify the good/bad split does not substantially change our results (e.g., mean versus median split versus tercile split, considering the "neutral" category in the middle an additional control).

[4] We also estimated models in which we specified first-order autoregressive AR(1) as the within-subject covariance structure for repeated measures in a default error structure where subjects are assumed to be independent. Results are consistent between the different specifications.

[5] The data set contains 179,794 design submissions of 56,612 individuals, all of which are included in the computation of key variables used in this study. However, the notion of learning is only meaningful for individuals who made at least two design submissions. Hence, the number of observations included in the panel regression includes only those individuals that made at least two design submissions. The analysis using *RatingError* includes only individuals who cast at least one rating between at least two submissions.

[6] Results of judgment and choice decisions are not necessarily identical as judgment can be ignored at the point of choice (Einhorn and Hogarth 1981). In fact, we find this to be the case in our empirical data in the sense that designs chosen for print are usually extremely highly rated but not necessarily the highest-rated one (i.e., the contest winner chosen by Threadless is sometimes the second or third highest rated design; see the online appendix for details).

References

Afuah A, Tucci CL (2012) Crowdsourcing as a solution to distant search. *Acad. Management Rev.* 37(3):355–375.

Al-Abood SA, Davids K, Bennett SJ (2001) Specificity of task constraints and effects of visual demonstrations and verbal instructions in directing learners' search during skill acquisition. *J. Motor Behav.* 33(3):295–305.

Almirall E, Lee M, Majchrzak A (2014) Open innovation requires integrated competition-community ecosystems: Lessons learned from civic open innovation. *Bus. Horizons* 57(3):391–400.

Andrieux M, Proteau L (2016) Observational learning: Tell beginners what they are about to watch and they will learn better. *Frontiers Psych.* 7:51.

Argote L (2012) *Organizational Learning—Creating, Retaining and Transferring Knowledge* (Springer Science & Business Media, New York).

Argote L, Epple D (1990) Learning curves in manufacturing. *Science* 247(4945):920–924.

Argote L, McEvily B, Reagans R (2003) Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management Sci.* 49(4):571–582.

Arrow KJ (1962) The economic learning implications of by doing. *Rev. Econom. Stud.* 29(3):155–173.

Badets A, Blandin Y (2004) The role of knowledge of results frequency in learning through observation. *J. Motor Behav.* 36(1):62–70.

Bandura A (1971) *Social Learning Theory* (General Learning Press, New York).

Bandura A (1977) *Social Learning Theory* (Prentice-Hall, Englewood Cliffs, NJ).

Bandura A, Jeffrey RW (1973) Role of symbolic coding and rehearsal processes in observational learning. *J. Personality Soc. Psych.* 26(1):122–130.

Banerjee AV (1992) A simple model of herd behavior. *Quart. J. Econom.* 107(3):797–817.

Barrett M, Oborn E, Orlikowski W (2016) Creating value in online communities: The sociomaterial configuring of strategy, platform, and stakeholder engagement. *Inform. Systems Res.* 27(4):704–723.

Bates D, Maechler M, Bolker BM, Walker S (2015) Fitting linear mixed-effects models using {lme4}. Accessed June 15, 2017, http://arxiv.org/abs/1406.5823.

Bauer J, Franke N, Tuertscher P (2016) Intellectual property norms in online communities: How user-organized intellectual property regulation supports innovation. *Inform. Systems Res.* 27(4):724–750.

Bechky BA (2003) Sharing meaning across occupational communities: The transformation of understanding on a production floor. *Organ. Sci.* 14(3):312–330.

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econom.* 100(5):992–1026.

Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the behavior of others: Conformity, fads, and informational cascades. *J. Econom. Perspect.* 12(3):151–170.

Blandin Y, Proteau L (2000) On the cognitive basis of observational learning: Development of mechanisms for the detection and correction of errors. *Quart. J. Experiment. Psych. Sect. A* 53(3):846–867.

Blandin Y, Lhuisset L, Proteau L (1999) Cognitive processes underlying observational learning of motor skills. *Quart. J. Experiment. Psych. Sect. A* 52(4):957–979.

Blohm I, Riedl C, Fuller J, Leimeister JM (2016) Rate or trade? Identifying winning ideas in open ideas sourcing. *Inform. Systems Res.* 27(1):27–48.

Boudreau KJ (2012) Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. *Organ. Sci.* 23(5):1409–1427.

Boudreau KJ, Lakhani KR (2015) "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Res. Policy* 44(1):4–19.

Boudreau KJ, Lacetera N, Lakhani KR (2011) Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Sci.* 57(5):843–863.

Boudreau KJ, Lakhani KR, Menietti M (2016) Performance responses to competition across skill levels in rank-order tournaments: Field evidence and implications for tournament design. *RAND J. Econom.* 47(1):140–165.

Bourdieu P (1990) *The Logic of Practice* (Stanford University Press, Stanford, CA).

Bower GH, Hilgard ER (1981) *Theories of Learning* (Prentice-Hall, Englewood Cliffs, NJ).

Brabham DC (2010) Moving the crowd at Threadless. *Inform. Comm. Soc.* 13(8):1122–1145.

Brown JS, Duguid P (1991) Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organ. Sci.* 2(1):40–57.

Buchanan JJ, Dean NJ (2010) Specificity in practice benefits learning in novice models and variability in demonstration benefits observational practice. *Psych. Res.* 74(3):313–326.

Camerer CF (2011) *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press, Princeton, NJ).

Carroll WR, Bandura A (1990) Representational guidance of action production in observational learning: A causal analysis. *J. Motor Behav.* 22(1):85–97.

Catalini C, Hui X (2017) Can capital defy the law of gravity? Investor networks and startup investment. Working paper, Massachusetts Institute of Technology, Cambridge, MA.

Dahl DW, Fuchs C, Schreier M (2014) Why and when consumers prefer products of user-driven firms: A social identification account. *Management Sci.* 61(8):1978–1988.

Dahlander L, Magnusson MG (2005) Relationships between open source software companies and communities: Observations from Nordic firms. *Res. Policy*. 34(4):481–493.

Dahlander L, Piezunka H (2014) Open to suggestions: How organizations elicit suggestions through proactive and reactive attention. *Res. Policy*. 43(5):812–827.

Darr ED, Argote L, Epple D (1995) The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Sci.* 41(11):1750–1762.

Einhorn HJ, Hogarth RM (1981) Behavioral decision theory: Processes of judgment and choice. *Annual Rev. Psych.* 32(1):53–58.

Epple D, Argote L, Murphy K (1996) An empirical investigation of the microstructure of knowledge acquisition and transfer through learning by doing. *Oper. Res.* 44(1):77–86.

Firestein S (2015) *Failure: Why Science Is So Successful* (Oxford University Press, Oxford, UK).

Fitzgerald J, Gottschalk P, Moffitt R (1998) An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *J. Human Resources*. 33(2):251–299.

Foster AD, Rosenzweig MR (1995) Learning by doing and learning from others: Human capital and technical change in agriculture. *J. Political Econom.* 103(6):1176–1209.

Füller J, Hutter K, Hautz J, Matzler K (2014) User roles and contributions in innovation-contest communities. *J. Management Inform. Systems* 31(1):273–308.

Gelman A, Hill J (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge, UK).

Gioia DA, Manz CC (1985) Linking cognition and behavior: A script processing interpretation of vicarious learning. *Acad. Management Rev.* 10(3):527–539.

Glaser, R (1989) Expertise and learning: How do we think about instructional processes now that we have discovered knowledge structures. Klahr D, Kotovsky K, eds. *Complex Information Processing: The Impact of Herbert A. Simon* (Lawrence Erlbaum Associates, Hillsdale, NJ), 269–282.

Greene WH (2003) *Econometric Analysis*, 5th ed. (Prentice-Hall, Upper Saddle River, NJ).

Hanna R, Mullainathan S, Schwartzstein J (2014) Learning through noticing: Theory and evidence from a field experiment. *Quart. J. Econom.* 129(3):1311–1353.

Harlow HF (1949) The formation of learning sets. *Psych. Rev.* 56(1):51–65.

Haunschild PR, Sullivan BN (2002) Learning from complexity: Effects of prior accidents and incidents on airlines' learning. *Admin. Sci. Quart.* 47(4):609–643.

Heyes C, Foster C (2002) Motor learning by observation: Evidence from a serial reaction time task. *Quart. J. Experiment. Psych. Section A* 55(2):593–607.

Hildebrand C, Häubl G, Herrmann A, Landwehr JR (2013) When social media can be bad for you: Community feedback stifles consumer creativity and reduces satisfaction with self-designed products. *Inform. Systems Res.* 24(1):14–29.

Howell ST (2017) Learning from feedback: Evidence from new ventures. National Bureau of Economic Research Working Paper 23874, New York University, New York.

Hsiang SM (2013) Visually-weighted regression. Working paper, University of California, Berkeley, Berkeley.

Huang Y, Singh PV, Srinivasan K (2014) Crowdsourcing new product ideas under consumer learning. *Management Sci.* 60(9):2138–2159.

Hutter K, Hautz J, Füller J, Mueller J, Matzler K (2011) Communitition: The tension between competition and collaboration in community-based design contests. *Creativity Innovation Management* 20(1):3–21.

James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning with Applications in R* (Springer, New York).

Jeppesen LB, Frederiksen L (2006) Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organ. Sci.* 17(1):45–63.

Jeppesen LB, Lakhani KR (2010) Marginality and problem-solving effectiveness in broadcast search. *Organ. Sci.* 21(5):1016–1033.

Jones DG, Endsley MR (2000) Overcoming representational errors in complex environments. *Human Factors* 42(3):367–378.

KC D, Staats BR, Gino F (2013) Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. *Management Sci.* 59(11):2435–2449.

King G, Tomz M, Wittenberg J (2000) Making the most of statistical analyses: Improving interpretation and presentation. *Amer. J. Political Sci.* 44(2):347–361.

Kornish LJ, Ulrich KT (2011) Opportunity spaces in innovation: Empirical analysis of large samples of ideas. *Management Sci.* 57(1):107–128.

Krogh Gv, Hippel Ev (2006) The promise of research on open source software. *Management Sci.* 52(7):975–983.

Lakhani KR, Wolf RG (2005) Why hackers do what they do: Understanding motivation and effort in free/open source software projects. Feller J, Fitzgerald B, Hissam S, Lakhani KR, eds. *Perspectives on Free and Open Source Software* (MIT Press, Cambridge, MA), 3–22.

Lave J, Wenger E (1991) *Situated Learning: Legitimate Peripheral Participation* (Cambridge University Press, New York).

Levina N, Arriaga M (2014) Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields. *Inform. Systems Res.* 25(3):468–488.

Levitt B, March JG (1988) Organizational learning. *Annual Rev. Sociol.* 14:319–340.

Lorenz J, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of crowd effect. *Proc. Natl. Acad. Sci. USA* 108(22):9020–9025.

Majchrzak A, Malhotra A (2016) Effect of knowledge-sharing trajectories on innovative outcomes in temporary online crowds. *Inform. Systems Res.* 27(4):685–703.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

Moffitt R, Fitzgerald J, Gottschalk P (1999) Sample attrition in panel data: The role of selection on observables. *Ann. Econom. Statist.* (55-56):129–152.

Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Science* 341(6146):647–651.

Nadler J, Thompson L, van Boven L (2003) Learning negotiation skills: Four models of knowledge creation and transfer. *Management Sci.* 49(4):529–540.

Nelson RR (1961) Uncertainty, learning, and the economics of parallel research and development efforts. *Rev. Econom. Statist.* 43(4):351–364.

Newell A, Rosenbloom PS (1981) Mechanisms of skill acquisition and the law of practice. Anderson JR, ed. *Cognitive Skills and Their Acquisition* (Lawrence Erlbaum Associates, Hillsdale, NJ), 1–55.

Newell A, Simon HA (1972) *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, NJ).

Nonaka I (1994) A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37.

O'Mahony S, Lakhani K (2011) Organizations in the shadow of communities. Marquis C, Lounsbury M, Greenwood R, eds. *Research in the Sociology of Organizations: Communities and Organizations*, vol. 33 (Emerald Group Publishing, Bingley, UK), 3–35.

Orlikowski WJ, Yates J (1994) Genre repertoire: The structuring of communicative practices in organizations. *Admin. Sci. Quart.* 39(4):541–574.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2015) Linear and non-linear mixed effects models. Accessed June 1, 2017, http://cran.r-project.org/package=nlme.

Preece J, Shneiderman B (2009) The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Trans. Human-Comput. Interaction* 1(1):13–32.

R Core Team (2015) R: A language and environment for statistical computing. Accessed May 3, 2017, http://www.r-project.org.

Riedl C, Woolley A (2017) Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and collaboration to crowd-based problem solving performance. *Acad. Management Discov.* 3(4):382–403.

Riedl C, Zanibbi R, Hearst MA, Zhu S, Menietti M, Crusan J, Metelsky I, Lakhani KR (2016) Detecting figures and part labels in patents: competition-based development of graphics recognition algorithms. *Internat. J. Document Anal. Recognition* 19(2):155–172.

Roberts JA, Hann I-H, Slaughter SA (2006) Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects. *Management Sci.* 52(7):984–999.

Robins JM, Rotnitzky A, Zhao LP (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* 90(429):106–121.

Rogers EM (2003) *The Diffusion of Innovations*, 5th ed. (Free Press, New York).

Organization Science, 2018, vol. 29, no. 6, pp. 1010–1032, © 2018 The Author(s)

Rosen S (1972) Learning by experience as joint production. *Quart. J. Econom.* 86(3):366–382.

Seidel VP, Langner B (2015) Using an online community for vehicle design: Project variety and motivations to participate. *Indust. Corporate Change* 24(3):635–653.

Singh PV, Tan Y, Youn N (2011) A hidden Markov model of developer learning dynamics in open source software projects. *Inform. Systems Res.* 22(4):790–807.

Stanko MA (2016) Toward a theory of remixing in online innovation communities. *Inform. Systems Res.* 27(4):773–791.

Sweller J (1988) Cognitive load during problem solving: Effects on learning. *Cognitive Sci.* 12(2):257–285.

Terwiesch C, Xu Y (2008) Innovation contests, open innovation, and multiagent problem solving. *Management Sci.* 54(9): 1529–1543.

Thomke SH (1998) Managing experimentation in the design of new products. *Management Sci.* 44(6):743–762.

Thorndike EL (1898) Animal intelligence: An experimental study of the associative processes in animals. *Psychol. Rev. Monograph Supplements* 2(4):i–109.

Thornton RA, Thompson P (2001) Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding. *Amer. Econom. Rev.* 91(5): 1350–1368.

Van Knippenberg D, Dahlander L, Haas MR, George G (2015) Information, attention, and decision making. *Acad. Management J.* 58(3):649–657.

Verbeek M, Nijman T (1992) Testing for selectivity bias in panel data models. *Internat. Econom. Rev.* 33:681–703.

von Krogh G, Spaeth S, Lakhani KR (2003) Community, joining, and specialization in open source software innovation: A case study. *Res. Policy* 32(7):1217–1241.

von Krogh G, Haefliger S, Spaeth S, Wallin MW (2012) Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quart.* 36(2):649–676.

Wenger E (1998) *Communities of Practice: Learning, Meaning, and Identity* (Cambridge University Press, New York).

Wiersma E (2007) Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Sci.* 53(12):1903–1915.

Wooldridge JM (2002) Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Econom. J.* 1(2):117–139.

Yates J, Orlikowski WJ, Woerner SL (2003) Virtual organizing: Using threads to coordinate distributed work. Sprague RH Jr, ed. *Proc. 36th Annual Hawaii Internat. Conf. System Sci.* (IEEE, Los Alamitos, CA), 1–10.

Yu L, Nickerson JV (2011) Cooks or cobblers? Crowd creativity through combination. *Proc. SIGCHI Conf. Human Factors Comput. Systems*. (ACM, New York), 1393–1402.

Christoph Riedl is assistant professor of information systems and network science at the D'Amore-McKim School of Business, and the College of Computer and Information Science at Northeastern University. He is also a faculty member at the Network Science Institute and a fellow at Harvard Institute of Quantitative Social Science. He received his PhD from Technische Universität München. His research interest is to understand how social and economic networks shape collaboration.

Victor P. Seidel is an associate professor at the F. W. Olin Graduate School of Business at Babson College. He is also a TECH Innovation Fellow at the Harvard John A. Paulson School of Engineering and Applied Sciences and a visiting scholar at Said Business School, University of Oxford. He received his PhD from Stanford. His research interests include organizational practices supporting product innovation, the role of online communities in innovation, and the application of design methods.