© *Academy of Management Journal*
2015, Vol. 58, No. 3, 856–880.
http://dx.doi.org/10.5465/amj.2012.0458

# DISTANT SEARCH, NARROW ATTENTION: HOW CROWDING ALTERS ORGANIZATIONS' FILTERING OF SUGGESTIONS IN CROWDSOURCING

HENNING PIEZUNKA
INSEAD

LINUS DAHLANDER
ESMT European School of Management and Technology

**In their search for innovation, organizations often invite suggestions from external contributors. Soliciting suggestions is a form of distant search, since it allows organizations to tap into knowledge that may not reside within their organizational boundaries. Organizations engaging in distant search often face a large pool of suggestions, an outcome we refer to as crowding. When crowding occurs, organizations, given a limited attention span, can attend to only a subset of suggestions. Our core argument is that crowding narrows the attention of organizations; that is, despite organizations' efforts to reach out to external contributors and access suggestions that capture distant knowledge, they are more likely to pay attention to suggestions that are familiar, not distant. We test our theory with a unique longitudinal dataset that captures how 922 organizations responded to 105,127 crowdsourced suggestions from external contributors. After distinguishing between three different dimensions of distance (content, structural, and personal), we find that (a) all three types of distance have independent negative effects on the likelihood of attention, (b) crowding amplifies these negative effects, and (c) there are differences among the effects' magnitudes. We elaborate on the broader implications of these findings for the literatures on attention, search, and crowdsourcing.**

> People think focus means saying "yes" to the thing you've got to focus on, but that's not what it means at all. It means saying "no" to the hundred other good ideas. (Steve Jobs, as quoted in Simons, 2010: 97)

During the devastating *Deepwater Horizon* accident, British Petroleum (BP) reached outside company boundaries by issuing an open call for suggestions from external contributors, and asking the general public what could be done to deal with the catastrophe. BP hoped that external contributors would find creative solutions to the pressing problem of the oil spill. A crowd of external contributors from a vast array of knowledge domains responded to the open call; the 120,000 suggestions BP received represented one of the largest pools of external suggestions ever documented (BP, 2010). However, BP's story is far from unique. Dating back as far as Napoleon, who used crowdsourcing to find solutions to famine and irrigation, and following the rise of the Internet and social media in more recent years, organizations often have reached outside of their boundaries to elicit suggestions from large and diverse crowds of external contributors (Dahlander & Piezunka, 2014; Robinson & Stern, 1998).

From a theoretical standpoint, the potential benefits of engaging in *distant search* by soliciting suggestions from crowds of external contributors are

We appreciate the very helpful comments from Gerry George and three anonymous reviewers, as well as the support of Steve Barley, Matt Bothner, Daisy Chung, Kathleen Eisenhardt, Martin Gargiulo, Dietmar Harhoff, Eric von Hippel, Lars Bo Jeppesen, Riitta Katila, Michael Lenox, Rory McDonald, Renee Rottner, Kurt Sandholtz, Catalina Stefanescu-Cuntze, and Chris Tucci. We also thank seminar participants at Aalborg University, ETH Zürich, INSEAD, London Business School, Ludwig Maximilian University, and the Academy of Management Conference 2012. We have received generous financial support from the Stanford Technology Ventures Program, the Peter Curtius-Stiftung (Peter Curtius Foundation), the Sloan Research Project on the Economics of Knowledge Contribution, and the Institute for Research in the Social Sciences at Stanford University.

well established. Two benefits stand out: First, distant search gives organizations access to knowledge that may not reside within their organizational boundaries—that is, the resulting suggestions might represent *distant knowledge* (Afuah & Tucci, 2012; Chesbrough, 2003). Existing research on knowledge spaces has introduced the concept of knowledge distance to capture how knowledge domains relate to one another (Fleming & Sorenson, 2004; George, Kotha, & Zheng, 2008; Kotha, George, & Srikanth, 2013). In referring to distant knowledge, we refer to knowledge that has a great distance from an organizations' current knowledge base (i.e., knowledge that is not familiar). While such knowledge is distant for an organization, its contributors might find the knowledge familiar. In fact, they might have highly refined understandings of such knowledge (Afuah & Tucci, 2012; Shane, 2000; von Hippel, 1986). Second, by soliciting suggestions from a crowd of external contributors, organizations gain access to a large number of suggestions (as illustrated by the BP example). Such a large number of suggestions allows organizations to be more choosy about which suggestions are worthy of their attention (Girotra, Terwiesch, & Ulrich, 2010; Gruber, MacMillan, & Thompson, 2008; Kornish & Ulrich, 2012) and to identify overarching trends (Surowiecki, 2005). Taken together, engaging in distant search by soliciting suggestions from external contributors provides access to numerous suggestions capturing distant knowledge, and can thus increase organizations' ability to innovate.

Once organizations have elicited suggestions from their contributors, the resulting knowledge needs to be filtered (Cooper, 1985, 2008; Salter, ter Wal, Criscuolo, & Alexy, forthcoming). Attention cannot be paid to all elicited suggestions because the attention span of organizations is limited (March & Simon, 1958; Ocasio, 1997; Sullivan, 2010). Intuitively, we expect that the attention of organizations will be focused on suggestions representing distant knowledge. Gaining access to such knowledge is one of the supposed key reasons for organizations to solicit suggestions from external contributors in the first place. Yet, research shows that it is challenging to attend to distant knowledge. Specifically, it is difficult for organizations to combine elements of their current knowledge base with newly elicited distant knowledge (Kotha et al., 2013). It is thus unclear whether the attention of organizations focuses on suggestions capturing distant knowledge ("wide attention") or on suggestions capturing knowledge that is adjacent to their current knowledge base ("narrow attention"). Our first research question addresses this unresolved issue by asking *whether organizations that elicit distant knowledge ultimately pay attention to distant knowledge*.

If organizations are successful in eliciting distant knowledge, they are exposed to numerous suggestions—a state we refer to as *crowding* (Ocasio, 2010: 1290). Crowding requires organizations to filter elicited suggestions particularly severely: The stronger the crowding, the more severe the filtering. Thus, when crowding occurs, organizations pay attention to only a small subset of the elicited suggestions. While this baseline effect (i.e., that crowding reduces the share of suggestions that receive attention) is well established in other contexts (Hansen & Haas, 2001; Sullivan, 2010), we wondered whether—and, if so, how—crowding affects how organizations filter suggestions. If crowding changes how organizations filter suggestions, such a change might affect their tendency to pay attention to distant knowledge. We thus ask, in our second research question, *whether organizations pay more or less attention to distant knowledge as they are exposed to crowding*.

Intuitively, we expect that organizations shift their attention toward suggestions representing distant knowledge as crowding increases because crowding widens organizations' search. Suggestions representing distant knowledge are often rare (Jeppesen & Lakhani, 2010), and crowding gives organizations access to more such suggestions (in absolute terms). In addition to organizations' interest in such suggestions, suggestions capturing distant knowledge are also likely to be more salient than other suggestions, and are, thus, more likely to attract organizations' attention (Li, Maggitti, Smith, Tesluk, & Katila, 2013). For example, in the large pool of suggestions elicited by BP, suggestions representing distant knowledge (e.g., genetically modifying bacteria to foster the breakdown of the spilled oil) were likely to stand out. This shift in organizations' attention toward suggestions representing distant knowledge fits the intuition that organizations with more suggestions (i.e., crowding) are more innovative (i.e., they pay more attention to distant knowledge). Overall, this perspective suggests that an increase in crowding lets organizations shift their attention toward suggestions that capture distant knowledge.

There is, however, an alternative scenario. It is possible that, as crowding increases, organizations become less likely to select suggestions that represent distant knowledge—a behavior we describe as a *narrowing of attention*. To cope with crowding, organizations may adjust their filtering to become simpler, and, as a result, filter out suggestions that capture distant knowledge. Such a narrowing of attention in response to crowding would be paradoxical and would counter the original intent of the organizations. However, we argue that this phenomenon does occur, which implies that organizations that solicit suggestions with the intention of gaining access to distant knowledge often end up filtering out that same knowledge once they are successful (i.e., once crowding occurs).

Examining our research question was a formidable empirical challenge. The first challenge was to access a complete set of suggestions (i.e., not only those that were implemented) that organizations had elicited from external contributors. After all, organizations' attention cannot be fully understood when the available data are limited to the suggestions to which they eventually pay attention (Denrell, 2003, 2005). We overcame this challenge by constructing a unique, large-scale, longitudinal dataset of suggestions that external contributors had submitted to a broad range of organizations. We created this dataset in collaboration with a private company that allowed us to study 105,127 suggestions submitted to 922 organizations between November 2007 and June 2011. Our second empirical challenge was to collect sufficiently detailed information about each of the suggestions. We met this challenge by documenting (a) each suggestion's meta-information (e.g., the identity of the external contributor; the suggestion date; and any public support, such as votes or comments, that the suggestion attracted from other external contributors subsequent to its submission); (b) each suggestion's indirect affiliation with other suggestions, to which it may have been linked via the engagement of external contributors (i.e., two suggestions were linked if one user voted, commented, or voted and commented on both suggestions); and (c) a suggestion's content, which we analyzed using a novel form of text analysis that built on recent work in information retrieval and examined characteristics on the level of the suggestions (e.g., suggestion length, number of positive or negative words, etc.) and on the level of the organization (e.g., similarity to other suggestions). In sum, to address our research question, we generated a unique dataset and explored it with cutting-edge network and text analysis methods.

Our results strongly support our core argument that *crowding leads organizations to narrow their attention*. Consistent with prior literature, we find that crowding reduces the chances of any single suggestion receiving an organization's attention. However, we also find that crowding prompts an organization to narrow its attention, such that suggestions representing distant knowledge are less likely to be accepted. We develop and test our core hypothesis by evaluating each suggestion along three dimensions of distance: the distances of (1) the suggestion's content (i.e., the actual description of the suggestion), (2) its structure (i.e., the types of contributors who support it), and (3) its personal connections (i.e., its relationship to the contributor from whom it originated). The results are consistent, albeit with different magnitudes of effects, across all three dimensions.

Our paper has implications for the literatures on attention, search, and crowdsourcing. With respect to *attention*, we examine how crowding changes what organizations pay attention to. Prior research has pointed out that organizations' attentional limits force them to engage in more severe filtering (Hansen & Haas, 2001; March & Simon, 1958; Ocasio, 1997; Sullivan, 2010). We find that crowding also changes *how* organizations allocate their attention by narrowing that attention and filtering out suggestions that capture distant knowledge. With respect to *search*, we examine why organizations often fail to pay attention to distant knowledge. Prior research has shown how organizations often succeed in gaining access to distant knowledge (Katila & Ahuja, 2002; Laursen & Salter, 2006; Rosenkopf & Nerkar, 2001), but then fail to make use of it (Kotha et al., 2013; Reitzig & Sorenson, 2013). We show that organizations' very success in eliciting many suggestions (i.e., achieving crowding) may contribute to their failure to pay attention to the distant knowledge they have collected. With respect to *crowdsourcing*, we focus on the filtering process that occurs subsequent to soliciting suggestions from a crowd. Prior research has focused on the phase and potential of soliciting suggestions from crowds (Afuah & Tucci, 2012; Dahlander & Piezunka, 2014; Jeppesen & Lakhani, 2010; Surowiecki, 2005); the current research examines organizations' handling of the suggestions they have solicited. By illustrating the challenges that organizations face in the filtering process, we show how organizations that do not handle filtering well may fail to tap into the full potential of crowdsourcing.

## THEORETICAL BACKGROUND AND HYPOTHESES

### The Challenge of Crowding

Crowding is challenging because organizations are limited in terms of their attention and their ability to process information. Such limits were noted early on in the field of organizational research (Cyert & March, 1963; Simon, 1947), and have sparked various streams of study, such as the attention-based view (Ocasio, 1997) and research on information overload (Huber & Daft, 1987). These streams deal with the origins, challenges, and consequences of the disparity between the myriad of stimuli to which organizations are exposed and the amount of stimuli they can actually process. Ocasio (1997: 193) explicated these "bandwidth" problems when he points out that organizations are "bounded in their capacity to attend to all (or even most) environmental stimuli that impinge, directly or indirectly, upon a particular situation." Such bandwidth problems have been documented in different contexts, such as the ways in which organizations process regulations (Sullivan, 2010) and retrieve information from knowledge management systems (Hansen & Haas, 2001). The key consequence of crowding is thus competition for organizations' attention, since "the attention devoted to a particular decision [...] depends on alternative claims of attention" (Cyert & March, 1963: 199). As such, we assume, as a non-controversial baseline, that organizational attention to any given suggestion will decline as crowding increases.

The research regarding the degree to which crowding changes how organizations filter—and, thus, what they pay attention to—is limited. Research on information overload has pointed out that organizations often miss valuable information because they are simultaneously exposed to an overload of worthless information. As a result, organizations' decisions become poorer, despite their access to more information (and their level of comfort regarding such access) (O'Reilly, 1980). Research on attention has also identified a shift in attention that occurs during the process of filtering. Sullivan (2010) observed, in her study of airlines' processing of new regulations, that an increase in new regulations shifted airlines' focus toward more urgent regulations. While these studies indicate that organizations change how they filter when crowding occurs, they do not resolve the question of whether an organization will allocate more or less attention to suggestions representing distant knowledge.

### Search for Distant Knowledge

Organizations reach beyond their boundaries to gain access to distant knowledge. Different organizations rely on different means to gain access to such knowledge: for example, hiring (Rosenkopf & Nerkar, 2001; Tzabbar, 2009), acquisition (Ahuja & Katila, 2001), and alliances (Stuart & Podolny, 1996). We study how organizations seek access to distant knowledge by soliciting suggestions from external contributors (Afuah & Tucci, 2012; Jeppesen & Lakhani, 2010). Organizations that gain access to distant knowledge become more innovative as a result (see Laursen (2012) for a comprehensive overview). The mechanism underlying this increase in innovativeness is that organizations combine the newly gained, distant knowledge with their previously held, familiar knowledge (Katila & Ahuja, 2002). It is the distant knowledge in particular that has the greatest potential for breakthrough innovations (Fleming, 2001; Fleming & Sorenson, 2004; Singh & Fleming, 2010). Gaining access to such knowledge also allows organizations to differentiate themselves from their competition (Katila, Chen, & Piezunka, 2012; Piezunka & Hannah, 2014). Thus, organizations that solicit suggestions from contributors strive for and benefit from gaining access to distant knowledge.

Other research offers a different picture. Many studies suggest that organizations often struggle to make use of distant knowledge. For example, business units often fail to attend to knowledge originating in business units to which they are only weakly tied (Hansen, 1999; Reitzig & Sorenson, 2013), scientific evaluation committees filter out research proposals that draw on distant knowledge domains (Boudreau, Guinan, Lakhani & Riedl, forthcoming), and universities struggle to commercialize interdisciplinary research that builds on disciplines that are distant from one another (Kotha et al., 2013). Recent lab experiments show that people who claim to like creative ideas are, in fact, much more likely to be biased against them (Mueller, Melwani, & Goncalo, 2012). Moreover, this bias escalates when people experience uncertainty (Mueller et al., 2012). Uncertainty is the norm in situations like the one we study here—in which the outcomes are manifold and difficult to assess. The resulting failure to make use of distant knowledge can be explained by organizations' difficulty in processing such knowledge. It is challenging for organizations to assess the validity and potential of distant knowledge. Moreover, it is also challenging for organizations to act on distant knowledge, since distant suggestions represent only the potential for future innovations, which will require significant investment to be realized (Kotha et al., 2013; Thursby & Thursby, 2002). For example, given its background in geoengineering, it would be difficult for BP to assess and implement a suggestion regarding genetically modifying bacteria to foster the breakdown of spilled oil. It is thus particularly challenging for organizations to make good use of distant domains of knowledge. As a result, coping with distant knowledge is often more costly for organizations than coping with familiar knowledge is, and distant knowledge is therefore more likely to be avoided altogether (Criscuolo, Haas, & George, 2013). In sum, organizations often fail to attend to distant knowledge, but it is unclear whether such a tendency holds for organizations that engage in distant search with the explicit intention of gaining access to distant knowledge—and whether crowding renders organizations more or less likely to pay attention to such knowledge.

## Hypotheses

We hypothesize that organizations generally filter out suggestions that capture distant knowledge—a tendency that is further amplified by crowding. We suggest that organizations engage in two practices when filtering suggestions. First, they *simplify* the filtering. March and Simon (1958) suggested that filtering "is simplified by omitting some criteria and paying particular attention to others." Evans and Foster (2011) showed how researchers simplify their processing of scientific research by focusing on meta-information, such as author, journal source, etc. In the same way, organizations are likely to focus on meta-information, such as the originating contributor or the headline, when processing suggestions. However, suggestions capturing distant knowledge are unlikely to include any meta-information that would be meaningful to a focal organization. For example, the suggestion's headline is likely to be difficult to understand.

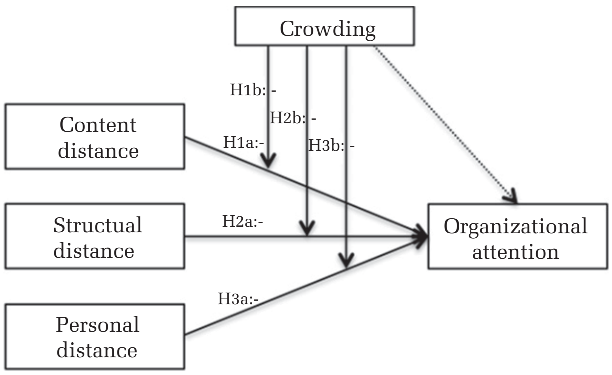
Second, organizations *rationalize* the filtering of knowledge. Rationalizing the filtering process might involve defining criteria for how to filter suggestions (e.g., focusing on suggestions in such a way that allows an organization to benefit from economies of scale and scope when attending to them). Rationalizing might also involve changes in the organization of the filtering process, such as establishing gates for different stages of the process (Cooper, 1985, 2008; Salter et al. forthcoming) or committees with voting rules (Knudsen & Levinthal, 2007). Such rationalization methods prevent an organization from engaging with distant knowledge, which often requires a certain "foolishness" (March, 2006). For example, the very fact that the knowledge is distant nearly always excludes the realization of economies of scale and scope, since a distant suggestion is rarely strongly related to other suggestions.

We argue that crowding amplifies organizations' tendency to simplify and rationalize—and, thus, to filter out suggestions representing distant knowledge. As crowding occurs, organizations look for ways to handle a larger number of suggestions. Research has shown that, while crowding increases the number of processed issues in absolute terms, it also reduces the share of processed issues (e.g., Sullivan, 2010). To process more suggestions while staying within their attention limits, organizations need to increase their processing efficiency. As crowding increases, organizations tend to simplify and rationalize, thus narrowing their attention.

We develop our core argument—that organizations filter out distant knowledge and that crowding further narrows organizations' attention (i.e., it allows less attention for suggestions that are distant)—by examining the effects of distance along multiple dimensions. We examine suggestions that are distant in terms of *content* (i.e., the actual description of a suggestion), *structure* (i.e., the types of contributors supporting it), and *person* (i.e., the contributor from whom it originates). Developing our core hypotheses along three dimensions illustrates the generalizability and robustness of our claim that crowding narrows attention. For each dimension of distance and its related hypothesis, our argumentation includes four components. First, we define each dimension of distance. Second, we explain why organizations often search for knowledge that is distant in terms of the focal type of distance. Third, we explain why organizations, despite their efforts to engage in distant search, are less likely to pay attention to suggestions characterized by the focal type of distance. Figure 1 summarizes our conceptual model. An added contribution of our conceptual model is that separating the three types of distance allows us to compare the types in terms of salience and their independent effects on organizations' attention.

***Content distance.*** We define content distance as the extent to which the content of a suggestion diverges from the content of suggestions to which an organization has previously paid attention. Content distance thus constitutes the reverse of similarity, in that two suggestions are distant if there is little (or no) similarity between their contents.

**FIGURE 1**
**Conceptual Framework**




*Note*: The solid lines illustrate the hypothesized effects. The dashed line represents the main effects, about which we do not directly theorize. We expect the three forms of distance (content, structural, and personal) to (a) have negative direct effects and (b) have their effects altered by crowding.

Gaining access to suggestions that are distant in terms of content is one of the main reasons organizations engage in distant search. Suggestions that are distant in terms of content have the potential to increase organizations' innovativeness. For example, the most innovative solution to the so-called "longitude problem"—namely, the marine chronometer—was different in terms of content from all previous suggestions because it applied the principles of clockwork, where the vast majority of other suggestions had applied astronomy (Sobel, 2007). Soliciting suggestions from external contributors is thus a fertile avenue through which to gain access to suggestions that are distant in terms of content. For example, Shane (2000) described how MIT's Technology Office solicited suggestions from external contributors that were distant in terms of content when looking for ways to deploy its three-dimensional printing technology. Organizations that solicit suggestions from external contributors often strive to and succeed in gaining access to suggestions that are distant in terms of content, which increases the variance in their choice set of possible avenues of action—thus, potentially increasing their innovativeness.

However, we suggest that organizations tend to filter out suggestions that are distant in terms of content once they are elicited—a tendency that is further amplified by crowding. When organizations filter, they tend to simplify, which is likely to steer attention away from suggestions that are distant in terms of content. Such suggestions are challenging to associate with any prior suggestions and difficult to categorize, since the receiving organization is unlikely to have received any similar suggestions. Distant suggestions are, by definition, not similar and, thus, fail to fall into any of the categories that organizations form when processing suggestions. Rationalizing the filtering also works against suggestions that are distant in terms of content because such suggestions offer little, if any, economies of scale and scope. When an invention combines knowledge from adjacent domains, the combination is likely to be more straightforward than the combination of knowledge from distant domains (Kotha et al., 2013). In fact, history is rife with examples in which organizations have filtered out distant content. For example, in the case of the previously discussed longitude problem, the scientific committee authorized by the British Parliament to administer an open call for solutions did not welcome the suggestion by the clockmaker John Harrison; in fact, it worked actively to exclude it. Harrison had to battle for years to receive acknowledgment for his suggestion. From examples like this one, we expect that organizations are less likely to pay attention to suggestions in which the content is distant from work they have previously undertaken. Thus, despite organizations' original ambition (i.e., to seek suggestions representing knowledge that is distant in terms of content), we suggest that organizations often employ a *content-based filter*, which is likely to discriminate against suggestions that are distant in terms of content—and that crowding further amplifies this tendency.

*Hypothesis 1. The effect of content distance on organizations' attention is (a) negative and (b) amplified by an increase in crowding.*

***Structural distance.*** We define structural distance as the extent to which a focal suggestion lacks links to other suggestions via the support of external contributors. Two suggestions are structurally distant if no (or only a few) external contributors support both of them. It is important to note that structural distance differs from content distance (Rodan & Galunic, 2004). Recent research has revealed the difference and emphasized the need to differentiate between structure and content, as it can be misleading to use structure as an indicator of content (Aral & Alstyne, 2011; Rodan & Galunic, 2004). It is therefore important to disentangle the two types of distance. Just because external contributors share similar views (i.e., their suggestions are structurally close) does not mean that the underlying content of their suggestions is closely related. For example, in U.S. politics, measures to restrict access to abortion and reduce income taxes are structurally close because they are often supported by the same people, but they are distant in terms of content. Thus, suggestions might be distant in terms of content, but not in terms of structure (and vice versa).

Gaining access to suggestions that are distant in terms of structure is one of the main reasons organizations engage in distant search. Suggestions that are distant in terms of structure have the potential to increase organizations' innovativeness. The link between structural distance and innovation has been well established (Granovetter, 1973; Hargadon & Sutton, 1997; Sgourev, 2013). Contributors on the periphery can work without feeling pressure to conform. Indeed, "marginal men" may be more likely to contribute path-breaking ideas (Merton, 1972), since they can think outside of established ways of thinking (Cattani & Ferriani, 2008). Being at the periphery may be so critical to contributors' innovativeness that they actually take on risks and make an effort to stay there (Salter, Criscuolo, & ter Wal, 2014).

However, we suggest that organizations tend to filter out structurally different suggestions once they are elicited—a tendency that is further amplified by crowding. Organizations simplify their filtering insofar as they forego trying to develop an overview and instead move sequentially from suggestion to suggestion. Consider the challenge of putting a puzzle together: If it is difficult to get an overview of all of the pieces, the player starts with a single piece and looks for a second piece that fits in the same area. In the same way, it is easier for organizations to pay attention to suggestions that are structurally related. This focus on structurally related suggestions is amplified by rationalizing, since addressing structurally related suggestions is likely to be more efficient than addressing structurally distant ones. Consider the example of a software developer that provides multiple versions of its software for different operating systems (e.g., Microsoft Windows and Apple OS). If the developer receives suggestions from users of both operating systems, the suggestions regarding each version are likely to be structurally related (i.e., users of one version are likely to support suggestions regarding the same version, but not suggestions regarding the other version, since users are unlikely to use both operating systems). When attending to a suggestion regarding one version, it is efficient for the software developer to also attend to other suggestions concerning the same version (i.e., other structurally related suggestions). Thus, despite organizations' original ambition to seek suggestions that capture knowledge that is distant in terms of structure, we suggest that organizations employ a *structure-based filter* that is likely to discriminate against suggestions that are structurally distant—and that crowding further amplifies this tendency.

*Hypothesis 2. The effect of structural distance on organizations' attention is (a) negative and (b) amplified by an increase in crowding.*

***Personal distance.*** We define personal distance as the extent to which the external contributor of a suggestion has previously received attention from the organization. A suggestion can thus be personally distant either because the external contributor has not made any suggestions in the past or because the organization has not paid attention to past suggestions. Note that a suggestion can be personally distant without being distant in either content or structure. For example, a contributor might make a suggestion for the first time in response to BP's call for solutions to the *Deepwater Horizon* accident. This contributor's first suggestion may be similar to other suggestions in terms of content (i.e., it may not be distant in terms of content) and might be supported by other external contributors who answered BP's call (i.e., it may not be distant in terms of structure).

Organizations engage in distant search with the particular intention of hearing from external contributors from whom they have not heard before. Suggestions from such contributors have the potential to increase organizations' innovativeness. For example, scholars of disruptive innovation have described how organizations fail to innovate because they continuously hear from the same set of customers (Adner & Levinthal, 2001; Christensen, 1997). Soliciting suggestions from external contributors is a way for organizations to hear from contributors that they have not yet heard from. These arguments propose that suggestions that are personally distant provide great innovation potential, and that, for this reason, organizations actively seek such suggestions.

However, we suggest that organizations tend to unintentionally filter out such suggestions once they are elicited—a tendency that is further amplified by crowding. Organizations simplify their filtering by focusing on meta-information, such as the submitting external contributor. Organizations are less likely to pay attention to suggestions when they do not know the originating contributor. The focus on suggestions from contributors with which the organizations have interacted before can also be rationalized. Prior interaction increases trust, fosters feelings of ease, and reduces the objectivity of evaluations (Lawler, 1992; Zajonc, 1968). That is, we think more highly of our close relations than we do of those with strangers. When an organization and a suggestion maker have engaged previously, the organization perceives the suggestion as having better quality. Organizations tend to allocate their attention to such relationships—even if these benefits are only perceived (Sorenson & Waguespack, 2006). For example, organizations are more likely to accept suggestions from employees or departments with whom or which they frequently interact (Hansen, 1999; Reitzig & Sorenson, 2013). Thus, despite organizations' original ambition to seek suggestions from personally distant contributors, we suggest that they employ a *person-based filter* that is likely to discriminate against suggestions that are distant in terms of structure—and that crowding amplify the negative effect.

*Hypothesis 3. The effect of personal distance on organizations' attention is (a) negative and (b) amplified by an increase in crowding.*

## METHODS

### Data

Our research question calls for a setting in which it is possible to trace the complete set of suggestions submitted to an organization, and not just those suggestions to which the organization ultimately pays attention. We thus built a large, unique, longitudinal dataset of suggestions submitted to a large number of organizations by numerous external contributors (see also Dahlander & Piezunka, 2014). We built this database in collaboration with a private software company that produces a software tool that organizations can embed in their websites to gather suggestions. The company is located on the west coast of the United States, and most of its customers are located in America and Western Europe. Its clients span various types of organizations across different industries, although the clientele is dominated by web companies (below, we explain how we account for this). The tool works much like a traditional suggestion box, allowing external contributors to submit their suggestions via a text field on a company's website. The default text that organizations use to encourage suggestions is: "I suggest to you..." In addition to submitting suggestions, external contributors can also vote and comment on suggestions made by others. This flexibility allows organizations to benefit from external contributors' new suggestions while simultaneously allowing external contributors to rank the importance of suggestions that have already been made. Our dataset included all information stored by the company. It provided us with full information on the actual suggestions, including each suggestion's full content; all comments on each suggestion; the identity of each external contributor who made, voted for, or commented on each suggestion; the accompanying conversations; the date of each interaction; and whether the organization eventually paid attention to (i.e., invested time and effort into) each suggestion.

Earlier work has also used the submission of suggestions as an empirical setting, but it has focused primarily on single organizations. For example, Bayus (2013) studied the Dell Idea Storm platform, and Reitzig and Sorenson (2013) examined the submission of suggestions to a European car manufacturer. In addition to increasing generalizability, a dataset that includes several organizations allows us to theorize about and exploit differences across organizations. The sample of organizations we examined received 1,077,669 suggestions between November 2007 and June 2011. The private company from which we retrieved the data launched in November 2007, so there is no risk of left censoring. We traced suggestions from organizations' first uses of the tool until the summer of 2011, when our data end. Suggestions to which the organizations accorded no action were considered right censored and were accounted for in our statistical approach. Our data thus offered the opportunity to longitudinally trace how organizations respond to suggestions, while accounting for both left and right censoring.

We cleaned the data in several ways. First, we removed all messages that the software tool coded as spam. Second, we removed all blank contributions (which, presumably, occurred because external contributors mistakenly submitted empty suggestions). Third, we focused on only those suggestions written in English, which was necessary in order for us to understand the suggestions' content and to merge the data with the corpora of positive and negative sentiments. Fourth, we removed suggestions contributed by organizations to their own forums (some organizations use this practice in an attempt to influence external contributors). Finally, we limited our dataset to those organizations for which we were able to match organizational characteristics with company information from ZoomInfo and CrunchBase. This limited our dataset considerably, but it also allowed us to implement more organization-level controls. In addition, limiting the dataset was necessary because some of the measures we constructed (such as the content measure, which we explain in greater detail below) were very time consuming to calculate. These precautions limited our dataset to 105,127 suggestions posted to 922 different organizations.

We supplemented these data with more qualitative information. We participated in several forums on our own by submitting suggestions, votes, and comments; we set up a tool ourselves and collected our own suggestions in order to better understand how the process works from an organization's perspective; and we conducted several informal interviews with the company that produces the tool in order to assess whether our readings were consistent with their understanding. These steps provided us with a better contextual awareness of how organizations use the tool to gain suggestions from outside of their organizational boundaries.

To further understand the organizations that used the tool, we gathered additional information about their characteristics. We considered several data sources before choosing ZoomInfo and TechCrunch's CrunchBase, which are websites that collect and aggregate detailed company information from other websites, such as company pages and news articles. Both databases typically focus on high-tech companies in the United States. ZoomInfo offered a relatively high number of matches to the organizations in our sample, as well as current information and a credible reputation among organizational scholars (Arora & Nandkumar, 2011; Graffin, Wade, Porac, & McNamee, 2008). According to its website, ZoomInfo "provides millions of just-verified, in-depth profiles on over 5 million business and 50 million employees" (ZoomInfo, 2012). Similarly, CrunchBase collects information from a vast number of websites. We parsed these two databases and matched their data with the organizations in our sample, giving us basic information about each company's size, geographic location, industry, etc. This strategy allowed us to gain access to additional data on a subset of organizations covered by both ZoomInfo and CrunchBase. We split the data in different ways to assess the strength of our inferences; these tests are detailed in the robustness checks section below.

### Dependent Variable

***Attention from the organization.*** Using the tool, organizations can assign certain statuses to suggestions based on whether the suggestions are selected. We developed a different dummy variables, indexed by $t$, which measured whether a suggestion received attention from an organization in calendar week $t$. We treated an organization's assignment of the status selected to a suggestion as attention from the organization. A suggestion left our risk set once it received a status of "selected."

### Independent Variables

***Crowding.*** Our basic premise is that crowding decreases the likelihood of attention and—more importantly for theorizing—changes the relative importance that organizations place on other cues. To measure crowding, we recorded how many suggestions were at risk (i.e., had not yet experienced a status change) in calendar week $t$. The degree of crowding varied with the number of new suggestions being posted (i.e., more suggestions increased crowding) and with an organization's completion and rejection of suggestions (i.e., more completions or rejections decreased crowding by decreasing the number of suggestions awaiting attention).

***Content distance.*** We used approaches developed in information retrieval to handle the content of the suggestions (for a similar approach, see Aral & Alstyne, 2011; Haas et al., 2014). We developed the programming in the statistical package R using the libraries *tm* and *SnowballC*. First, we processed the content of the suggestions by converting all words into lowercase and by removing all punctuation, white space, and non-alphabetic characters. Second, we removed all "stop words," such as "if" and "when," which are so common in the English language that they have little informational content. Third, we stemmed the text by changing each word to its stem or root form (e.g., "fished," "fisher," and "fishes" all became "fish"). Stemming is commonly used to enable comparisons of content across different documents.

After conducting this basic word transformation, we created a corpus for each organization using the suggestions made in calendar week $t$. We represented each suggestion as a "bag of words" that is agnostic to word order. A vector of length $\boldsymbol{W}$ represented each suggestion, where $\boldsymbol{W}$ is the number of unique words. Following earlier research in information retrieval (Jurafsky & Martin, 2009), we wanted to place greater weight on words that were more unique to certain suggestions and lower weight on words that were common across suggestions. The intuition is that two suggestions are more similar if they share words that are less frequently used in an organization's overall corpus. Each cell in the vector is set to $\log(w + 1) * \log(1 + (\text{frequency of suggestions} / \text{frequency of suggestions with } w))$, where $\log(w + 1)$ is the log of the frequency of each word $w$ in a suggestion and the second multiplier is the inverse frequency of word $w$ in an organization's corpus. We combined these vectors into an $\boldsymbol{S} * \boldsymbol{W}$ suggestion–word matrix for each organization and week $t$, where $\boldsymbol{S}$ denotes suggestions. From this, we developed an $\boldsymbol{S} * \boldsymbol{S}$ matrix, in which each cell is the cosine similarity between each vector in the $\boldsymbol{S} * \boldsymbol{W}$ matrix in week $t$. The cosine similarity ranges from 0 to 1 in positive space (i.e., the space in which 1 denotes that two suggestions are completely similar). From each organization's corpus, we used subsets of the data to compare those combinations that were at risk during different time periods. We used the row sum of the $\boldsymbol{S} * \boldsymbol{S}$ matrix in week $t$ to get a measure of content similarity. Two suggestions are thus similar to the extent to which they share words, especially words that are used less frequently in an organization's corpus. This measure captures similarity, and our theorizing is about content distance, so we multiplied this value by negative one.

***Structural distance.*** Our data on suggestions and votes from external contributors created a two-mode network. We sought to construct measures of the relationships among suggestions based on whether the same external contributors voted on multiple suggestions. By converting the two-mode data to one-mode data, we constructed a network of how suggestions were related to other suggestions based on votes in an organization's community. The one-mode projections created a fully connected group of all suggestions linked to a given external contributor. However, it is possible to give ties different weights depending on the number of co-occurrences (i.e., the number of external contributors who voted on the same suggestions). Newman (2001) developed this reasoning in his work on scientific collaboration; his argument rested on the suggestion that ties between scientists are stronger when there are fewer co-authors on scientific papers. He proposed a measure in which he discounted for the number of co-authors using the formula:

$$w_{ij} = \sum\nolimits_{p} \frac{1}{N_p - 1}$$

Here, $w_{ij}$ denotes the weight of a tie between two scholars, and $N_p$ is the number of authors on paper $p$. In our context, this means that two suggestions that received votes from the same external contributor achieved a tie with a weight of 1. We discounted the effect of external contributors that voted for many different suggestions, since they contributed less to the weight of the tie linking two suggestions. We also recalculated the measure with unweighted networks and found similar results. We operationalized these measures using the *tnet* library of R and drawing on the Newman projection of two-mode networks (Newman, 2001). We calculated the networks for each organization on a weekly basis (which required us to develop thousands of networks to calculate the network measure used in the paper).

We proposed that suggestions that are connected to other suggestions are more likely to get attention. We defined structural distance as the extent to which two suggestions exhibit similar associations with external contributors. To measure the structural distance of a suggestion, we measured the weighted degree centrality of a given suggestion in week $t$ in the one-mode projection of the suggestion network. This essentially captures the number of suggestions to which a given suggestion is linked, normalized by the size of the community. Our theorizing is about structural distance, so we multiplied this value by negative one.

***Personal distance.*** External contributors that have previously interacted with an organization are more likely to trigger attention from the organization. Such individuals have often learned from experience how to frame suggestions to gain attention within an organization's particular environment. Prior experience also accounts for a "messenger bias" (Menon & Blount, 2003), which favors those individuals that have provided suggestions in the past. Indeed, organizations often tend to listen to a small group of external contributors (von Hippel, 1986). We measured the number of implemented suggestions that an external contributor contributed to the focal organization before submitting a focal suggestion, then multiplied this number by negative one.

### Controls

We also included several variables to control for alternative explanations arising from suggestion-, individual-, or organization-specific factors that could affect the likelihood that a suggestion attracted organizational attention. Suggestion-level controls were the length of suggestions, positive and negative sentiments in the suggestion, positive and negative sentiments in the comments, share of votes, and number of votes. Individual-level controls comprised tenure and anonymity. Organizational-level controls were the age of the forum, the size of the organization, its geographical location, and its industry.

***Length of suggestions.*** A suggestion's content must be thorough enough for an organization to credibly understand the suggestion. At the same time, a suggestion with too many words is time-consuming to digest. We captured these intuitions by measuring the number of words in each suggestion. It is plausible that there are decreasing returns to the number of words; that is, an increase in the number of words has a greater effect when there are fewer words than when there are more words. To test this, we included a squared effect for the number of words.

***Positive and negative sentiments in the suggestion.*** Suggestions with a positive tone are likely to receive more attention (Reitzig & Sorenson, 2013). To code a suggestion's tone, we extracted all words in the suggestion and matched them with a list of positive and negative words developed by Liu, Hu, and Cheng (2005), who identified a comprehensive list of both positive and negative words with different variants (including the most common misspellings). For each suggestion in our database, we thus developed the ratios of positive and negative words, respectively, to the overall number of words.

***Positive and negative sentiments in the comments.*** The comments that suggestions receive from other external contributors provide signals to an organization regarding whether the suggestions are popular. More than the simple number of comments, the sentiments in the comments potentially influence whether a suggestion will receive attention. We thus controlled for the positive and negative sentiments of comments for each suggestion, measured as the mean of the proportion of positive / negative words for a suggestion's comments in week *t*. We created this variable by analyzing the content of each comment and matching it to the database developed by Liu and colleagues (2005). We then constructed the proportion of positive / negative words for each comment and took the mean of this proportion in week *t*.

***Share of votes.*** We theorized about structural distance, but we also had to account for the fact that some suggestions are simply more popular than are others. The underlying logic of this assumption is that organizations use external contributors not only to generate suggestions, but also to identify those suggestions that are important to many external contributors (similar to a "wisdom-of-the-crowd" logic). We measured the share of total votes—out of an organization's complete set of active suggestions—that each focal suggestion received in week *t*. To accomplish this, we counted the cumulative number of votes on suggestion *s* in week *t* and divided this number by the organization's total cumulative votes in that week. We counted the share rather than the raw number because the organizations sorted suggestions by the number of votes; thus, one would expect suggestions with higher shares to be more likely to get selected. This variable allowed us to clearly identify the importance of suggestion centrality by controlling for the sheer number of votes received by each suggestion.

***Number of comments.*** We expected suggestions to often be under-specified (von Hippel, 1986), so that further comments on the suggestions would trigger organizations' attention. To assess this, we measured the cumulative number of comments from external contributors on a suggestion in week *t*.

***Tenure.*** With increased tenure in a suggestion community, individuals gain firsthand experience about "ways of doing things" (Brown & Duguid, 1991). This can affect their ability both to develop suggestions and to frame those suggestions in line with ongoing community debates (Hargadon & Douglas, 2001). We would expect individuals with longer tenure in a community to attract more attention when posting new suggestions, so we measured tenure as the number of days elapsed between an individual's first vote, comment, or suggestion in the forum and the day a focal comment was submitted.

***Anonymity.*** The tool allows individuals to submit their suggestions anonymously. There are various reasons for posting anonymously; for example, an external contributor might not want to be associated with a suggestion he or she believes to be provocative. We would expect organizations to be suspicious of suggestions made anonymously. We thus coded a dummy variable that takes the value of 1 if an external contributor was anonymous when proposing his or her suggestion.

***Age of the forum.*** Organizations have vastly different amounts of experience in terms of working with external contributors, and some organizations are thus more practiced in listening to their users. We would expect organizations that have become accustomed to using the suggestion tool to be more likely to pay attention to external contributors. We thus measure the number of days elapsed since an organization received its first suggestion via the suggestion tool.

***Size of the organization.*** The size of an organization may affect the likelihood that a suggestion receives attention, since it is plausible that smaller organizations have fewer resources to commit to sifting through suggestions. We accounted for this by using separate dummy variables for small organizations (fewer than 20 employees) and for larger organizations.

***Received VC funding.*** We included a dummy variable to capture whether an organization had received VC funding according to CrunchBase.

***Geographical location.*** We developed a dummy variable to represent whether an organization was based in the United States.

***Industry of the organization.*** There are broad differences across industries with respect to the likelihood of innovation by external contributors (Morrison, Roberts, & von Hippel, 2000). Some organizations may thus be more accustomed than others to listening to external contributors. To measure this effect, we created industry dummy variables using the broad industry classification scheme detailed on CrunchBase.

Table 1 summarizes the variables used in the study, their definitions, and the data sources used to construct them. It is worth noting that all of the independent variables are time variant and traced on a weekly basis after the initial suggestion.

### Estimation Strategy

Our dataset is longitudinal, and it traces each suggestion and its likelihood of receiving attention on a weekly basis. To test our hypotheses, we used Cox proportional hazard models that allowed us to make use of the continuous data at our disposal and to account for the fact that some suggestions were, at any given time, pending attention (and, thus, treated as right censored). The hazard rate for the *j*th subject is:

$$h(t|x_j) = h_0(t)\exp(x_j\beta_x)$$

where we estimate the $\beta_x$ coefficients. The Cox proportional hazard models are agnostic about the assumptions of the shape of the hazard function, as long as the shapes are the same for all subjects *j*.

We used the Cox proportional hazard models to estimate whether a suggestion would receive attention (i.e., would be selected by an organization). A suggestion entered the risk set at the time that it was initiated and left the risk set when it was selected by the organization. The day we observed our first suggestion was set to time 0. The Cox proportional hazard regressions have no intercept, since they are subsumed into the baseline hazard. The coefficients can be interpreted as the change in the hazards for a one-unit change in the underlying covariate.

Our hypotheses required us to make use of time-varying covariates. For instance, the amount of crowding a suggestion faces changes continuously, and we had to account for this over time. We thus split our dataset into weekly records (more finely grained data would be too computationally intensive, since we had to construct networks and measures of content similarity over time). The use of weekly data for our sample created approximately 5 million weekly observations.

## RESULTS

Table 2 illustrates the descriptive statistics, including means, standard deviations, and correlations. Correlations among variables are modest, which reduces the concern of multicollinearity. Figure 2 shows unconditional Kaplan–Meier survival estimates, illustrating that suggestions are more likely to get attention in their first weeks and that attention levels out after about ten weeks. Figure 2B scales the *y*-axis differently to illustrate this pattern more clearly.

Table 3 shows the results of the Cox proportional hazard regressions predicting the time elapsed before a suggestion receives attention from an organization. These models use Huber–White clustered standard errors at the individual level to account for heteroscedasticity. The control variables are grouped into suggestion-, individual-, and organization-level controls in the tables. Model 1 is the baseline model, with controls available at the suggestion, individual, and organizational levels. Model 2 adds the crowding variable. The main independent variables are added separately in Models 3, 4, and 5. We then combine the crowding variable and the main independent variables in Model 6 in order to compare the sizes of the coefficients of the main variables. We continue by assessing the interaction effects between the different measures of distance and crowding in separate regressions in Models 7, 8, and 9. Model 10 is the full model, showing all of the main and interaction effects together.

### The Baseline Effect of Crowding

A basic premise of our analysis is that crowding reduces the likelihood of any suggestion gaining

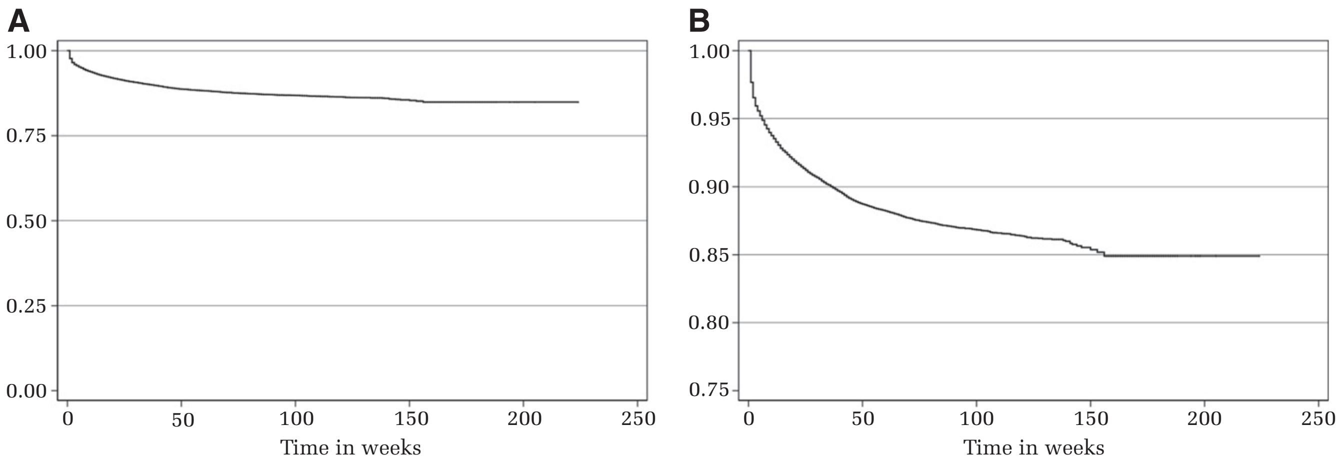
**TABLE 1**
**Description of Variables and Their Data Sources**

| Dependent variable | Definition | Data source |
|---|---|---|
| Selected | Dummy = 1 if the suggestion is implemented by the organization at time *t*. | Proprietary data source |
| **Suggestion-level control variables** | | |
| Suggestion length | Number of words in the suggestion. | Proprietary data source |
| Positive sentiments of suggestion | Proportion of positive words in the suggestion (i.e., the number of positive words / total number of words in the suggestion). | Proprietary data source |
| Negative sentiments of suggestion | Proportion of negative words in the suggestion (i.e., the number of negative words / total number of words in the suggestion). | Proprietary data source |
| Positive sentiments of comments on suggestion | Mean of the proportion of positive words in the comments on the suggestion at time *t*. | Proprietary data source |
| Negative sentiments of comments on suggestion | Mean of the proportion of negative words in the comments on the suggestion at time *t*. | Proprietary data source |
| Number of comments | Cumulative number of comments on the suggestion at time *t*. | Proprietary data source |
| Share of votes | Share of supporting votes in the community at time *t*, calculated as the cumulative number of votes on a suggestion divided by overall number of cumulative votes in the community. | Proprietary data source |
| **Individual-level control variables** | | |
| Tenure in community | Tenure in days of an individual since his or her first posting, vote, or comment at the time of posting the suggestion. | Proprietary data source |
| Anonymous | Dummy = 1 if an individual is anonymous when posting the suggestion. | Proprietary data source |
| **Organizational-level control variables** | | |
| Age of forum | Age of the forum in days at the time the suggestion is posted. | Proprietary data source |
| Size of organization | Dummy variables capturing different size categories for number of employees: small company (baseline) = 1–20 employees; larger organization = more than 20 employees. | CrunchBase and ZoomInfo |
| Received VC funding | Dummy = 1 if the organization has received a venture capital investment. | CrunchBase and ZoomInfo |
| Industry | Dummy variables capturing 19 different industries, such as biotech, clean tech, education, software, and web. | CrunchBase and ZoomInfo |
| U.S. location | Dummy = 1 if the organization is based in the United States. | CrunchBase and ZoomInfo |
| **Independent variables** | | |
| Crowding | Number of competing suggestions in the organization's community at time *t*. | Proprietary data source |
| Content distance | The row sum of the cosine similarity of a focal suggestion compared to all other previously accepted suggestions at time *t*, multiplied by −1. | Proprietary data source |
| Structural distance | Weighted degree centrality in the one-mode projection of the votes network for the organization's community at time *t*, multiplied by −1. | Proprietary data source |
| Personal distance | Number of previously selected or rejected suggestions posted by the individual *i* posting the focal suggestion, multiplied by −1. | Proprietary data source |

**TABLE 2**
**Descriptive Statistics**

| Variable | Mean | *SD* | Min. | Max. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Suggestion-level controls variables:** | | | | | | | | | | | | | | | | | | | | |
| 1 Suggestion length | 22.233 | 21.754 | 0 | 1138 | 1 | | | | | | | | | | | | | | | |
| 2 Positive sentiments of suggestion | 0.043 | 0.073 | 0 | 1 | −0.03 | 1 | | | | | | | | | | | | | | |
| 3 Negative sentiments of suggestion | 0.023 | 0.054 | 0 | 1 | 0.01 | −0.05 | 1 | | | | | | | | | | | | | |
| 4 Positive sentiments of comments on suggestion | 0.014 | 0.059 | 0 | 1 | 0.03 | 0.04 | −0.02 | 1 | | | | | | | | | | | | |
| 5 Negative sentiments of comments on suggestion | 0.006 | 0.036 | 0 | 1 | 0.02 | 0.01 | 0.05 | 0.09 | 1 | | | | | | | | | | | |
| 6 Share of votes | 0.012 | 0.056 | 0 | 1 | −0.04 | 0.01 | −0.03 | 0.03 | 0.00 | 1 | | | | | | | | | | |
| 7 Number of comments | 0.959 | 12.155 | 0 | 5410 | 0.02 | 0.00 | 0.00 | 0.06 | 0.05 | 0.05 | 1 | | | | | | | | | |
| **Individual-level control variables:** | | | | | | | | | | | | | | | | | | | | |
| 8 Suggestion maker is anonymous | 0.355 | 0.478 | 0 | 1 | −0.19 | 0.05 | 0.03 | −0.05 | −0.04 | 0.08 | −0.03 | 1 | | | | | | | | |
| 9 Tenure in community | 1.731 | 7.8 | 0 | 137 | 0.08 | −0.01 | −0.02 | 0.00 | 0.00 | −0.03 | 0.00 | −0.15 | 1 | | | | | | | |
| **Organizational-level controls variables:** | | | | | | | | | | | | | | | | | | | | |
| 10 U.S. location | 0.581 | 0.493 | 0 | 1 | −0.01 | 0.05 | −0.03 | 0.03 | 0.00 | 0.01 | 0.00 | 0.00 | 0.06 | 1 | | | | | | |
| 11 Small organization | 1.448 | 0.497 | 1 | 2 | 0.07 | −0.05 | 0.00 | −0.01 | 0.00 | −0.09 | 0.01 | −0.31 | 0.11 | 0.26 | 1 | | | | | |
| 12 Age of forum | 34.592 | 28.547 | 0 | 168 | 0.01 | 0.00 | 0.03 | −0.04 | −0.01 | −0.15 | −0.02 | −0.03 | 0.17 | −0.08 | −0.16 | 1 | | | | |
| **Independent variables:** | | | | | | | | | | | | | | | | | | | | |
| 13 Crowding | 1306.94 | 1517.1 | 1 | 4703 | 0.02 | −0.04 | 0.07 | −0.05 | 0.01 | −0.17 | 0.00 | −0.04 | −0.03 | −0.46 | −0.04 | 0.23 | 1 | | | |
| 14 Content distance | 0.008 | 0.944 | −6.73 | 4.65 | 0.35 | −0.02 | 0.01 | −0.02 | −0.02 | −0.01 | −0.02 | 0.07 | −0.02 | 0.00 | 0.00 | 0.02 | 0.00 | 1 | | |
| 15 Structural distance | −0.092 | 0.238 | −2 | 0 | 0.01 | −0.01 | 0.04 | −0.08 | −0.02 | −0.37 | −0.06 | −0.01 | −0.02 | −0.07 | 0.06 | 0.25 | 0.27 | −0.04 | 1 | |
| 16 Personal distance | −0.084 | 0.577 | −20 | 0 | −0.09 | 0.00 | 0.01 | −0.03 | −0.01 | 0.01 | 0.00 | 0.09 | −0.32 | 0.01 | 0.00 | 0.01 | 0.05 | −0.03 | −0.06 | 1 |

**FIGURE 2**
**Kaplan–Meier Survival Estimates**




*Note*: These graphs show the Kaplan–Meier survival estimates for suggestions selected by organizations. The left-hand graph (Fig. 2A) shows the Kaplan–Meier estimate with the hazard from 0 to 1, and the right-hand graph (Fig. 2B) shows the estimate for a narrower range of the hazard to illustrate the steeper decline in the first weeks. Organizations are more likely to pay attention to a suggestion in the first weeks after it is posted, and the curve flattens significantly after the first ten weeks.

attention. As a baseline, we expected that crowding would create more alternatives for organizations to choose among, making attention for any single suggestion less likely. Across our different estimations, we found a very consistent result: Crowding decreases the likelihood of attention. The results from Model 2 show that, when crowding increased by one standard deviation, the hazard of being selected decreased by 33% ($\exp(-.3965)$). Having established this baseline finding, we tested our hypotheses that crowding moderates the effects of the different types of distance above and beyond its main negative effect.

## Hypothesis 1

The first hypothesis suggested (a) that content distance would reduce the likelihood of attention and (b) that crowding would amplify this effect. More specifically, we expected suggestions that organizations had implemented in the past to be strong predictors of what would be selected in the future. We found strong support for claim (a), since the main effect was negative and significant ($-.0797$, $p < .01$). Lending support to (b), this negative effect increased with crowding, a result that was significant at the 1% level ($-0.0367$, $p < .01$). Comparing two suggestions when crowding was at its mean, an increase of one standard deviation in the content distance (holding all other variables constant) yielded a hazard equal to $\exp(-0.0797 * 1 -0.0347 * 1 * 0) = .92$, an 8% decrease in the hazard. In comparison, when crowding was at one standard deviation above its mean, the same shift yielded a hazard ratio equal to $\exp(-.0797 * 1 -0.0347 * 1 * 1) = .89$, an 11% decrease in the hazard. Thus, the effect of content distance on the hazard of a suggestion being selected was more negative when crowding was high.

## Hypothesis 2

Our second hypothesis captured the essence of structural distance: namely, the extent to which a focal suggestion is related to other suggestions. We proposed (a) that structural distance would reduce the likelihood of attention and (b) that crowding would amplify this effect. We found support for our (a) hypothesis (that structurally distant suggestions are less likely to receive attention): Model 4 shows that a one-standard-deviation increase in structural distance decreased the hazard by about 5% ($\exp(-.0526)$). Results from the final model show that crowding amplified the effect of structural distance ($-0.8506$, $p < .01$), providing support for our (b) hypothesis (that crowding further decreases the likelihood of organizations paying attention to structurally distant suggestions). Note that the main effect increases when the interaction effect is included and that the negative interaction effect is one

**TABLE 3**
**Cox Hazard Risk Models Predicting Selection**

| **Suggestion-level control variables:** | **Model 1** | **Model 2** | **Model 3** | **Model 4** | **Model 5** | **Model 6** | **Model 7** | **Model 8** | **Model 9** | **Model 10** |
|---|---|---|---|---|---|---|---|---|---|---|
| Suggestion length | 0.0320† | 0.0359* | −0.0026 | 0.0309† | 0.0228 | −0.0078 | −0.001 | 0.0341† | 0.0272 | −0.0104 |
| | (0.0174) | (0.0179) | (0.0179) | (0.0175) | (0.0172) | (0.0181) | (0.0182) | (0.0180) | (0.0176) | (0.0183) |
| Suggestion length squared | −0.0079 | −0.0086 | −0.0049 | −0.0079 | −0.0077 | −0.0054 | −0.0053 | −0.0085 | −0.0083 | −0.0051 |
| | (0.0053) | (0.0057) | (0.0045) | (0.0053) | (0.0052) | (0.0048) | (0.0048) | (0.0056) | (0.0055) | (0.0047) |
| Positive sentiments in suggestions | 0.0106 | 0.0078 | 0.0093 | 0.0113 | 0.0106 | 0.0069 | 0.0065 | 0.0078 | 0.0078 | 0.0062 |
| | (0.0094) | (0.0094) | (0.0095) | (0.0094) | (0.0093) | (0.0094) | (0.0095) | (0.0094) | (0.0094) | (0.0094) |
| Negative sentiments in suggestions | 0.0411** | 0.0475** | 0.0423** | 0.0422** | 0.0400** | 0.0485** | 0.0486** | 0.0473** | 0.0465** | 0.0475** |
| | (0.0081) | (0.0079) | (0.0082) | (0.0081) | (0.0081) | (0.0080) | (0.0080) | (0.0080) | (0.0079) | (0.0080) |
| Positive sentiments in the comments of suggestions | 0.1189** | 0.1181** | 0.1190** | 0.1178** | 0.1181** | 0.1165** | 0.1182** | 0.1152** | 0.1172** | 0.1145** |
| | (0.0048) | (0.0048) | (0.0048) | (0.0048) | (0.0049) | (0.0049) | (0.0048) | (0.0049) | (0.0049) | (0.0049) |
| Negative sentiments in the comments of suggestions | 0.0561** | 0.0590** | 0.0559** | 0.0559** | 0.0563** | 0.0588** | 0.0588** | 0.0570** | 0.0591** | 0.0569** |
| | (0.0047) | (0.0048) | (0.0047) | (0.0047) | (0.0047) | (0.0048) | (0.0048) | (0.0049) | (0.0048) | (0.0049) |
| Share of votes | 0.0687** | 0.0625** | 0.0675** | 0.0577** | 0.0711** | 0.0564** | 0.0612** | 0.0642** | 0.0644** | 0.0655** |
| | (0.0066) | (0.0068) | (0.0067) | (0.0070) | (0.0066) | (0.0071) | (0.0068) | (0.0069) | (0.0067) | (0.0069) |
| Number of comments | 0.0105** | 0.0122** | 0.0105** | 0.0102** | 0.0105** | 0.0119** | 0.0122** | −0.0072** | 0.0122** | −0.0069** |
| | (0.0019) | (0.0019) | (0.0019) | (0.0019) | (0.0019) | (0.0019) | (0.0019) | (0.0022) | (0.0019) | (0.0022) |
| **Individual-level control variables:** | | | | | | | | | | |
| Suggestion maker is anonymous | 0.0714** | 0.0492† | 0.0705** | 0.0763** | 0.0778** | 0.0588* | 0.0477† | 0.0799** | 0.0553* | 0.0835** |
| | (0.0262) | (0.0257) | (0.0262) | (0.0262) | (0.0259) | (0.0253) | (0.0257) | (0.0259) | (0.0254) | (0.0255) |
| Tenure in community | 0.0073 | −0.005 | 0.0078 | 0.004 | −0.0444* | −0.0561** | −0.0044 | −0.0176 | −0.0578** | −0.0663** |
| | (0.0182) | (0.0186) | (0.0182) | (0.0183) | (0.0202) | (0.0202) | (0.0186) | (0.0189) | (0.0201) | (0.0203) |
| **Organizational-level control variables:** | | | | | | | | | | |
| U.S. location | −0.1276** | −0.2529** | −0.1296** | −0.1285** | −0.1198** | −0.2457** | −0.2551** | −0.2778** | −0.2445** | −0.2719** |
| | (0.0243) | (0.0246) | (0.0243) | (0.0244) | (0.0244) | (0.0246) | (0.0246) | (0.0247) | (0.0245) | (0.0245) |
| Received VC funding | −0.1632** | 0.0093 | −0.1638** | −0.1480** | −0.1526** | 0.0246 | 0.0094 | 0.0510† | 0.0167 | 0.0571* |
| | (0.0264) | (0.0281) | (0.0264) | (0.0266) | (0.0265) | (0.0280) | (0.0281) | (0.0284) | (0.0279) | (0.0281) |
| Small organization | −0.4251** | −0.3282** | −0.4237** | −0.4125** | −0.4132** | −0.3097** | −0.3262** | −0.3650** | −0.3184** | −0.3537** |
| | (0.0262) | (0.0258) | (0.0261) | (0.0263) | (0.0258) | (0.0254) | (0.0258) | (0.0264) | (0.0253) | (0.0259) |
| Age of forum | −0.3027** | −0.2099** | −0.3003** | −0.2864** | −0.2951** | −0.1921** | −0.2075** | −0.1868**** | −0.2036** | −0.1798** |
| | (0.0125) | (0.0127) | (0.0125) | (0.0127) | (0.0125) | (0.0127) | (0.0126) | (0.0127) | (0.0126) | (0.0126) |
| Industry dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Independent variables:** | | | | | | | | | | |
| Crowding | | −0.3965** | | | | −0.3880** | −0.4000** | −0.0779** | −0.3808** | −0.0743* |
| | | (0.0193) | | | | (0.0190) | (0.0194) | (0.0285) | (0.0203) | (0.0291) |
| H1a: Content distance | | | −0.0672** | | | −0.0680** | −0.0811** | | | −0.0797** |
| | | | (0.0092) | | | (0.0091) | (0.0101) | | | (0.0102) |
| H2a: Structural distance | | | | −0.0526** | | −0.0349** | | −0.7144** | | −0.7013** |
| | | | | (0.0097) | | (0.0105) | | (0.0434) | | (0.0438) |
| H3a: Personal distance | | | | | −0.2901** | −0.2785** | | | −0.3828** | −0.3414** |
| | | | | | (0.0621) | (0.0618) | | | (0.0668) | (0.0698) |

**TABLE 3**
**(Continued)**

| Suggestion-level control variables: | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| H1b: Crowding × Content distance | | | | | | | −0.0347** | | | −0.0367** |
| | | | | | | | (0.0115) | | | (0.0118) |
| H2b: Crowding × Structural distance | | | | | | | | −0.8625** | | −0.8506** |
| | | | | | | | | (0.0539) | | (0.0541) |
| H3b: Crowding × Personal distance | | | | | | | | | −0.1694* | −0.1216 |
| | | | | | | | | | (0.0753) | (0.0820) |
| Wald $\chi^2$ | 3038.98 | 3603.09 | 3122.82 | 3100.85 | 3073.36 | 3732.61 | 3665.6 | 3943.89 | 3636.88 | 4013.9 |
| Number of subjects | 105127 | 105127 | 105127 | 105127 | 105127 | 105127 | 105127 | 105127 | 105127 | 105127 |
| Number of failures | 10629 | 10629 | 10629 | 10629 | 10629 | 10629 | 10629 | 10629 | 10629 | 10629 |
| Time at risk | 5039413 | 5039413 | 5039413 | 5039413 | 5039413 | 5039413 | 5039413 | 5039413 | 5039413 | 5039413 |

*Note:* Huber–White robust standard errors in brackets clustered at the individual level. Two-tailed tests. Standardized coefficients for continuous variables.
$^{+}$ $p < 0.1$
* $p < 0.05$
** $p < 0.01$

of great magnitude. Comparing two suggestions when crowding was at its mean, an increase of one standard deviation in structural distance (holding all other variables constant) yielded a hazard of exp (−0.7013 * 1 −0.8506 * 1 * 0) = .50, a 50% decrease in the hazard. In comparison, when crowding was at one standard deviation above its mean, the same shift yielded a hazard ratio of exp(−0.7013 * 1 - 0.8506 * 1 * 1) = .21, a 79% decrease in the hazard. In conclusion, the effect of structural distance is more negative when crowding is high.

### Hypothesis 3

Our third hypothesis proposed that personal distance from a suggestion maker makes an organization less likely to pay attention to his or her suggestion. We expected (a) that personal distance would reduce the likelihood of attention and (b) that crowding would amplify this effect. We found support for (a), in that personal distance decreased the likelihood of attention from the organization. Model 5 shows that a one-standard-deviation increase in personal distance decreased the hazard by about 25% (exp(−.2901)). While the interaction between personal distance and crowding is significant in Model 9, it loses its significance in the full Model 10. Against our reasoning in (b), we thus conclude that crowding does not amplify the effect of personal distance beyond its main negative effect. However, in several of the robustness checks, this effect is consistent with Hypothesis 3b.

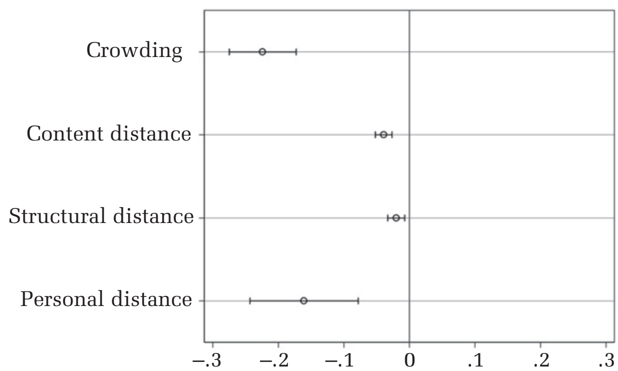
An added contribution of our paper stems from the fact that our distinction among different distances allowed us to assess and compare the magnitudes of the coefficients. Figure 3 displays the marginal effects with 95% confidence intervals using the results from Model 5. All marginal effects are statistically different from 0, but they differ in their magnitudes. In particular, the effect of personal distance is greater than the effects of content and structural distance.

### Supplementary Analyses and Robustness Checks

We developed several robustness checks to strengthen our inferences. Table 4 displays some of these checks.

First, we conducted a median split of the crowding variable by creating a dummy to assess whether the effects persist in this simpler model and to ease interpretation. We find consistent support for all hypotheses: All the main effects of distance are

**FIGURE 3**
**Effect Sizes of the Independent Variables**




*Note*: This figure illustrates the coefficients using the results from Model 5, which combines only the main effects of the independent variables. The circle represents the estimate of the coefficient, and the lines represent the 95% confidence intervals. Confidence intervals that overlap 0 (displayed by the vertical line in the figure) are insignificant.

negative and significant, and the interaction effects with the crowding dummy are also negative and significant.

Second, we re-estimated our models using the full sample of suggestions (about 550,000 suggestions), including suggestions made to organizations for which we lacked additional information. We estimated the models without the content distance variable (since this took too much computation time to estimate for the full sample). The full sample was more heterogeneous than our original sample, but the results were robust with respect to both the signs and the magnitudes of the coefficients.

Third, we wanted to ensure that our findings were present even in the absence of the many controls we included. Model 12 thus includes all of the independent variables and interactions without any of the controls. All of the coefficients remain significant and keep their expected signs, illustrating that our results are robust in a more parsimonious model.

Fourth, as we acknowledged earlier, our sample is heavily dominated by web companies, which may be more accustomed to involving externals than companies in other industries. The dominance of web companies was amplified by our decision to focus on organizations for which we were able to find additional sources of information from CrunchBase or ZoomInfo. We re-ran our analysis using only those companies that were Internet based and that operated in the high-tech arena, since it was plausible that removing non-web, non-high-tech organizations would affect our conclusions. Model 13 shows the results for web companies only. It is worth noting here that our sample was so web-dominated that Model 13 had nearly the same number of companies as the full sample (we removed 8,055 suggestions, bringing the total down to 97,072 in this model).

## DISCUSSION

Addressing our two research questions has led us to two main findings. With respect to our first research question, we find that organizations tend to filter out suggestions that capture distant knowledge. We have confirmed this effect across the three dimensions of distance: content, structural, and personal distance. With respect to our second research question, we find that crowding amplifies organizations' tendency to filter out suggestions that represent distant knowledge. Below, we elaborate on the implications of these findings.

### Attention: Crowding Narrows Attention

Our main contribution to the research on *attention* lies in the establishment of a new mechanism: *crowding narrowing attention*. Prior research on attention examines the criteria that guide organizations' attention (e.g., Haas et al., 2014; Hansen & Haas, 2001) and how crowding reduces organizations' attention to any single suggestion (Ocasio, 1997; Sullivan, 2010). Yet, earlier research neglects the interaction between these two effects. Studying this interaction, we find that crowding reduces organizations' attention to suggestions that are distant in terms of content and structure. In other words, rather than expanding their perspective, organizations may respond to crowding in the same way that human eyes respond to too much sun: They can't see through the glare without squinting. We expect the mechanism of crowding narrowing attention to hold in various organizational processes, including the generation of alternatives and their subsequent filtering (e.g., creating a pool of job applicants or conducting a call for proposals). The existence of too many alternatives might allow organizations to filter out the more distant alternatives.

Understanding that crowding narrows organizations' attention also informs our understanding of why too much information can decrease decision-making performance (O'Reilly, 1980). Stinchcombe (1990) asserted that organizations grow toward sources of crucial information in the same way that

**TABLE 4**
**Supplementary Analyses and Robustness Checks**

| Suggestion-level control variables: | Model 11 Median split of crowding | Model 12 With only independent variables | Model 13 Web companies |
|---|---|---|---|
| Suggestion length | −0.0144 | | −0.0014 |
| | (0.0177) | | (0.0244) |
| Suggestion length squared | −0.0046 | | −0.0115 |
| | (0.0045) | | (0.0088) |
| Positive sentiments in suggestions | 0.0066 | | 0.0043 |
| | (0.0094) | | (0.0101) |
| Negative sentiments in suggestions | 0.0464** | | 0.0463** |
| | (0.0081) | | (0.0084) |
| Positive sentiments in the comments of suggestions | 0.1122** | | 0.1158** |
| | (0.0049) | | (0.0052) |
| Negative sentiments in the comments of suggestions | 0.0578** | | 0.0515** |
| | (0.0048) | | (0.0053) |
| Share of votes | 0.0598** | | 0.0653** |
| | (0.0069) | | (0.0075) |
| Number of comments | 0.0065** | | −0.0070** |
| | (0.0019) | | (0.0021) |
| **Individual-level control variables:** | | | |
| Suggestion maker is anonymous | 0.0664** | | −0.0021 |
| | (0.0251) | | (0.0272) |
| Tenure in community | −0.0520** | | −0.0583** |
| | (0.0193) | | (0.0218) |
| **Organizational-level control variables:** | | | |
| U.S. location | −0.1914** | | −0.3126** |
| | (0.0244) | | (0.0259) |
| Received VC funding | −0.0276 | | −0.0018 |
| | (0.0277) | | (0.0301) |
| Small organization | −0.3322** | | −0.4065** |
| | (0.0262) | | (0.0274) |
| Age of forum | −0.2203** | | −0.1917** |
| | (0.0127) | | (0.0132) |
| Industry dummies | | | Yes |
| **Independent variables:** | | | |
| Crowding | | −0.3183** | −0.0619* |
| | | (0.0292) | (0.03) |
| Crowding above median dummy | −0.2198** | | |
| | (0.0323) | | |
| H1a: Content distance | −0.0570** | −0.0827** | −0.0821** |
| | (0.0107) | (0.0097) | (0.011) |
| H2a: Structural distance | −0.0341** | −0.4952** | −0.6842** |
| | (0.0105) | (0.0488) | (0.0454) |
| H3a: Personal distance | −0.2210** | −0.3236** | −0.3108** |
| | (0.0620) | (0.0634) | (0.0785) |
| H1b: Crowding × Content distance | | −0.0474** | −0.0403** |
| | | (0.0136) | (0.0121) |
| H1b: Crowding above median dummy × Content distance | −0.0595** | | |
| | (0.0174) | | |
| H2b: Crowding × Structural distance | | −0.4991** | −0.8320** |
| | | (0.0589) | (0.0559) |
| H2b: Crowding above median dummy × Structural distance | −0.4582** | | |
| | (0.0408) | | |
| H3b: Crowding × Personal distance | | −0.1465† | −0.1168 |
| | | (0.0794) | (0.0891) |
| H2b: Crowding above median dummy × Personal distance | −0.2573** | | |
| | (0.0739) | | |
| Wald $\chi^2$ | 3567.1 | 1218.55 | 3563.17 |
| Log likelihood | −117989.27 | −117793.75 | −105298.9 |

TABLE 4
(Continued)

| Suggestion-level control variables: | Model 11 Median split of crowding | Model 12 With only independent variables | Model 13 Web companies |
|---|---|---|---|
| Number of subjects | 105127 | 105127 | 97072 |
| Number of failures | 10629 | 10629 | 9570 |
| Time at risk | 5039413 | 5039413 | 4714277 |

*Note*: Huber–White robust standard errors in brackets clustered at the individual level. Two-tailed tests. Standardized coefficients for continuous variables.

† $p < 0.1$

* $p < 0.05$

** $p < 0.01$

plants grow toward the sun. However, collecting more information eventually keeps organizations from paying attention to knowledge with which they are not already familiar. So, by collecting too much information, an organization increases the chances that it will overlook knowledge that is novel and valuable but otherwise "hard to place."

### Search: Crowding Out Distant Knowledge

We show that even organizations with the particular intention of gaining access to distant knowledge by soliciting suggestions from external contributors may never attend to these suggestions. The organizations we studied strived to gain access to distant knowledge, but eventually filtered out such knowledge. The step of gaining access to distant knowledge must thus be considered separately from the step of paying attention to distant knowledge. The finding that crowding narrows attention strikes us as particularly unfortunate—and even ironic—given that those organizations that are exposed to crowding are among the most successful at soliciting suggestions from external contributors. However, it is these organizations in particular that fail to act on distant suggestions, precisely because they have too many suggestions to review. Crowding can thus paradoxically narrow organizations' attention in subsequent suggestion filtering and "crowd out" distant knowledge. Narrowing the focus of attention to the point at which distant knowledge gets insufficient attention—so that it effectively remains outside an organization's "field of vision"—reduces an organization's ability to innovate (Miller, 1993). We refer to this phenomenon as the "wide-lens trap." Adner (2012) introduced the analogy of a wide lens to illustrate the way in which organizations widen their search for innovation and draw on more diverse knowledge. While wide lenses have the physical characteristic of broadening the dimensions included, they simultaneously amplify differences between objects in the foreground and those in the background, so that things that are close seem even closer and things that are distant seem even more distant.

Organizations' failure to innovate is sometimes attributed to their failure to generate new knowledge. Our results suggest a potential alternative and a somewhat paradoxical explanation. It might be that organizations succeed in generating a particularly large amount of new knowledge, but that they fail to pay attention to the knowledge that has the most potential for innovation (Fleming, 2001; Fleming & Sorenson, 2004). Research provides numerous examples of how firms have "missed" ideas because there were too many ideas to pay attention to (e.g., the canonical case of Xerox's failure to commercialize many of the innovations developed by its PARC research center). We suggest that firms might fail to innovate not despite, but because of, their success in generating new knowledge, which eventually leads them to pay attention to more familiar knowledge.

Our findings have implications for the research on search-based competition (Greve & Taylor, 2000; Katila & Chen, 2008; Piezunka & Hannah, 2014). The organizations we studied solicited suggestions via their websites, which are public. This public format allows organizations' competitors to see contributors' suggestions as well as the organizations' responses. Competitors may use such insights to differentiate their search efforts from those of competing organizations (Greve & Taylor, 2000; Katila et al., 2012). Moreover, competitors might focus their attention particularly on those suggestions that have been filtered out by competing organizations. Competitors can thus benefit from the oversights of a competing organization (Piezunka &

Hannah, 2014). While organizations sometimes reveal such information intentionally for good reason (Alexy, George, & Salter, 2013; Pacheco-de-Almeida & Zemsky, 2012), in the case of crowdsourcing, such reveals would be unintended and uncontrolled. Taken together, organizations that solicit suggestions from external contributors via crowdsourcing might help their competitors achieve breakthrough innovations.

### Crowdsourcing: Lulled and Lured by the Crowds

Crowdsourcing constitutes a particular approach to sampling knowledge. For illustrative purposes, compare the following two knowledge-sourcing strategies. In *network-based knowledge sourcing*, an organization seeks suggestions from individual contributors in its network. For instance, von Hippel (1986) suggested that organizations focus on selected contributors. By seeking out specific individual contributors, organizations reduce the amount of redundant information they must process (i.e., they focus mostly on bridging structural holes, rather than on engaging all of their contributors) (Burt, 2004; Granovetter, 1973; Hansen, 1999). An organization thus has access to a small but very diverse set of suggestions with few redundancies.

In contrast, in *crowd-based knowledge sourcing*, the organization seeks information from all newcomers. Given the potential for larger amounts of sourced information (as compared to the results of network-based sourcing), crowd-based sourcing more often results in crowding and, thus, as we have showed, the narrowing of an organization's attention. The paradoxical result is that crowdsourcing can lead organizations to focus on only a subset of their contributors—those who share the same knowledge as the organizations. Organizations with greater access to external suggestions may therefore be no more innovative than peer firms whose approach to external search is more ambivalent. In other words, while organizations can learn from crowds, they can also be lured or lulled by them: lured into wasting attention on the process of discerning good ideas from bad, and lulled into believing that the ideas expressed most often or most loudly are also the best.

We also contribute to the literature on crowdsourcing by examining the stages that follow the sourcing of suggestions. We show that organizations might fail to harness the full potential of crowdsourcing due to inadequate filtering mechanisms. Our research thus underscores the necessity of adjusting various organizational processes when seeking to benefit from crowdsourcing (Foss, Laursen, & Pedersen, 2011). For example, rather than ignoring contributors to whose suggestions they cannot pay attention, organizations should benevolently reject them (Piezunka & Dahlander, 2015). We thus suggest that scholars of crowdsourcing examine the phases subsequent to the sourcing itself.

### Methodological Implications: Integration of Networks and Content

Distinguishing among content, structural, and personal distance also offers new methodological insights by linking methods on networks to research in information retrieval and computer science on content. This is important because network literature is often "content-agnostic": that is, it assumes that diverse networks provide access to diverse information (Aral & Alstyne, 2011). As Rodan and Galunic (2004: 545) note, "Network structure has been used as a proxy for information and knowledge heterogeneity, although the latter is never directly measured." Recent advancements in computer science and information retrieval have prompted scholars to think more deeply about the content being channeled through networks. By parsing content from structure, we have been able to show that the two forms of distance exhibit *separate* independent effects. The separation is important because two suggestions can be structurally close (i.e., they can have the same supporters) but be different in terms of what they propose.

### Managerial Implications

While crowding appears to have the unintended consequence of narrowing organizations' attention, managers can actively resist the tendency to narrow attention by taking steps at the stages of soliciting and filtering external suggestions. At the stage of generating alternatives, managers might constrain the number of alternatives that are generated. Prior research has often argued that generating more alternatives is superior and that organizations should always try to engage more external contributors, whereas our findings counter that organizations might benefit from casting a smaller net. For example, instead of mobilizing a large crowd, an organization might choose to rely on a smaller group of external contributors—ideally, contributors who are more likely to provide distant knowledge. Simple tactics, such as choosing a single contributor randomly (if the community is fragmented), could force an organization to work with a diverse array of contributors. By constraining the generation of alternatives, organizations can avoid the need to use strict filters to sift through the mass of suggestions, which has the unintended effect of narrowing the array of ideas that the firm ultimately considers.

In the stage during which organizations filter suggestions and decide which alternatives to pay attention to, an organization interested in the benefits of distant search will need procedures that better facilitate the processing of large quantities of suggestions. It may also be useful to define a goal ratio for how many familiar and distant suggestions an organization wants to implement to avoid the effect of one crowding out the other. An organization can also ensure that distant knowledge "survives" by establishing criteria through which it prioritizes suggestions that represent distant knowledge. For example, organizations might always begin by examining suggestions from external contributors from whom they have never heard.

### Limitations and Future Work

While we strongly believe in our findings, and have attempted to consider and rule out alternative explanations, it is important to be transparent about potential limitations. We believe our theory will hold in the presence of the following scope conditions: (a) all suggestions and ideas are visible, not only to the organization, but also to externals; and (b) externals can engage with any suggestion by voting, commenting, or offering refinements. These scope conditions are necessary because our theorizing rests on the cues an organization takes from its external contributors and from the relationships among these contributors. If the initiative is simply a suggestion box without social mechanisms, organizations cannot use relationship cues to cull their suggestions.

Crowding has another unintended consequence that is regularly overlooked: In addition to the few external contributors who receive positive responses from organizations, there may be legions of engaged but ignored (or rejected) external contributors. These individuals can potentially work against an organization by developing and expressing their negative views publicly, potentially tainting the organization's reputation. Prior research, which has focused on the upsides of crowding, has neglected these hidden costs. We suggest that future research theorizes more actively about such costs by studying how they may harm organizations' relationships with external contributors and endanger future search efforts to engage such contributors. Only then can researchers provide insights into how organizations can alleviate hidden costs and assess whether such costs outweigh potential benefits.

## REFERENCES


Adner, R. 2012. ***The wide lens: A new strategy for innovation***. New York, NY: Penguin Group.

Adner, R., & Levinthal, D. 2001. Demand heterogeneity and technology evolution: Implications for product and process innovation. ***Management Science***, 47: 611–628.

Afuah, A., & Tucci, C. 2012. Crowdsourcing as solution to distant search. ***Academy of Management Review***, 37: 355–375.

Ahuja, G., & Katila, R. 2001. Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. ***Strategic Management Journal***, 22: 197–220.

Alexy, O., George, G., & Salter, A. 2013. *Cui bono*? The selective revealing of knowledge and its implications for innovative activity. ***Academy of Management Review***, 38: 270–291.

Aral, S., & Alstyne, M. V. 2011. The diversity–bandwidth trade-off. ***American Journal of Sociology***, 117: 90–171.

Arora, A., & Nandkumar, A. 2011. Cash-out or flameout! Opportunity cost and entrepreneurial strategy: Theory and evidence from the information security industry. ***Management Science***, 57: 1844–1860.

Bayus, B. L. 2013. Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. ***Management Science***, 59: 226–244.

Boudreau, K., Guinan, E., Lakhani, K., & Riedl, C. Looking across and looking beyond the knowledge frontier: Intellectual distance and resource allocation in science. *Management Science* (forthcoming).

BP. 2010. Suggestion site closure. http://www.restorethegulf.gov/release/2010/09/27/suggestion-site-closure (accessed July 5, 2011).

Brown, J. S., & Duguid, P. 1991. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovating. ***Organization Science***, 2: 40–57.

Burt, R. S. 2004. Structural holes and good ideas. ***American Journal of Sociology***, 110: 349–399.

Cattani, G., & Ferriani, S. 2008. A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the Hollywood film industry. ***Organization Science***, 19: 824–844.

Chesbrough, H. 2003. ***Open innovation: The new imperative for creating and profiting from technology***. Boston, MA: Harvard Business School Press.

Christensen, C. M. 1997. ***The innovator's dilemma***. Boston, MA: Harvard Business School Press.

Cooper, R. G. 1985. Selecting winning new product projects: Using the NewProd system. ***Journal of Product Innovation Management***, 2: 34–44.

Cooper, R. G. 2008. Perspective: The Stage-Gate® idea-to-launch process: Update, what's new, and NexGen systems. ***Journal of Product Innovation Management***, 25: 213–232.

Cyert, R. M., & March, J. G. 1963. ***A behavioral theory of the firm***. Malden, MA: Blackwell.

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

Denrell, J. 2003. Vicarious learning, undersampling of failure, and the myths of management. ***Organization Science***, 14: 227–243.

Denrell, J. 2005. Selection bias and the perils of benchmarking. ***Harvard Business Review***, 83: 114–119.

Denrell, J., & Kovacs, B. 2008. Selective sampling of empirical settings in organizational studies. ***Administrative Science Quarterly***, 53: 109–144.

Evans, J. A., & Foster, J. G. 2011. Metaknowledge. ***Science***, 331: 721–725.

Fleming, L. 2001. Recombinant uncertainty in technological search. ***Management Science***, 47: 117–132.

Fleming, L., & Sorenson, O. 2004. Science as a map in technological search. ***Strategic Management Journal***, 25: 909–928.

Foss, N. J., Laursen, K., & Pedersen, T. 2011. Linking customer interaction and innovation: The mediating role of new organizational practices. ***Organization Science***, 22: 980–999.

George, G., Kotha, R., & Zheng, Y. 2008. Entry into insular domains: A longitudinal study of knowledge structuration and innovation in biotechnology firms. ***Journal of Management Studies***, 45: 1448–1474.

Girotra, K., Terwiesch, C., & Ulrich, K. T. 2010. Idea generation and the quality of the best idea. ***Management Science***, 56: 591–605.

Graffin, S. D., Wade, J. B., Porac, J. F., & McNamee, R. C. 2008. The impact of CEO status diffusion on the economic outcomes of other senior managers. ***Organization Science***, 19: 457–474.

Granovetter, M. S. 1973. The strength of weak ties. ***American Journal of Sociology***, 78: 1360–1380.

Greve, H., & Taylor, A. 2000. Innovations as catalysts for organizational change: Shifts in organizational cognition and search. ***Administrative Science Quarterly***, 45: 54–80.

Gruber, M., MacMillan, I., & Thompson, J. 2008. Look before you leap: Market opportunity identification in emerging technology firms. ***Management Science***, 54: 1652–1665.

Haas, M., Criscuolo, P., and George, G. 2014. Which problems to solve? Online knowledge sharing and attention allocation in organizations. http://amj.aom.org/content/early/2015/01/05/amj.2013.0263.abstract

Hansen, M. T. 1999. The search–transfer problem: The role of weak ties in sharing knowledge across organization subunits. ***Administrative Science Quarterly***, 44: 82–111.

Hansen, M. T., & Haas, M. R. 2001. Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. ***Administrative Science Quarterly***, 46: 1–28.

Hargadon, A., & Sutton, R. I. 1997. Technology brokering and innovation in a product development firm. ***Administrative Science Quarterly***, 42: 716–749.

Hargadon, A. B., & Douglas, Y. 2001. When innovations meet institutions: Edison and the design of electric light. ***Administrative Science Quarterly***, 46: 476–501.

Huber, G. P.,, & Daft, R. L.1987. The information environments of organizations. In F. M. Jablin, L. L. Putnam, K. H. Roberts & L. W. Porter (Eds.), ***Handbook of organizational communication: An interdisciplinary perspective***: 130–164. Newbury Park, CA: Sage Publications.

Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and problem-solving effectiveness in broadcast search. ***Organization Science***, 21: 1016–1033.

Jurafsky, D., & Martin, J. H. 2009. ***Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition***. Boston, MA: MIT Press.

Katila, R., & Ahuja, G. 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. ***Academy of Management Journal***, 45: 1183–1194.

Katila, R., & Chen, E. L. 2008. Effects of search timing on product innovation: The value of not being in sync with rivals. ***Administrative Science Quarterly***, 53: 593–625.

Katila, R., Chen, E. L., & Piezunka, H. 2012. All the right moves: How entrepreneurial firms compete effectively. ***Strategic Entrepreneurship Journal***, 6: 116–132.

Knudsen, T., & Levinthal, D. A. 2007. Two faces of search: Alternative generation and alternative evaluation. ***Organization Science***, 18: 39–54.

Kornish, L., & Ulrich, K. 2012. The importance of the raw idea in innovation: Testing the sow's ear hypothesis. ***Journal of Marketing Research***, 51: 14–26.

Kotha, R., George, G., & Srikanth, K. 2013. Bridging the mutual knowledge gap: Coordination and the commercialization of university science. ***Academy of Management Journal***, 56: 498–524.

Laursen, K., & Salter, A. 2006. Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. ***Strategic Management Journal***, 27: 131–150.

Laursen, K. 2012. Keep searching and you'll find: What do we know about variety creation through firms' search activities for innovation? ***Industrial and Corporate Change***, 21: 1181–1220.

Lawler, E. J. 1992. Affective attachments to nested groups: A choice-process theory. ***American Sociological Review***, 57: 327–339.

Li, Q., Maggitti, P. G., Smith, K. G., Tesluk, P. E., & Katila, R. 2013. Top management attention to innovation: The role of search selection and intensity in new product introductions. ***Academy of Management Journal***, 56: 893–916.

Liu, B., Hu, M., & Cheng, J.2005. Opinion observer: Analyzing and comparing opinions on the web. In A. Ellis (Ed.), ***Proceedings of the 14th international conference on World Wide Web***: 342–351. New York, NY: Association for Computing Machinery.

March, J. 2006. Rationality, foolishness, and adaptive intelligence. ***Strategic Management Journal***, 27: 201–214.

March, J. G., & Simon, H. A. 1958. ***Organizations***. New York, NY: John Wiley & Sons.

Menon, T., & Blount, S. 2003. The messenger bias: A relational model of knowledge valuation. ***Research in Organizational Behavior***, 25: 137–186.

Merton, R. K. 1972. Insiders and outsiders: A chapter in the sociology of knowledge. ***American Journal of Sociology***, 78: 9–47.

Miller, D. 1993. The architecture of simplicity. ***Academy of Management Review***, 18: 116–138.

Morrison, P., Roberts, J., & von Hippel, E. 2000. Determinants of user innovation and innovation sharing in a local market. ***Management Science***, 46: 1513–1527.

Mueller, J. S., Melwani, S., & Goncalo, J. A. 2012. The bias against creativity: Why people desire but reject creative ideas. ***Psychological Science***, 23: 13–17.

Newman, M. E. J. 2001. The structure of scientific collaboration networks. ***Proceedings of the National Academy of Sciences of the United States of America***, 98(2): 404–409.

O'Reilly, C. A., III, 1980. Individuals and information overload in organizations: Is more necessarily better? ***Academy of Management Journal***, 23: 684–696.

Ocasio, W. 1997. Towards an attention-based view of the firm. ***Strategic Management Journal***, 18(S1): 187–206.

Ocasio, W. 2010. Attention to attention. ***Organization Science***, 22: 1286–1296.

Pacheco-de-Almeida, G., & Zemsky, P. B. 2012. Some like it free: Innovators' strategic use of disclosure to slow down competition. ***Strategic Management Journal***, 33: 773–793.

Piezunka, H., & Dahlander, L. 2015. Benevolent rejections - Fostering engagement and effectiveness in search. Working Paper.

Piezunka, H., & Hannah, D. 2014. Getting ahead by falling behind—A set theoretic model of search-based competition. *Academy of Management Best Paper Proceedings*, January 2014: 13211.

Reitzig, M., & Sorenson, O. 2013. Biases in the selection stage of bottom-up strategy formulation. ***Strategic Management Journal***, 34: 782–799.

Robinson, A. G., & Stern, S. 1998. ***Corporate creativity: How innovation and improvement actually happen***. San Francisco, CA: Berrett-Koehler Publishers.

Rodan, S., & Galunic, C. 2004. More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness. ***Strategic Management Journal***, 25: 541–562.

Rosenkopf, L., & Nerkar, A. 2001. Beyond local search: Boundary-spanning, exploration, and impact in the optical disc industry. ***Strategic Management Journal***, 22: 287–306.

Salter, A., Criscuolo, P., & ter Wal, A. 2014. Going underground: Bootlegging and individual innovative performance. ***Organization Science***, 25: 1287–1305.

Salter, A., ter Wal, A., Criscuolo, P., & Alexy, O. 2014. Open for Ideation: Individual-level Openness and Idea Generation in R&D, forthcoming in *Journal of Product Innovation Management*.

Sgourev, S. V. 2013. How Paris gave rise to Cubism (and Picasso): Ambiguity and fragmentation in radical innovation. ***Organization Science***, 24: 1601–1617.

Shane, S. 2000. Prior knowledge and the discovery of entrepreneurial opportunities. ***Organization Science***, 11: 448–469.

Simon, H. A. 1947. ***Administrative behavior: A study of decision-making processes in administrative organizations*** (3rd ed.). New York, NY: Free Press.

Simons, R. 2010. Stress-test your strategy. ***Harvard Business Review***. 88: 92–100.

Singh, J., & Fleming, L. 2010. Lone inventors as sources of breakthroughs: Myth or reality? ***Management Science***, 56: 41–56.

Sobel, D. 2007. ***Longitude: The story of a lone genius who solved the greatest scientific problem of his time***. London, UK: HarperCollins.

Sorenson, O., & Waguespack, D. M. 2006. Social structure and exchange: Self-confirming dynamics in Hollywood. ***Administrative Science Quarterly***, 51: 560–589.

Stinchcombe, A. 1990. ***Information and organizations***. Berkeley, CA: University of California Press.

Stuart, T., & Podolny, J. 1996. Local search and the evolution of technological capabilities. ***Strategic Management Journal***, 17: 21–38.

Sullivan, B. N. 2010. Competition and beyond: Problems and attention allocation in the organizational rule-making process. ***Organization Science***, 21: 432–450.

Surowiecki, J. 2005. ***The wisdom of crowds***. New York, NY: Anchor Books.

Thursby, J. G., & Thursby, M. C. 2002. Who is selling the ivory tower? Sources of growth in university licensing. ***Management Science***, 48: 90–104.

Tzabbar, D. 2009. When does scientist recruitment affect technological repositioning? ***Academy of Management Journal***, 52: 873–896.

von Hippel, E. 1986. Lead users: A source of novel product concepts. ***Management Science***, 32: 791–805.

Zajonc, R. B. 1968. Attitudinal effects of mere exposure. ***Journal of Personality and Social Psychology***, 9: 1–27.

ZoomInfo. 2012. ***ZoomInfo database***. http://www.zoominfo.com/business/overview/database.html(accessed December 1 2012).

**Henning Piezunka** (henning.piezunka@insead.edu) is an assistant professor of entrepreneurship at INSEAD. He received his PhD in management science and engineering from Stanford University. His research focuses on search and competition.

**Linus Dahlander** (linus.dahlander@esmt.org) is an associate professor at ESMT European School of Management and Technology in Berlin, Germany, and the holder of the KPMG Chair in Innovation. His research focuses on networks, communities, and innovation. He received a PhD from Chalmers University of Technology in Gothenburg, Sweden, and was a postdoctoral fellow at Stanford University.