Attracting solutions in crowdsourcing contests: The role of knowledge distance, identity disclosure, and seeker status

Patrick Pollok, Dirk Lüttgens, Frank T. Piller

RWTH Aachen University, School of Business & Economics, TIME Research Area, Templergraben 55, 52056 Aachen, Germany

Abstract

We investigate in the context of crowdsourcing how seekers can increase open innovation performance, measured as received solver attention, by making two strategic decisions: selecting innovation tasks that are well suited for crowdsourcing and choosing between the potentials of status signalling through identity disclosure versus enjoying the benefits of anonymity. Drawing on uncertainty reduction theory, we suggest that a well-articulated problem statement reduces uncertainties of potential solvers and increases their willingness to participate. We argue that the ability of seekers to draft high-quality problem statements depends on the distance between the problem domain and their current knowledge stock. An analysis of 637 crowdsourcing projects finds that problem-seeker knowledge distance and received solver attention are curvilinear related such that moderate levels of knowledge distance maximize solver participation. However, high-status seekers who engage in identity-based status signalling are able to benefit from crowdsourcing across all levels of problem-seeker knowledge distance.

1. Introduction

In the last decade, crowdsourcing has gained relevance as a central mechanism of open innovation (Afuah and Tucci, 2012; Franke et al., 2013; Gong, 2017). Crowdsourcing initiatives are often administrated by specialized intermediaries (Dahlander and Piezunka, 2014; Diener and Piller, 2013; Lopez-Vega et al., 2016). In the domain of technical development and problem solving, platforms like NineSigma or InnoCentive have become attractive to firms seeking external inflows of knowledge for their innovation processes. A major challenge for seeker firms is to elicit attention from the solver population and to attract a sufficient number of solution proposals (Bayus, 2013; Dahlander and Piezunka, 2014; Girotra et al., 2010; Haas et al., 2015). Intermediaries support this process by engaging large established communities of potential contributors and providing dedicated communication infrastructure to effectively disseminate the technology needs of their clients (Felin and Zenger, 2014).

In this study we investigate how seekers’ decisions in the problem-drafting stage influence the confidence of potential solvers that they will succeed in a particular crowdsourcing challenge. We draw on uncertainty reduction theory (URT) (Berger, 1979; Berger and Calabrese, 1975) and signalling theory (Spence, 1973, 2002) to investigate factors influencing the perceived uncertainty of solvers (hence, lowering their confidence), which in turn will impact their level of participation. URT suggests that the initial interaction between two parties, who strive to build a collaborative relationship, is motivated primarily by the goal of reducing uncertainty (Berger and Calabrese, 1975). Communication plays a key role in this process, as it is through communication that uncertainty is reduced (Gibbs et al., 2011). In crowdsourcing, the core means of communication at this stage is the so-called “request for proposals (RFP)”, i.e. the problem statement drafted by the seeker. Using the RFP as the unit of analysis in this study, we argue that solvers are searching for signals and clues in the RFP document that can reduce their uncertainty whether it is worthwhile to submit a proposal or not. Seekers, in turn, can influence the perceived uncertainty of solvers by making deliberate decisions about the RFP. We investigate two of these decisions: the distance of the problem selected by a seeker for crowdsourcing and choosing between the potentials of status signalling through identity disclosure versus enjoying the benefits of anonymity.

This research extends other work on crowdsourcing that has examined how contest holders can influence participation and submission behavior of potential contributors by modifying specific factors in a crowdsourcing contest. These parameters include the amount and structure of the award (Boudreau et al., 2011; Erat and Krishnan, 2012),
the contest mode (Bockstedt et al., 2016), duration (Yang et al., 2010), or solver-pool size (Boudreau et al., 2011; Terwiesch and Xu, 2008). However, little attention has been paid so far to decisions about the crowdsourced problem and information revealed about the seeker organization. We address this gap by investigating how seeker firms can successfully reduce uncertainties perceived by solvers and increase their confidence that their efforts may pay-off when participating in a crowdsourcing contest. This would maximize the chances of capturing the attention of a large number of potential solvers. One choice in this regard refers to a prominent dilemma in crowdsourcing described as the “paradox of openness” (Laursen and Salter, 2014; Arora et al., 2016). In order to capture the attention of the crowd and to entice solution proposals, seekers have to reveal sufficient technical details about the problem itself and the opportunities that come with solving it (Alexy et al., 2013). At the same time, seekers have to be anxious not to divulge current technological deficiencies and future development trajectories to the public domain, current customers, and to potential competitors (Lopez-Vega et al., 2016). The dilemma here is that, if too little or the wrong information is disclosed in an RFP, potential solvers might not be willing to contribute and submit a solution.

Our paper makes a number of important contributions. First, we use URT to frame the relationship between seekers and solvers. In particular, we complement the understanding of “problem revealing” (Alexy et al., 2013) in innovation by explaining how seekers’ decisions about information disclosure in crowdsourcing affect the decisions of potential knowledge providers to submit a solution proposal in response to an open call for solutions. Secondly, we investigate how the quality of articulating the problem statement influences a seeker’s ability to attract solvers to a crowdsourcing project. We find that the quality of the articulation of the crowdsourcing problem depends on the distance between a seeker’s current (technological) knowledge stock and the knowledge domain of the selected problem (we use the term ‘problem-seeker knowledge distance’ for this relationship in the following). An analysis of 637 crowdsourcing projects hosted by NineSigma, a leading crowdsourcing intermediary, suggests that moderate levels of distance between the current knowledge stock of a seeker and the crowdsourced problem maximize the level of solver participation. We relate this finding to the (in)ability of seekers to formulate compelling problem statements suited to reducing the uncertainty of solvers in situations with too low or too great a knowledge distance. Thirdly, we contribute to the discussion of the benefits and challenges of revealing one’s identity in open innovation (Mahr et al., 2015). Complementing Alexy et al. (2013), we consider identity disclosure and status signalling as a form of revealing important contextual information. We find that this is an effective mechanism to attract solvers to a specific problem statement. Seekers with a high financial status appear as attractive collaboration partners and are able to benefit from crowdsourcing along all levels of problem-seeker knowledge distance when they reveal their identity in the RFP.

2. Theory and hypotheses

Uncertainty reduction theory (URT) (Berger and Calabrese, 1975) and arguments from signalling (Spence, 1973; Podolny, 1993) and status theory (Azoulay et al., 2014; Merton, 1968; Simcoe and Waguespack, 2011) guided the development of our conceptual model of the relationship between seekers’ (information) disclosure decisions and received solver attention (contest participation) in crowdsourcing. URT was originally formulated by Berger and Calabrese (1975) to examine initial interactions between strangers and has been applied to a variety of relationships in research on interpersonal and organizational communication as well as social and economic exchange relationships (Berger, 2011). According to URT, parties initiating a relationship strive to increase confidence by engaging in communication behavior that may take the form of interrogation or self-disclosure. The theory holds that the level of uncertainty for each party involved in the interaction decreases as the amount of communication between them increases (Ducaroz et al., 2016; Gibbs et al., 2011). Evidence in support of URT in contexts similar to our study of crowdsourcing has come from research on relationships in various online settings, where potential exchange partners evaluate each other based on limited informational cues available, including online dating websites (Gibbs et al., 2011), social and professional network sites (Toma, 2014), and e-commerce applications (Larrimore et al., 2011). For instance, Larrimore et al. (2011) study peer-to-peer lending where lenders make funding decisions based on the information provided in a loan request. They find that borrowers who disclose more information and details in their loan request are more likely to get funded. These borrowers are able to effectively reduce uncertainty and increase lender confidence in their intention to repay the loan. Similarly, Ma et al. (2017) in a study of the peer-to-peer sharing platform Airbnb, describe how hosts use self-disclosure to signal different underlying attributes that increase confidence by potential tenants, influencing their decision-making positively.

This uncertainty reduction mechanism is relevant for crowdsourcing of technical problems, too. We conceptualize crowdsourcing as a communication process between seekers and solvers, where seekers reveal problem information and contextual information to encourage potential collaboration partners (solvers) to initiate an exchange relationship and hence contribute to technical problem solving (Alexy et al., 2013; Lüttgens et al., 2014). Participation from a solver perspective implies allocation of attention to the crowdsourced problem by evaluating, investigating, drafting, and submitting a solution proposal. Attention allocation generally refers to focusing time and cognitive resources on a stimulus (Ashcraft, 1998; Kahneman, 1973; Haas et al., 2015). In the context of crowdsourcing, this stimulus is the request for proposals (RFP) document used to disseminate the R&D problem to be solved (Lopez-Vega et al., 2016). RFPs are non-confidential documents containing the problem statement, technological specifications, evaluation criteria, and information on project timing, business opportunities, and seeker firm (Feitler et al., 2012; Lüttgens et al., 2014). Solvers’ decision to participate and submit a solution proposal involves uncertainties that are related to the crowdsourcing problem (Boudreau et al., 2011; Schäfer et al., 2017). We restrict our investigation to these problem-related uncertainties since they are directly linked to information contained in the RFP documents and can be actively influenced by seekers’ information disclosure behavior. Following URT, we argue that the level of uncertainty perceived by solvers depends on the problem-related information seekers decide to reveal in the RFP document.

Beyond problem-related information, some seekers further decide to disclose contextual information in the RFP. In particular, seekers can choose whether they include their company’s identity in the RFP or not, which also signals their financial status. Drawing on status and signalling theory (Spence, 1973; Podolny, 1993), we argue that the relationship between problem-seeker knowledge distance and received solver attention is moderated by the identity disclosure decision and the financial status of the seeker firm issuing the RFP. In particular, we propose that the negative effects of either very low or very high problem-seeker knowledge distance on attention are weaker in RFPs from high-status seekers who engage in identity-based status signalling (Simcoe and Waguespack, 2011). Our conceptual model of solvers’ reactions to seekers’ information-revealing behavior in crowdsourcing is summarized in Fig. 1.

2.1. The relationship between problem-seeker knowledge distance and received solver attention

Over the last decade, an increasing number of studies has adopted a problem-solving perspective on innovative search to explain how search for solutions can effectively be governed depending on the type of innovation problem (Atuah and Tucci, 2012; Felin and Zenger, 2014; Macher, 2006; Nickerson et al., 2012; Nickerson and Zenger, 2004).
Defining the "problem" as the basic unit of analysis, this literature differentiates innovation problems by dimensions such as complexity, novelty, the hiddenness of solution knowledge, or the extent to which problems are well structured, to guide the choice between different strategies to match solutions to problems once the latter have been formulated. Problem formulation is seen as a cognitive process. The challenges associated with problem formulation are rooted in the assumption that problem holders (seekers) formulate their technology needs based on the technical knowledge they possess in the domains related to the problem (Baer et al., 2013). Hence, the extent to which a problem is clearly specified and comprehensively described (i.e., has a high articulation quality) can be expressed as a function of the problem-related knowledge available to the seeker.

We use the concept of problem-seeker knowledge distance to describe how the quality of the problem formulation in RFPs depends on seekers' expertise in the problem domain. Problem-seeker knowledge distance is the relatedness between the current technological knowledge base of the seeker firm and the knowledge domain of the RFP (Lopez-Vega et al., 2016; Piezunoka and Dahlander, 2015). The less the problem domain of the RFP is related to the technological knowledge base of the seeker firm, the larger the problem-seeker knowledge distance (distant problem). When, on the other hand, the knowledge domain of the RFP is closely related to a seeker's existing knowledge base, problem-seeker knowledge distance is low (local problem).

We suggest that this distance is linked to solver attention as it affects the quality of the problem articulation by the seeker, which will in turn influence the level of problem-related uncertainty perceived by potential solvers. Recent research on participation motives in crowdsourcing shows that solvers behave strategically and make their selection decision primarily based on self-interest and an assessment of the benefits they expect to obtain from participation (Bockstedt et al., 2016; Franke et al., 2013; Terwiesch and Xu, 2008). Especially uncertainties related to the crowdsourcing task itself affect solvers' participation behavior (Ales et al., 2017; Boudreau et al., 2011). When the perceived probability of winning becomes too small due to increased problem-related uncertainty, solvers are discouraged to participate (Terwiesch and Xu, 2008).

In this paper, we refer to problem related uncertainty as the degree of confidence solvers have regarding (1) the match between their own solution skills and the solution requirements formulated in the RFP (technical problem solving confidence), and (2) the match between their own perceptions and the seeker's perception of solution quality (evaluation confidence). We argue that technical problem solving confidence (increasing with problem-seeker knowledge distance) and evaluation confidence (decreasing with problem-seeker knowledge distance) are two latent mechanisms that shape the distance-attention relationship in crowdsourcing. In particular, we propose that they are two opposing (linear) functions interacting to generate an inverted U-shaped relation between problem-seeker knowledge distance and received solver attention (Ang, 2008; Haans et al., 2016):

At low levels of problem-seeker knowledge distance (local problems), problem statements are more likely to be over-specified. Seekers' existing knowledge in the problem domain can lead to tunnel vision (Franklin and CI, 2013) that drives them to restrict the solution space, allowing potential solvers to use only narrow sets of alternative solution approaches (Baer et al., 2013; Lopez-Vega et al., 2016). Further, when seekers possess sufficient expertise in the problem domain, they tend to have relatively high expectations regarding the optimal solution, which are reflected in rather strict specifications and high degrees of task difficulty (Erat and Krishnan, 2012; Lakhani et al., 2007; Lopez-Vega et al., 2016). This high precision in the RFP does not leave much room for creative technical solutions and poses the risk for solvers that they cannot meet all technical specifications listed in the document (Katok and Siemsen, 2011; Lopez-Vega et al., 2016). Hence, when solution criteria and specifications are too strict, there is a high likelihood that solvers' technical skills do not match the solution requirements formulated in the RFP. Hence, the technical problem solving confidence perceived by potential solvers is low. At the same time, however, such a highly specified RFP with strict but unambiguous solution criteria gives solvers a fairly good idea about the sort of solution the seeker is looking for (Erat and Krishnan, 2012). Hence, there is a low likelihood of a mismatch seekers' and solvers' quality perceptions, and potential solvers will perceive high levels of evaluation confidence.

When a seeker selects a problem for crowdsourcing that has a high level of problem-seeker knowledge distance, the seeker lacks knowledge and expertise in the area of this problem. Hence, we expect that the resulting RFP would impose less strict technical requirements and leave more room for creative solutions by potential solvers. Problem specificity will decrease. Such problem statements provide a larger solution space and appear to be technically more feasible (Baer et al., 2013; Gross and Sproull, 2004). With less restrictive problem statements, the likelihood that solvers' technical skills and abilities match the technical requirements that are listed in the RFP is high. Hence, solvers will perceive higher levels of technical problem solving confidence. This assumption is in line with literature on job choice intentions (Edwards, 1994; Jansen and Kristof-Brown, 2006; Resick et al., 2007; Saks and Ashforth, 1997), which suggests that job seekers make career decisions based on perceptions of how well their knowledge, skills, and abilities match the requirements formulated in a job description. This research found that the more detailed and specific the required abilities and skills of an applicant are outlined in the job description, the lower the number of job applications. Fewer candidates are confident that their abilities match the specific requirements formulated in the job opening (Mason and Belt, 1986; Roberson et al., 2005).

On the flipside, low problem specificity leaves potential solvers without a clear understanding what solution the seeker is exactly searching for. It provides few information about the evaluation criteria the seeker will employ to select the winning solution. When solvers perceive problem statements as fuzzy and incompletely specified with regards to solution requirements, the risk of a mismatch between the solver and seeker about what constitutes a perfect solution is high. Solvers perceive a low level of evaluation confidence. It is hard for them to assess upfront whether investing time and effort in solving the problem will finally pay off, and they will more likely refrain from submitting solutions (Bockstedt et al., 2016; Erat and Krishnan, 2012).

While technical problem-solving confidence will increase with problem-seeker knowledge distance and lead to higher attention, evaluation confidence will decrease with problem-seeker knowledge distance and lead to lower attention, and vice versa. The combined effect of these two opposing latent functions suggests a relationship between problem-seeker knowledge distance and received solver attention that takes the form of an inverted U-shape (Haans et al., 2016). Hence, we propose:

**Hypothesis 1.** Problem-seeker knowledge distance is curvilinear related to received solver attention (taking an inverted U-shape), with the highest level of solver attention occurring at an intermediate level of problem-seeker knowledge distance.
2.2. Identity disclosure decision: The moderating roles of identity information disclosure and seeker status

Along with the decision which problem to select for crowdsourcing and how to draft the corresponding RFP document, seekers also have to decide which contextual information they reveal in the RFP. In intermediary-administrated crowdsourcing, seeker firms can choose either to remain anonymous or to reveal their firm’s “identity”. Accordingly, an RFP document can also contain the seeker firm’s name and its industry (Figs. A1 and A2 in Appendix A provide an example). However, instead of making intensive use of the opportunity to provide additional contextual information in the RFP, the prevalent strategy among seekers seems to be to remain anonymous (Lüttgens et al., 2014; Mahr et al., 2015).

Choosing anonymity over identity disclosure can be a reasonable choice from the perspective of a seeker firm. Revealing a seeker firm’s identity informs its competitors of its development pipeline. At the same time, a public RFP also indicates to customers of a seeker that their supplier has open technical challenges. Also, potential solvers evaluating a crowdsourcing task may become biased either technically (a potential solver from the food industry may not consider its knowledge relevant for an aerospace company) or ethically (identifying a seeker as a defense company, for instance, may deter some solvers from submitting solutions). Hence, anonymity might increase the chances of receiving valuable solutions from distant knowledge fields. The opportunity to stay anonymous is actually one of the core reasons why seeker firms use the services of a crowdsourcing intermediary in the first place (Diener and Piller, 2013; Mahr et al., 2015).

A closer inspection of the consequences of staying anonymous, however, raises the question whether this concealing decision might affect solver attention in the opposite way. First, organizational research has recognized trustworthiness as an important issue in interfirm relationships (Gulati, 1995; Pavlou, 2002). If seekers remain anonymous, information asymmetries between seekers and solvers arise. Solvers have only limited information about the seeker’s true characteristics and intentions (Alexy et al., 2013; Silveira and Wright, 2010), but have to disclose valuable solution information and knowledge when participating in a contest. Disclosing the identity of a seeker serves here as an important signal of its trustworthiness, which we expect to affect the behavioral intentions of solvers to enter the (crowdsourcing) relationship (Colquitt et al., 2012; Larrimore et al., 2011; Toma, 2014).

Second, identity disclosure leads to a further important clue for solvers: the status of the seeker. RFPs are an invitation to collaborate (Alexy et al., 2013). In so doing, solvers will strive to partner with attractive seekers and will, all else being equal, prefer RFPs from these seekers over requests from less attractive potential partners. Using status as a proxy for partner attractiveness, literature on partnering patterns has shown that the value of a potential collaboration or exchange increases with partner status (Baum et al., 2005; Jensen, 2006). Status broadly refers to the prestige ascribed to individuals or organizations because of the position they occupy in a social hierarchy (Jensen and Roy, 2008; Piazza and Castellucci, 2014). Research focusing on the signalling value of status suggests that identity information serve as effective status signals (Higgins et al., 2011; Kovacs and Sharkey, 2014; Merton, 1968; Simcoe and Waguespack, 2011). This stream of literature provides empirical evidence regarding the effects of identity-based status signals on attention and performance. Simcoe and Waguespack (2011), for example, investigate the performance effects of identity-based status signals in an open standards community, the Internet Engineering Task Force (IETF). They find that IETF submission announcements attract significantly more attention and are more likely to get published when the name of a high-status author is visible (i.e., her identity is disclosed). This finding corresponds to the Matthew Effect (Merton, 1968), suggesting that high-status actors receive disproportionately more attention compared to low-status actors when their identity is known. Similarly, Higgins et al. (2011) present evidence that the names of star scientists who are associated with a firm are perceived as effective status signals by potential investors. Biotechnology firms with affiliated Nobel laureates are able to raise significantly more money in an IPO than firms without these star scientists. In the context of our study, we focus on status in financial terms, i.e., whether a seeker is seen as a particular financially attractive partner, as solvers aim at selling their technical solution to the seeker or entering a research collaboration.

Status literature further claims that the importance of status signals increases with the level of uncertainty attached to a specific outcome (Higgins et al., 2011; Podolny, 2005; Kovacs and Sharkey, 2014; Sharkey and Kovacs, 2017; Simcoe and Waguespack, 2011). Consistent with this view we suggest that status signals from high-status seeker firms will reduce the negative impact of problem-related uncertainty on solver participation, especially at the extreme ends of the knowledge distance spectrum, where the overall level of perceived problem-related uncertainty is high and confidence of solvers is low. In particular, we argue that at low levels of problem-seeker knowledge distance, potential solvers will be less deterred by strict technical specifications when the RFP comes from a financially attractive seeker. Despite the low levels of technical problem-solving confidence, seeker firms with a high status revealing their identity will appear more interesting to work with and will attract more submissions compared to their low status counterparts. Similarly, at high levels of problem-seeker knowledge distance, despite the inherent risk of a potential mismatch between solvers’ and seekers’ perceptions of solution quality (low evaluation confidence), solvers will more likely submit a solution to an RFP from a high status seeker in the hope that a firm with sufficient financial resources could potentially become a partner.

We thus suggest that the slopes of the inverted U-shaped relationship between problem-seeker knowledge distance and received solver attention are flatter for high-status seekers who engage in identity revealing than for low status seekers who disclose their identity in the RFP.

Hypothesis 2. The inverted U-shaped relationship between problem-seeker knowledge distance and received solver attention is moderated by seeker identity disclosure in the case of high status firms such that the relationship becomes flatter for high status firms as compared to low status firms.

3. Data and method

3.1. Data

The data we analyze in this paper was constructed from multiple sources:

1. Crowdsourcing project data. In collaboration with the crowdsourcing intermediary NineSigma, we created a unique dataset of project-level data containing detailed information about each crowdsourcing project hosted by NineSigma in the five-year period between 2009 and 2014 (e.g., number of solutions received, financial rewards, size and scope of the intermediaries’ email campaign). In this period of time, 238 seeker firms conducted 889 RFPs. Moreover, the project dataset provides information on the seeker firms holding the challenges (e.g., seeker firm name, location, industry). Appendix B provides a detailed description of NineSigma’s crowdsourcing process.

2. Patent data and firm-level secondary information. To determine seeker firms’ technological position, we use patent class data provided in Bureau van Dijk’s (BvD) Orbis database. The Orbis patent data originates from the European Patent Office’s PATSTAT database and classifies patents according to the international patent classification code (IPC). To account for some weaknesses of the secondary information captured in the intermediary’s project database, we
collected additional firm-level controls such as NACE industry codes, firm-size data, and financial information. We then matched secondary and patent data using firm BdD ID number as the unique firm identifier.

3 Technology need statements (RFP documents). We analyzed the content of the RFP documents of the respective projects and extracted information regarding solution requirements and seeker status. The initial data were cleaned and reduced in several ways. We restricted our initial sample of 889 RFPs conducted by 238 seeker firms between 2009 and 2014 to manufacturing firms (NACE category C) in order to remove potential 'noise' caused by RFPs of only limited technical nature, such as challenges hosted on behalf of seeker firms from the service sector. The remaining 701 RFPs were then classified according to three-digit IPC patent classes using the IPCCAT categorization system offered by the World Intellectual Property Organization (WIPO). The IPCCAT automated classification procedure is based on the Balanced Winnow algorithm and was initially designed to assist patent offices in classifying patent documents according to the IPC system (Foglia, 2007). 20 RFPs could not be classified in any particular patent class. Another 44 projects were eliminated due to missing patent and secondary information at the firm level. These criteria left us with a final sample of 637 crowdsourcing projects from 183 different firms.

3.2. Measurement

3.2.1. Dependent variable: received solver attention

We measure solver attention as the number of solutions an RFP is able to attract. As every submission exposes potential solvers to the uncertainties inherent in the respective RFP, the number of solutions received is a suitable measure to assess solvers’ immediate response to seekers’ information-revealing behavior in crowdsourcing. In addition, crowdsourcing literature suggests a positive quantity-quality association where the likelihood of high-quality solutions increases with the level of attention received from solvers in terms of solution quality (Bockstedt et al., 2016; Girotra et al., 2010). The number of solutions an RFP is able to attract can thus be regarded as a relevant performance measure in the crowdsourcing literature suggesting a positive quantity-quality association where the likelihood of high-quality solutions increases with the number of solutions an RFP is able to attract. As every submission exposes potential solvers to the seekers’ information-revealing behavior in crowdsourcing. In addition, crowdsourcing literature suggests a positive quantity-quality association where the likelihood of high-quality solutions increases with the level of attention received from solvers in terms of solution quality.

3.2.2. Independent variable: problem-seeker knowledge distance

Following a long tradition in the economics and strategy literature, we use patents as a proxy for knowledge (Colombelli and Quatraro, 2018; Gilsing et al., 2008; Rosenkopf and Almeida, 2003; Sampson, 2007). We rely on patent class-data from the International Patent Classification system (IPC) to measure the distance between the knowledge stock of a seeker and an RFP. The IPC represents the entire field of patents for invention (WIPO, 2018:3). This “technological knowledge space” is considered to be coherent and not a mere patchwork of isolated knowledge fragments (Geels, 2014; McNamee, 2013; Olsson and Frey, 2002; Sampson, 2007; Saviotti, 2007).

Like any classification system, however, the IPC system draws artificial boundaries between patent classes and treats every pair of patents as equally (un)related once they are categorized into different classes. McNamee, 2013; Moeen et al., 2013). The taxonomy hence assumes that e.g. class F03, which represents ‘Machines or engines for liquids; wind, spring, or weight motors; producing mechanical power or a reactive propulsive thrust, not otherwise provided for’, is equally (un) related to class F04, which describes “Positive-displacement machines for liquids; pumps for liquids or elastic fluids”, as it is to class A22, which represents “Butchering; meat treatment; processing poultry or fish”. This assumption may cause potential bias and measurement error when traditional distance measures are applied (Huo and Motobashi, 2014; McNamee, 2013). Previous research has therefore suggested to use relatedness adjusted distance measures that account for the potential variance in the relatedness among patent classes to more accurately capture knowledge distance (Alstott et al., 2017; McNamee, 2013). Following earlier work on industry-, product-, and technology-relatedness (Alstott et al., 2017; Breschi et al., 2003; Nesta, 2008; Quatraro, 2010; Teece et al., 1994), we build on a co-occurrence based procedure and calculate our measure of problem-seeker knowledge distance in two steps.

Step 1: For each possible pair of patent classes, we count the number of seeker firms holding patents in the two classes. Let $S_{ik} = 1$ if seeker firm $k$ holds at least one patent in class $i$ (where $i = 1, ..., 129$, represents the three-digit IPC class), and 0 otherwise. The number of seeker firms with patents in class $i$ is $n_i = \sum_k S_{ik}$, and the number of seeker firms patenting in both classes $i$ and $j$ is $Y_{ij} = \sum_k S_{ik} S_{jk}$. We then adjust the resulting $(129 \times 129)$ matrix of observed co-occurrences $Y_{ij}$ with the hypergeometric distribution of co-occurrences of patent classes that we would expect under the random hypothesis. The mean and the variance of this hypergeometric random variable $X_{ij}$ is as follows:

$$\mu_{ij} = E(x_{ij}) = \frac{n_i n_j}{K}$$

$$\sigma_{ij}^2 = \mu_{ij} \left(1 - \frac{n_i}{K} \right) \left(1 - \frac{n_j}{K}\right) / K$$

Based on the above, we derive the relatedness between each pair of patent classes as the difference between observed and expected co-occurrences and normalize the relatedness scores to range between 0 and 1.

$$\eta_{ij} = \frac{(Y_{ij} - \mu_{ij})}{\sigma_{ij}}$$

The resulting $(129 \times 129)$ matrix $M$ consists of the relatedness scores of each possible pair of patent classes. When observed numbers are higher than expected, we assume a non-random relationship between the two patent classes. Hence, relatedness scores with a value of zero indicate that the respective patent classes are independent, whereas values close to one indicate that two technology fields are highly related (Bryce and Winter, 2009; Teece et al., 1994).

Step 2: The relatedness matrix is then used to derive an adjusted distance measure based on the Euclidean distance between the vector of patent classes related to the RFP (i.e. $M_{vRFP}$, the product of the relatedness matrix and the RFP vector) and the relatedness adjusted patent portfolio (i.e. $M_{vP}$, the product of $M$ and seekers’ patent portfolio vector). Problem-seeker knowledge distance is then calculated as:

$$ED = \sqrt{(M_{vP} - M_{vRFP})^T (M_{vP} - M_{vRFP})}$$

where $v_{RFP}$ is defined as the RFP-vector and $v_{P} (k_{p1}, k_{p2}, ..., k_{p129})$ is defined as the seeker firm’s patent portfolio-vector capturing the share of patents the seeker firm held in class $i$ in the five-year period prior to the RFP-project. Problem-seeker knowledge distance (ED) thus compares a seeker’s patent portfolio to the RFP vector, with both vectors being adjusted for the expected knowledge relatedness between patent classes in the IPC taxonomy. The measure is finally normalized to range between 0 and 1.

The Euclidean distance has been commonly used in previous studies as a measure to capture the intellectual distance among members of project selection panels (Crisculo et al., 2017), the technological distance between firms (e.g., Ahuja, 2000; Benner and Waldfogel, 2008; Rosenkopf and Almeida, 2003; Song et al., 2003), or to compare two vectors representing the patent portfolios of firms and industries (Hohberger et al., 2015). As a robustness check, we re-ran our analyses using an alternative knowledge distance measure based on the Cosine angle which has been used in studies on innovation to measure firms’ technological positions relative to alliance partners (Sampson, 2007).
the content distance of technical documents (Piezunka and Dahlander, 2015), or the expertise distance between knowledge providers and seekers (Hwang et al., 2015). The results are consistent with those of the Euclidean distance reported in the results section of this paper (see Appendix D for details).

3.2.3. Moderator variables

Identity disclosure. We constructed an indicator variable taking the value 1 if the seeker firm’s name was disclosed in the RFP and 0 otherwise. As an illustration, Appendix A provides an example of a non-anonymous and an anonymous RFP from our field data. To ensure accuracy, we checked our coding of the RFP documents with entries in the project database of the intermediary.

Status. In our study context of intermediary-administrated crowdsourcing, seeker firms that choose to remain anonymous in an RFP can send a status signal indicating their rank on either the Fortune 500 or Fortune Global lists. Accordingly, and in line with recent research on organizational status (Piazza and Castellucci, 2014; Podolny, 2005; Simcoe and Waguespack, 2011), we use the position of a seeker firm in Fortune’s rankings of the top 1000 (US and global) firms by revenue to proxy for seeker firms’ financial status. In cases of anonymous RFPs, we take the Fortune rank that is indicated in the RFP document (see Fig. A2 in Appendix A for an example). In cases where seekers’ identity is disclosed in the RFP, we look up seekers’ position on the Fortune list in the respective year of the RFP. This information was used to create a five-point categorical status measure (Status) coded 0 if the seeker firm was unranked, 1 if ranked at positions between 1000 and 501, 2 between 500-101, 3 in the 100-51 range, and 4 if the seeker firm was among the top 50 of either Fortune or Fortune Global lists.

3.2.4. Control variables

We include two proxies to capture the specificity and strictness of a crowdsourcing problem (RFP). First we use Num req to control for the number of solution criteria and technical requirements that are addressed in the RFP document. The second proxy, Num ex, is calculated as the number of solution approaches explicitly labelled “not of interest”. The variable thus measures to what extent the solution space of the problem is restricted in the RFP. Exist is a dummy variable indicating whether the seeker firm asks the intermediary to exclude distinct organizations (e.g., competitors) or solvers from certain industries from the request (1) or not (0). This list is used internally by the NineSigma team when soliciting responses from potential solvers directly. Reward is the monetary compensation awarded for winning solutions in the RFP document. We use average exchange rates in the year of the RFP to convert rewards in international currencies into USD. Further, we controlled for the extent and intensity of the intermediary’s email campaign to disseminate the RFP to potential solvers. Feitler et al. (2012) provide a detailed description of NineSigma’s use of customized email lists to invite potential solvers to respond to clients’ technology needs. NineSigma uses such email campaigns to complement the online release of RFP documents on their website. We use the number of email invitations (Mailcount) and reminder emails (Remindercount) sent by NineSigma to control for the effect of email volume on solver attention received. In order to adjust for skewed distributions, we use the square root values of Reward, Mailcount, and Remindercount in the analysis.

To control for effects of crowding, we count the number of RFPs of the same three-digit IPC class that were launched in the 30 days before and after the focal RFP was launched (Competing RFPs). To account for seekers’ technological strength, we include the total number of patents held by each seeker firm in the five-year period prior to the RFP (Patent stock). Portfolio composition is operationalized as the number of different patent classes in a seeker’s patent portfolio relative to the total number of 129 possible three-digit IPC patent classes. Finally, we control for the possible impact of an economic downturn because it may influence firms’ innovation and search behavior. In line with previous studies (Hud and Hussinger, 2015; Ugur et al., 2016) we include an indicator (Crisis year) taking the value 1 if the RFP was initiated during the recent financial crisis (project year = 2009) and 0 otherwise. Because some of the control variables were skewed, we used square root and logarithmic transformations (indicated by superscripts a and b) to arrive at distributions that more closely approximate the assumption of normality.

4. Analysis and results

Table 1 provides a descriptive overview of our main variables, and their correlations are reported in Table 2. The average RFP in our sample defines 6.36 technical requirements and excludes 2.41 solution approaches to specify the technology need. It has 2.5 competing RFPs that solicit solutions to similar technology problems and receives 16.65 solution proposals from the solver pool. The seeker’s identity is disclosed in only 14 per cent of RFPs, whereas an exclusion list is used in 63 per cent of RFPs. With an average RFP-to-patent portfolio distance of 0.49 (Euclidean distance), the average technology need in the sample is located at the centre of the knowledge distance spectrum. Finally, an average seeker firm scores 1.73 out of a possible maximum of 4 on the seeker-status index. The results of the hierarchical ordinary least squares (OLS) regressions are reported in Tables 3 and 4. All models are estimated using robust standard errors, clustered by seeker firm to account for potential within firm correlations. Model 1 in Table 3 is the baseline model and only includes the control variables of seeker firm and RFP attributes. Model 2 adds the linear and squared terms of the independent variable to test the curvilinear relationship between problem-seeker knowledge distance and solver attention. Models 3 to 8 in Table 4 split the sample to test the interaction of problem-seeker knowledge distance and disclosure at different values of seeker status. We use three cut-off values (seekers among the top 500, top 100, and top 50 of the Fortune list) to distinguish high-status seekers from all other seekers in the sample. These cut-off values are not sample-dependent, such as tests at one standard deviation below and above the median or mean, and represent meaningful focal values for seeker status at which the interaction between distance and disclosure can be measured (Spiller et al., 2013).

Some interesting findings emerge from the baseline model (Model 1). The coefficient of Competing RFPs is positive and significant suggesting that crowding has a positive effect on received solver attention. This is surprising since high numbers of similar RFPs could be expected to have a negative effect on the attention each RFP receives in terms of solutions from the solver pool due to “bandwidth problems” (Piezunka and Dahlander, 2015). Instead, requests soliciting solutions to related problems seem to reinforce the crowd’s interest in the respective problem domain and lead to more solutions to each of the related requests. However, the number of approaches described as being ‘not of interest’ in the RFP (Num ex) as well as the count of requirements and solution criteria listed in the RFP (Num req) do not bear a significant relationship with received solver attention.

The other controls perform largely as expected. The effect of Reward has a positive, albeit marginal, impact on received solver attention. Exclusion list has a negative and significant coefficient. The crisis indicator (Crisis year) is positive and significant. The coefficient of the number of reminder emails sent (Remindercount) is also positive and significant, but here the effect on received attention is also marginal.

**Hypothesis 1.** Predicts a curvilinear relationship between problem-seeker knowledge distance and received solver attention. The linear term of the knowledge distance variable (ED) in Model 2 is positive and significant (b = 1.927; p < 0.01) and the squared term (ED sq) is negative and significant (b = -1.698; p < 0.05). Although providing
initial support for Hypothesis 1, these results are not sufficient to confirm a curvilinear relationship between the independent and the dependent variable (Haans et al., 2016). Besides significant coefficients of the expected sign, two additional conditions need to be met to confirm an inverted U-shape. First, the lower part of the slope (left of the inflection point) has to be positive and significant and the upper part of the slope negative and significant. Second, the inflection point has to be located within the range of the data (between 0 and 1).

Applying the *U test* command in Stata 12 (Lind and Mehlum, 2010) confirms that the inflection point lies at a value of 0.567 (x = −β1/2β2 = −1.927 / (2 + 1.698) = 0.567) and falls within the data range of ED. The results further show a positive and significant positive lower bound slope (1.927; p < 0.005), and a negative and significant upper bound slope (−1.468; p < 0.05). Thus, we find support for Hypothesis 1.

**Fig. 2** reports the fitted values between problem-seeker knowledge distance (ED) and *Received Solver Attention* based on Model 2, and shows that RFPs with moderate distance (ED = 0.567) to the existing knowledge stock of the seeker firm yield maximal levels of solver attention. Values on the right-hand side of the distance scale are associated with relatively wide confidence bands, indicating that the data in our sample is much sparser at high levels of problem-seeker knowledge distance.

Models 3 through 8 test Hypothesis 2, which suggests that the prediction of received solver attention by the interaction between knowledge distance and seeker identity disclosure is contingent upon the level of seeker status. To test this effect, we split the sample at three levels of the seeker status variable. Models 3 and 4 analyze split-samples of high-status seekers (*Status* = 2: RFPs from seekers that are among the top 500 of the Fortune ranking) and low-status seekers (*Status* < 2: all other seeker firms), respectively.

Results show that in the high-status sample (Model 3), both interaction terms, linear and squared, are statistically significant (*b* = −9.784, *p* = 0.064; *b* = 10.667, *p* = 0.044), while in the low-status case the interaction terms between knowledge distance and identity disclosure are insignificant. When we move the status threshold up to values of *Status* = 3 and assign seekers to the high-status group only when they are among the top 100 of the Fortune list (Model 5), and to the low-status group otherwise (Model 6), we obtain identical results. Again, for seekers ranking high in the Fortune list, both interaction terms are significant and have the expected signs (*b* = −10.680, *p* = 0.04; *b* = 11.643, *p* = 0.025). There is no significant interaction effect of *Disclosure* on the distance-attention relationship for low-status seeker firms. Finally, Models 7 and 10 split the sample at values of *Status* = 4. As predicted, the coefficients of the interaction terms are not significant in the low-status sample. For the high-status group (*Status* > 4: RFPs from seekers in the top 50 of the Fortune list) we find the coefficients of the linear and the quadratic interaction terms to be significant (*b* = −19.816, *p* = 0.062; *b* = 21.869, *p* = 0.027). This suggests an interaction of problem-seeker knowledge distance, disclosure, and status when the threshold for high status is very strict.
Table 3
OLS Models (1–2) predicting the number of submitted solution proposals.

|                    | Model 1 Baseline | Model 2 H1 Main effect |
|--------------------|------------------|------------------------|
|                    | b/p              | se                     | b/p              | se                     |
| Num_ex*            | −0.033           | (0.041)                | −0.028           | (0.042)                |
| Num_req*           | 0.037            | (0.054)                | 0.038            | (0.055)                |
| Reward*            | 0.000*           | (0.000)                | 0.000*           | (0.000)                |
| Exlist             | −0.116*          | (0.116)                | −0.016*          | (0.208)                |
| Mailcount*         | 0.002            | (0.002)                | 0.002            | (0.002)                |
| Remindercount*     | 0.003*           | (0.002)                | 0.003*           | (0.002)                |
| Competing RFPs     | 0.028**          | (0.046)                | 0.036**          | (0.002)                |
| Patent countb      | 0.014            | (0.011)                | 0.015            | (0.011)                |
| Portfolio composition | −0.223       | (0.124)                | −0.283           | (0.124)                |
| Crisis year        | 0.185*           | (0.012)                | 0.201**          | (0.005)                |
| Disclosure         | 0.168            | (0.073)                | 0.161            | (0.071)                |
| Status             | 0.037+           | (0.010)                | 0.035            | (0.011)                |
| ED                 | 1.097            | (0.017)                | 1.027            | (0.022)                |
| ED_sq              | −1.698           | (0.023)                | −0.686           | (0.023)                |
| Constant           | 2.046***         | (0.000)                | 1.517***         | (0.000)                |
| N (observations)   | 637              | 637                    | 637              | 637                    |
| N (seeker firms)   | 183              | 183                    | 183              | 183                    |
| F                  | 5.290            | 5.573                  | 5.573            | 5.573                  |
| Model significance | 0.000            | 0.000                  | 0.000            | 0.000                  |
| Log likelihood     | −580.58          | −577.22                | −577.22          | −577.22                |
| AIC                | 1.664            | 1.659                  | 1.659            | 1.659                  |
| R2                 | 0.098            | 0.108                  | 0.108            | 0.108                  |
| Adj R2             | 0.081            | 0.087                  | 0.087            | 0.087                  |

Notes. The models were estimated using OLS regressions. Coefficients are unstandardized. Huber-White robust standard errors clustered by seeker firm are in parentheses, p-values in brackets. Two-tailed tests are reported: * * * indicates significance at the 0.1% level, ** indicates significance at the 1% level, * indicates significance at the 5% level, and + indicates significance at the 10% level. a Indicates a square root transformation. b Indicates a logarithmic transformation.

Together, these results support Hypothesis 2 and suggest that the moderating effect of Disclosure is contingent upon the level of seeker’s Status (Dawson and Richter, 2006).

5. Discussion

In this paper, we study two factors influencing the uncertainty solvers perceive when making a decision whether to allocate attention to a particular crowdsourcing challenge or not, respectively the confidence they perceive whether it could be worthwhile to participate: the quality of the problem articulation and the additional contextual information revealed in an RFP. We argue that both factors are the result of dedicated decisions by the seeker and affect crowdsourcing performance, measured in terms of solvers’ participation behavior. Building on the problem-solving perspective (Baer et al., 2013; Foss et al., 2016; Nickerson and Zenger, 2004), our research complements earlier studies on knowledge seekers’ choices between different modes of open innovation governance (Afuah and Tucci, 2012; Felin and Zenger, 2014; Van de Vrande et al., 2009). We match problem characteristics to the extant knowledge stock of a seeker firm to explain the effectiveness of crowdsourcing with regard to attracting a sufficient number of contributors. Uncertainty reduction theory (URT) provides a useful perspective on the interaction process between seekers and solvers, as it helps us to understand the different types of uncertainty that drive solvers participation behavior. Our core argument is that seekers can influence solvers’ confidence in the decision to allocate attention to crowdsourcing problems by means of strategic revealing decisions when drafting the RFP document.

By providing empirical evidence of the link between knowledge distance and problem formulation, we offer a more nuanced understanding of the established perception in the literature that crowdsourcing is a useful approach to solving technical problems that are distant from a seeker’s current technological knowledge stock (Afuah and Tucci, 2012). Our data indicates that crowdsourcing is neither an adequate mechanism for problems that are completely unrelated to the current field of expertise of a seeker (distant problems), nor for those that are supposed to provide solutions to problems closely related to their current R&D activities (local problems). Seekers can expect optimal returns from crowdsourcing in terms of attracting high numbers of solution proposals when they select problems where sufficient domain knowledge is in place to formulate the technology need comprehensively enough so that knowledgeable solvers can recognize and understand the technical challenge. Still, the problem domain should not be so familiar that seekers risk being overly restrictive when defining technical specifications and solution requirements.

However, we find that seeker firms with superior (financial) status are able to attract solutions at the extreme ends of the knowledge distance spectrum when their identity is disclosed in an RFP, while anonymous seekers are not able to leverage their high status. While in general seeker firms suffer from diminishing returns to knowledge distance, this condition does not apply for high-status seekers who engage in identity-based signalling. Confirming recent work on the link between status and attention (Azoulay et al., 2014; Simcoe and Waguespack, 2011), our results indicate that seeker identity disclosure may convey information about seekers’ underlying qualities. Superior identity-based status signals provide an additional stimulus for solvers to submit solution proposals. By transmitting a high status signal, low levels of solver confidence at the two extreme points (local vs. distant problems) can be overcompensated. An explanation for this effect provides the understanding of status as a social construct (Jensen, 2006; Jensen and Roy, 2008), which suggests that higher status not only affects the perception of the focal actor, but also the perception of its affiliates. For solvers it might be more attractive to submit proposals to seeker organizations that are associated with high (financial) status, as their own status can benefit from their affiliation with a higher-status con
Table 4
OLS Models (3–8) predicting the number of submitted solution proposals.

| Status        | Model 3 H2 | Model 4 H2 | Model 5 H2 | Model 6 H2 | Model 7 H2 | Model 8 H2 |
|---------------|------------|------------|------------|------------|------------|------------|
| Status > = 2  | b/p        | se         | b/p        | se         | b/p        | se         |
| Num_ex⁶       | −0.059     | 0.026      | −0.081     | 0.033      | −0.123+    | 0.012      |
| (0.242)       | (0.050)    | (0.046)    | (0.051)    | (0.046)    | (0.064)    | (0.063)    |
| Num_re⁶       | 0.029      | 0.009      | 0.059      | 0.058      | 0.009      | 0.03       |
| (0.697)       | (0.073)    | (0.078)    | (0.074)    | (0.077)    | (0.076)    | (0.100)    |
| Reward⁶       | 0.000+     | 0.000+     | 0.000+     | 0.000+     | 0.000      | 0.000      |
| (0.884)       | (0.000)    | (0.000)    | (0.000)    | (0.000)    | (0.000)    | (0.000)    |
| Exlist        | −0.078     | −0.195**   | −0.077     | −0.198**   | −0.059     | −0.138*    |
| (0.428)       | (0.097)    | (0.074)    | (0.459)    | (0.103)    | (0.009)    | (0.680)    |
| Mailcount⁶    | 0.004+     | 0.001      | 0.004+     | 0.001      | 0.007*     | 0.002      |
| (0.090)       | (0.002)    | (0.002)    | (0.002)    | (0.002)    | (0.002)    | (0.002)    |
| Remindercount⁶| 0.001      | 0.005*     | 0.001      | 0.005*     | −0.003     | 0.004*     |
| (0.624)       | (0.002)    | (0.002)    | (0.002)    | (0.002)    | (0.002)    | (0.002)    |
| Competing RFPs| 0.034+     | 0.008      | 0.008      | 0.008+     | −0.022     | 0.044***   |
| (0.078)       | (0.019)    | (0.013)    | (0.020)    | (0.009)    | (0.014)    | (0.032)    |
| Patent stock⁶ | −0.102*    | −0.025     | −0.097*    | 0.023      | −0.081     | −0.003     |
| (0.028)       | (0.045)    | (0.041)    | (0.046)    | (0.413)    | (0.581)    | (0.144)    |
| Portfolio composition | 1.052* | −0.328 | 1.035* | −0.327 | 1.282 | −0.046 |
| (0.028) | (0.466) | (0.318) | (0.476) | (0.381) | (1.220) | (0.866) |
| Crisis year   | 0.301*     | 0.152*     | 0.332*     | 0.139+     | 0.361+     | 0.161*     |
| (0.029)       | (0.032)    | (0.070)    | (0.128)    | (0.053)    | (0.071)    | (0.094)    |
| Disclosure    | 2.247+     | −0.29      | 2.464*     | −0.454     | 4.963+     | 1.12       |
| (0.624)       | (1.224)    | (1.008)    | (1.208)    | (1.014)    | (2.616)    | (0.906)    |
| ED            | 1.691      | 1.929      | 1.827      | 1.850*     | 10.233     | 2.010**    |
| (0.312)       | (1.658)    | (0.753)    | (1.661)    | (0.757)    | (0.148)    | (0.686)    |
| ED_sq         | −1.841     | −1.48+     | −2.062     | −1.393+    | −11.548+   | −1.658*    |
| (0.284)       | (1.703)    | (0.800)    | (1.679)    | (0.885)    | (0.894)    | (0.690)    |
| ED x Disclosure| −0.764+ | 1.761 | −1.018     | −1.68*     | 1.525      | 1.825      |
| (0.064)       | (5.174)    | (0.721)    | (4.914)    | (0.040)    | (5.064)    | (4.973)    |
| ED_sq x Disclosure| 10.667* | −1.394 | 11.643* | −2.226 | 21.869+ | 4.706 |
| (0.044)       | (5.178)    | (5.416)    | (5.053)    | (0.025)    | (5.687)    | (5.524)    |
| Constant      | 2.044***   | 1.133**    | 1.954***   | 1.141**    | 0.177      | 1.31***    |
| (0.000)       | (0.444)    | (0.358)    | (0.461)    | (0.002)    | (0.356)    | (0.929)    |
| N observations| 232        | 405        | 222        | 415        | 88         | 549        |
| N (seeker firms)| 59      | 156        | 55         | 158        | 21         | 178        |
| F              | 5.952      | 3.875      | 14.408     | 3.597      | 56.193     | 4.582      |
| Model significance | 0.000 | 0.000      | 0.000      | 0.000      | 0.000      | 0.000      |
| Log likelihood | −200.94    | −364.62    | −189.96    | −374.68    | −61.35     | −497.92    |
| AIC            | 1.870      | 1.880      | 1.856      | 1.883      | 1.758      | 1.872      |
| R2             | 0.141      | 0.123      | 0.166      | 0.119      | 0.346      | 0.103      |
| Adj R2         | 0.081      | 0.09       | 0.106      | 0.086      | 0.21       | 0.077      |

Notes. The models were estimated using OLS regressions. Coefficients are unstandardized. Huber-White robust standard errors clustered by seeker firm are in parentheses, p-values in brackets. Two-tailed tests are reported: ***indicates significance at the 0.1% level, **indicates significance at the 1% level, *indicates significance at the 5% level, and + indicates significance at the 10% level. aIndicates a square root transformation. bIndicates a logarithmic transformation.

Fig. 2. Predicted relationship between problem-seeker knowledge distance and number of solution proposals (received solver attention).

Note. Fig. 2 is based on Model 2 in Table 3.
It describes the trade-off between the benefits of openness and disclosure in terms of incentivizing more external contributors to participate and the downsides of openness in terms of increased uncertainties like the risk of revealing critical information to competitors. In the context of crowdsourcing, the paradox of openness suggests that if too little or the wrong problem-related information is disclosed in an RFP, potential solvers might not be willing to contribute and submit a solution. Our results, however, indicate that the negative effects of limited problem-related information in RFPs can be attenuated by the disclosure of contextual information. Our research hence contributes to a better understanding of this paradox.

6. Limitations and future research

Our study has some limitations that provide opportunities for future research. First, we focus on a single crowdsourcing intermediary. Although the search process used by NineSigma is similar to the ones employed by other intermediaries, there are some details of NineSigma’s approach that demand additional study with other intermediaries before generalizing our results broadly. For example, the approach to protect and transfer intellectual property as well as the transparency regarding the identity of seekers and solvers differ among various crowdsourcing service providers (Diener and Piller, 2013). For intermediaries that require a transfer of IP when solvers submit a solution proposal, the impact of some of our effects may change. For example, the risk of misappropriation (moral hazard uncertainty) as perceived by potential solvers will be higher in a setting where solvers’ solution proposals are not determined as non-confidential. Future research might replicate our analysis using other intermediaries to investigate whether our results also hold in such settings.

Second, the data and the analytical approach adopted in this study do not allow us to measure the relative contribution of trust versus status in a single model. Future research should use experiments to more clearly discern between the effects of trust and status on attention allocation in crowdsourcing. Another way to disentangle the two effects would be to conduct research that uses solvers as informants. Survey research and qualitative research with potential solvers would allow for a direct assessment of solvers’ uncertainty perceptions and help to gain a deeper understanding of their immediate consequences.

Third, our dependent variable measures the attention an RFP has received from the solver community. Future research could use more refined measures for solver contributions, while accounting for aspects such as pecuniary investments from solvers in developing their solutions. Such measures would also allow investigation of how these investments influence the perceived value and quality of a solution form the perspective of the seeker.

In conclusion, we find that crowdsourcing is still a fascinating area of research, as it fundamentally changes the way innovating firms seek technical input to their development approaches. We believe that our study could contribute to this debate by both enhancing the number of important governance factors involved and providing empirical evidence of the link between knowledge distance and crowdsourcing success (in terms of received solver attention).

Acknowledgments

The authors thank the handling editor, Keld Laursen, and two anonymous reviewers for their very helpful suggestions and guidance throughout the review process. The paper also benefited from helpful comments by Christian Hopp, Oliver Salge, and the participants at the RWTH Aachen University Innovation Research Seminar, the World Open Innovation Conference, and the Academy of Management Conference.
Appendix A. Exemplary RFP documents

Figs. A1 and A2

Fig. A1. Non-anonymous RFP.

Note. Fig. A1 shows an RFP where seeker firm’s identity is disclosed i.e., identity-based status signal. In such cases, seekers’ position on the Fortune list in the respective year was taken from fortune.com.

Fig. A2. Anonymous RFP with status information.

Note. Fig. A2 shows an example of an anonymous RFP where seeker status is signalled by the use of a status label (seeker’s rough position in Fortune list of most profitable companies).
Appendix B. NineSigma’s crowdsourcing model

Intermediaries like NineSigma work with their clients (seeker firms) through all stages of a crowdsourcing process. The figure below illustrates a typical workflow of a crowdsourcing project with the intermediary NineSigma (Fig. B1).

**Problem selection:** A “problem owner” of the seeker firm takes the initial decision to engage in crowdsourcing i.e., a research manager who is seeking a solution to a given task. Not all technical problems are equally suited for crowdsourcing (Afuah and Tucci, 2012). Depending on the nature of the task, some intermediaries are better suited for one specific problem than other (Diener and Piller, 2013). Once a seeker contacts NineSigma to commission an RFP, a program manager from NineSigma coordinates the project. Program managers are experienced project managers with a strong background in industrial R&D. 

**Problem formulation:** The first activity is to transfer the technical problem into a written problem statement, the request for proposals (RFP) document (see Figs. A1 and A2 for examples). RFPs are non-confidential documents comprising a brief description of the technology need, possible solution approaches, approaches not of interest for the seeker, evaluation criteria for successful solutions as well as information on project timing and business opportunities for the winning solver. While this problem-drafting activity is a core activity of the seeker firm, NineSigma supports this stage with sample RFPs, template documents, and feedback based on their experience from previous projects. Often, as a result of this feedback process, complex technical issues are divided into smaller sub-problems (Jeppesen and Lakhani, 2010; Lopez-Vega et al., 2016). Depending on the nature and complexity of the respective technical problem, this stage may require several iterations.

**Open call:** Upon completion, the RFP document is broadcast to a subset of NineSigma’s network of potential solvers (in total more than two million actors). NineSigma uses a customized email campaign to disseminate the RFP to several thousand potential solvers (Feitler et al., 2012). In addition, the RFP is openly released on the website of NineSigma, soliciting responses from solvers outside NineSigma’s solver community. NineSigma also frequently encourages its network partners to share RFPs in their own network.

While individuals (scientists, retired experts, individual researchers) are also part of the NineSigma’s solver base, the majority of its potential contributors are specialized technology providers, component manufacturers, applied research institutes, and other specialized organizations. Recipients of an RFP then self-select to the task and respond to the RFP. Participating solvers then submit a brief, non-confidential and non-enabling description of their proposed solution (in comparison with other intermediaries, NineSigma’s crowdsourcing approach focuses on the match between the actors, and only indirectly on the identification of a technical solution. Hence, all documents used for the broadcasting process are required to be non-confidential, as the exchange of confidential information only follows after the match is formed and after signing an NDA). NineSigma provides a template for solvers to structure their response. In general, submitted solution proposals compete with each other, and the proposal meeting the previously defined performance criteria best will receive an award, an offer for acquisition, or a contract for further development. Different from other crowdsourcing intermediaries (like Innocentive), the exact reward or financial incentive is often not specified in the RFP, but subject to further negotiations once the most promising solution proposal is selected. Accordingly, the decision to allocate attention and participate involves considerable uncertainty for solvers. In order to participate, solvers must first allocate time and effort to problem solving and then disclose their solution despite unknown odds.

**Evaluation:** After the submission deadline has passed, proposals are evaluated. The program managers facilitate the evaluation process by preparing a summary report of all submissions, which also includes contact information of the solvers. In general, only those proposals meeting the predefined solution criteria are forwarded in full text to the seeker. Next, seekers have to prioritize the proposals received according to a
classification scheme provided by NineSigma. If necessary, the intermediary will solicit additional information regarding the solution from the respective solvers.

**Acquisition:** Once interesting solutions have been identified, connections between the seeker firm and "winning" solvers are established. NineSigma sometimes helps seekers to engage in follow-up communication with the selected solvers and proposed milestones and activities for the technology acquisition or cooperation process. This back-end support particularly makes sense for anonymous RFPs as it ensures that the seeker’s identity is kept private as long as possible. But in general, this acquisition process or other activities following a successful matchmaking between a seeker and solver are governed by these partners directly, with little inclusion of the intermediary.

Appendix C. Summary statistics

See Fig. C1

![Fig. C1. Distribution of seeker status.](image)

**Appendix D. Supplementary analyses and robustness checks**

See Figs. D1 and D2 and Tables D1 and D2.

![Fig. D1. Predicted values of cosine distance.](image)

**Appendix E. Supplementary data**

Supplementary material related to this article can be found, in the online version, at [doi:https://doi.org/10.1016/j.respol.2018.07.022](https://doi.org/10.1016/j.respol.2018.07.022).
Fig. D2. Interaction of status, identity disclosure, and problem-seeker knowledge distance on received solver attention using cosine distance.

Table D1
OLS regressions using cosine distance as independent variable (Model 9).

| Main effect | b/p | se |
|-------------|-----|----|
| Num_ex*     | −0.028 | [0.502] | (0.042) |
| Num_req*    | 0.039 | [0.486] | (0.055) |
| Reward*     | 0.000* | [0.011] | (0.000) |
| Exist       | −0.117** | [0.025] | (0.052) |
| Mailcount*  | 0.002 | [0.253] | (0.002) |
| Remindcount*| 0.003* | [0.049] | (0.002) |
| Competing RFPs | 0.034** | [0.003] | (0.011) |
| Patent stock  | 0.016 | [0.509] | (0.024) |
| Portfolio composition | −0.259 | [0.312] | (0.255) |
| Crisis year | 0.205** | [0.064] | (0.070) |
| Disclosure | 0.163 | [0.113] | (0.102) |
| Status | 0.034 | [0.119] | (0.022) |
| CD | 1.293** | [0.066] | (0.467) |
| CD_sq | −1.700** | [0.008] | (0.638) |
| Constant | 1.814*** | [0.000] | (0.267) |
| N (observations) | 637 | | |
| N (seeker firms) | 183 | | |
| F | 5.670 | | |
| Model significance | 0.000 | | |
Table D1 (continued)

| Model 9 H1 | Main effect |
|------------|-------------|
|            | b/p        | se         |
| Log likelihood | 577.43 |
| AIC         | 1.680 |
| R2          | 0.107 |
| Adj R2      | 0.087 |

Notes. The model was estimated using OLS regressions. Coefficients are unstandardized. Huber-White robust standard errors clustered by seeker firm are in parentheses, p-values in brackets. Two-tailed tests are reported. **Indicates significance at the 0.1% level, *Indicates significance at the 10% level. Indicates a square root transformation. Indicates a logarithmic transformation.

Table D2

OLS regressions using cosine distance as independent variable (Models 10–15).

| Model 10 H2 | Status > = 2 | Model 11 H2 | Status < 2 | Model 12 H2 | Status > = 3 | Model 13 H2 | Status < 3 | Model 14 H2 | Status > = 4 | Model 15 H2 | Status < 4 |
|-------------|--------------|-------------|------------|-------------|--------------|-------------|------------|-------------|--------------|-------------|------------|
|             | b/p          | se          |            | b/p         | se          |            | b/p         | se          |             | b/p         | se         |
| Num_exa     | −0.062       | 0.027       | −0.083     | 0.035       | −0.12+      | 0.013      |
|             | [0.220]      | (0.050)     | [0.046]    | (0.047)     | [0.076]     | (0.064)    |
| Num_req     | 0.028        | 0.027       | 0.099      | 0.022       | 0.06        |
|             | [0.704]      | (0.073)     | [0.253]    | (0.078)     | [0.203]     | (0.101)    |
| Reward      | 0.000*       | 0.000*      | 0.000*     | 0.000*      | 0.000*      |
|             | [0.072]      | (0.000)     | [0.062]    | (0.000)     | [0.082]     | (0.000)    |
| Exist       | −0.078       | −0.2**      | −0.078     | −0.203**    | −0.041      | −0.141+    |
|             | [0.427]      | (0.097)     | [0.007]    | (0.074)     | [0.074]     | [0.786]    |
| Mailcount   | 0.002        | 0.001       | 0.005*     | 0.001       | 0.007*      | 0.002      |
|             | [0.056]      | (0.002)     | [0.033]    | (0.002)     | [0.042]     | (0.002)    |
| Remindercount | 0.01      | 0.02*       | 0.001      | −0.003      | 0.004*      |
|             | [0.649]      | (0.002)     | [0.039]    | (0.002)     | [0.047]     | (0.002)    |
| Competing RFPs | 0.033+   | 0.036**     | 0.037+     | 0.034*      | −0.018      | 0.042**    |
|             | [0.094]      | (0.020)     | [0.008]    | (0.013)     | [0.072]     | (0.021)    |
| Patent stock | −0.106*     | 0.013       | −0.032*    | −0.027      | −0.088      | 0.001      |
|             | [0.021]      | (0.045)     | [0.299]    | [0.008]     | [0.046]     | [0.032]    |
| Portfolio composition | 1.064* | −0.333     | 1.063*     | −0.281      | 1.348       | −0.055     |
|             | [0.021]      | (0.448)     | [0.298]    | [0.319]     | [0.457]     | [0.314]    |
| Crisis year | 0.313*       | 0.155*      | 0.346**    | 0.142*      | 0.354       | 0.168*     |
|             | [0.022]      | (0.133)     | [0.027]    | (0.069)     | [0.008]     | [0.126]    |
| Disclosure  | 0.714        | 0.002       | 0.812      | −0.035      | 0.909       | 0.399      |
|             | [0.163]      | (0.505)     | [0.994]    | (0.269)     | [0.117]     | [0.510]    |
| CD          | 1.521        | 1.545**     | 1.559      | 1.538**     | 1.893       | 1.541**    |
|             | [0.368]      | (1.676)     | [0.002]    | [0.491]     | [0.383]     | (1.771)    |
| CD_sq       | −2.77        | −1.727***   | −2.963     | −1.702**    | −5.67       | −1.882***  |
|             | [0.283]      | (2.555)     | [0.007]    | (0.631)     | [0.265]     | (2.631)    |
| CD x Disclosure | −5.634   | −6.352     | 1.938      | −6.324      | −2.602      |
|             | [0.157]      | (3.933)     | [0.611]    | [2.852]     | [0.110]     | (3.911)    |
| CD_sq x Disclosure | 11.16+    | −1.912     | 12.56+     | −2.795      | 15.3        | 4.612      |
|             | [0.091]      | (6.497)     | [0.717]    | (5.275)     | [0.054]     | (6.375)    |
| Constant    | 2.229***     | 1.415***    | 2.154***   | 1.407***    | 2.26**      | 1.601***   |
|             | [0.000]      | (0.345)     | [0.000]    | [0.323]     | [0.000]     | [0.319]    |
| N (observations) | 232    | 405        | 222        | 415         | 88          | 549        |
| N (seeker firms) | 59        | 156        | 55         | 158         | 21          | 178        |
| F           | 4.843        | 4.121       | 12.295     | 3.709       | 66.541      | 4.690      |
| Model significance | 0.000 | 0.000       | 0.000      | 0.000       | 0.000       | 0.000      |
| Log likelihood | −200.99  | −364.56     | −189.91    | −374.50     | −62.07      | −497.90    |
| AIC         | 1.871        | 1.879       | 1.855      | 1.882       | 1.774       | 1.872      |
| R2          | 0.141        | 0.124       | 0.167      | 0.120       | 0.335       | 0.103      |
| Adj R2      | 0.081        | 0.09        | 0.106      | 0.087       | 0.197       | 0.077      |

Notes. The models were estimated using OLS regressions. Coefficients are unstandardized. Huber-White robust standard errors clustered by seeker firm are in parentheses, p-values in brackets. Two-tailed tests are reported. **Indicates significance at the 0.1% level, *Indicates significance at the 1% level, Indicates significance at the 5% level, and + indicates significance at the 10% level. Indicates a square root transformation. Indicates a logarithmic transformation.
organizations' filtering of suggestions in crowdsourcing. Acad. Manage. J. 58 (3), 856–880.
Podolny, J.M., 1993. A status-based model of market competition. Am. J. Sociol. 98 (4), 829–872.
Podolny, J.M., 2005. Status Signals: A Sociological Study of Market Competition. Princeton University Press, Princeton, NJ.
Quatraro, F., 2010. Knowledge coherence, variety and economic growth: manufacturing evidence from Italian regions. Res. Policy 39 (10), 1289–1302.
Resick, C.J., Baltes, B.B., Shantz, C.W., 2007. Person-organization fit and work-related attitudes and decisions. J. Appl. Psychol. 92 (5), 1446.
Roberson, Q.M., Collins, C.J., Oreg, S., 2005. The effects of recruitment message specificity on applicant attraction to organizations. J. Bus. Psychol. 19 (3), 319–339.
Rosenkopf, L., Almeida, P., 2003. Overcoming local search through alliances and mobility. Manage. Sci. 49 (6), 751–766.
Saks, A.M., Ashforth, B.E., 1997. A longitudinal investigation of the relationships between job information sources, applicant perceptions of fit, and work outcomes. Personnel Psychol. 50 (2), 395–426.
Sampson, R.C., 2007. R&D alliances and firm performance: the impact of technological diversity and alliance organization on innovation. Acad. Manage. J. 50 (2), 364–386.
Saviotti, P.P., 2007. On the dynamics of generation and utilization of knowledge: the local character of knowledge. Struct. Change Econ. Dyn. 18 (4), 387–408.
Schäfer, S., Antons, D., Lüttgens, D., Piller, F., Salge, T.O., 2017. Talk to your crowd: principles for effective communication in crowdsourcing. Res. Technol. Manage. 60 (4), 33–42.
Sharkey, A.J., Kovacs, B., 2017. The many gifts of status: how attending to audience reactions drives the use of Status. Manage. Sci. https://doi.org/10.1287/mnsc.2017.2879.
Silveira, R., Wright, R., 2010. Search and the market for ideas. J. Econ. Theory 145 (4), 1550–1573.
Simcoe, T.S., Waguespack, D.M., 2011. Status, quality, and attention: what's in a (missing) name? Manage. Sci. 57 (2), 274–296.
Song, J., Almeida, P., Wu, G., 2003. Learning-by-Hiring: when is mobility more likely to facilitate interfirm knowledge transfer? Manage. Sci. 49 (4), 351–365.
Spence, M., 1973. Job market signalling. Q. J. Econ. 87 (3), 355–374.
Spence, M., 2002. Signalling in retrospect and the informational structure of markets. Am. Econ. Rev. 92 (3), 434–459.
Spiller, S.A., Fitzsimons, G.J., Lynch Jr., J.G., McClelland, G.H., 2013. Spotlights, floodlights, and the magic number zero. J. Market. Res. 50 (2), 277–288.
Teeler, D.J., Rumelt, R., Doi, G., Winter, S., 1994. Understanding corporate coherence: theory and evidence. J. Econ. Behav. Organ. 23 (1), 1–30.
Terwiesch, C., Xu, Y., 2008. Innovation contests, open innovation, and multi-agent problem solving. Manage. Sci. 54 (9), 1529–1543.
Toma, C.L., 2014. Counting on friends: cues to perceived trustworthiness in facebook profiles. Proceedings of the Conference on Weblogs and Social Media (ISWSM). Ugur, M., Trushin, E., Solomon, E., 2016. Inverted-U relationship between R&D intensity and survival: evidence on scale and complementarity effects in UK data. Res. Policy 45 (7), 1474–1492.
Van de Vrande, V., Vanhaverbeke, W., Duysters, G., 2009. External technology sourcing: the effect of uncertainty on governance mode choice. J. Bus. Venturing. 24 (1), 62–80.
WIPO, 2018. Guide to the International Patent Classification. Available at: http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf. Accessed on 25 April 2018.
Yang, Y., Chen, P., Pavlou, P., 2009. Open innovation: an empirical study of online contests. Proc. 30th Internat. Conf. Inform. Systems (Association for Information Systems, Atlanta).
Yang, Y., Chen, P.-Y., Banker, R., 2010. Impact of past performance and strategic bidding on winner determination of open innovation contest. Workshop Inf. Syst. Econ. 11–12.
Dr. Patrick Pollok is an Assistant Professor at the Research Area TIME (Technology, Innovation, Marketing, and Entrepreneurship) at RWTH Aachen University. Previously, he has been an innovation manager at Forschungszentrum Jülich, a member of the Helmholtz Association, and one of Europe’s largest interdisciplinary research centers. His research focuses on organizational search, learning, and innovation. Patrick obtained a Ph.D. in management and a M.Sc. in Business Administration from RWTH Aachen University.
Dr. Dirk Lüttgens is an Assistant Professor at the Research Area TIME (Technology, Innovation, Marketing, and Entrepreneurship) at RWTH Aachen University. His research focuses on open innovation, business model innovation, and the implications of the current digital transformation on firms. Dirk obtained a Ph.D. from RWTH Aachen University, worked at the University of Applied Sciences in Luzern, Switzerland, and has been a lecturer in several executive programs.
Dr. Frank T. Piller is a professor for management at RWTH Aachen University, where he heads the Technology & Innovation Management Group and is the academic director of the EMBA program. His research interests include open and user innovation, managing disruptive innovation, and implications of new information technologies for product development. Frank’s research has been published in Journal of Product Innovation Management, R&D Management, Academy of Management Perspectives, Journal of Operations Management, MIT Sloan Management Review, amongst others. Frank obtained a Ph.D. from the University of Wuerzburg and worked at the TUM Business School, HKUST, and the MIT Sloan School of Management.