Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

Distance Dilemma: The Effects of Knowledge Distance on Solvers' Participation in Innovation Crowdsourcing

Christoph Ihl
Hamburg University of Technology
TUHH Institute of Entrepreneurship
christoph.ihl@tuhh.de

Robin Kleer
Berlin University of Technology
Technology and Innovation Management Group
robin.kleer@tu-berlin.de

Jan Reerink
Hamburg University of Technology
TUHH Institute of Entrepreneurship
jan.reerink@tuhh.de

Abstract

Open innovation is exemplified in crowdsourcing platforms that allow firms to broadcast R&D problems to a wide range of potential solvers. Empirical evidence to date suggests that solvers from distant fields have higher chances to make winning contributions in crowdsourcing contests. Knowledge distance, however, can also negatively affect solvers' decision to participate in the first place. In particular, we propose three theoretical mechanisms: (1) lower initial attention paid to distant problems, (2) higher uncertainty about value creation and capture, (3) adverse costs and benefits expected from solving distant problems. To investigate these effects, we measure the distance between potential solvers and R&D problems in the nano science and technology knowledge space by the means of topic modeling with over 900.000 scientific papers and actual requests for proposals (RfPs). In a discrete choice experiment, we invite solvers to inspect randomly assigned RfPs of high and low knowledge distance in order to measure their willingness to submit a solution conditional on contractual provisions. Our findings reveal that scientists indeed pay lower attention towards distant problems. To be compensated for adverse costs and benefits, distant solvers demand higher award money and the right to license also to third parties. However, solvers in proximity rather than distance to the problem domain have higher demand for contract provisions that reduce uncertainty to better protect their "home" knowledge base.

Overall, we shed light on managing an important trade-off in innovation crowdsourcing: while more distant solvers could make valuable contributions, they are more difficult to attract.

Jelcodes:O32,O31

Distance Dilemma: The Effects of Knowledge Distance on Solvers' Participation in Innovation Crowdsourcing

Abstract

Open innovation is exemplified in crowdsourcing platforms that allow firms to broadcast R&D problems to a wide range of potential solvers. Empirical evidence to date suggests that solvers from distant fields have higher chances to make winning contributions in crowdsourcing contests. Knowledge distance, however, can also negatively affect solvers' decision to participate in the first place. In particular, we propose three theoretical mechanisms: (1) lower initial attention paid to distant problems, (2) higher uncertainty about value creation and capture, (3) adverse costs and benefits expected from solving distant problems. To investigate these effects, we measure the distance between potential solvers and R&D problems in the nano science and technology knowledge space by the means of topic modeling with over 900.000 scientific papers and actual requests for proposals (RfPs). In a discrete choice experiment, we invite solvers to inspect randomly assigned RfPs of high and low knowledge distance in order to measure their willingness to submit a solution conditional on contractual provisions. Our findings reveal that scientists indeed pay lower attention towards distant problems. To be compensated for adverse costs and benefits, distant solvers demand higher award money and the right to license also to third parties. However, solvers in proximity rather than distance to the problem domain have higher demand for contract provisions that reduce uncertainty to better protect their "home" knowledge base. Overall, we shed light on managing an important trade-off in innovation crowdsourcing: while more distant solvers could make valuable contributions, they are more difficult to attract.

Keywords:

Innovation contests, broadcast search, knowledge distance

1 Introduction

Many companies open up their innovation process to gain access to external knowledge from different domains (Laursen & Salter, 2006). Open innovation characterizes an innovation process that operates as an open search and solution process beyond technical and organizational boundaries (Chesbrough, 2006; Dahlander & Gann, 2010). The rationale behind open innovation is to overcome problems of local search and industry blindness (Stuart & Podolny 1996; Rosenkopf & Nerkar, 2001). While locally bounded search may be advantageous when current problems are similar to old ones, a limited search space only leads to obvious solutions and rarely to radical advancements (Rosenkopf & Almeida, 2003).

As a management approach, open innovation offers different methods and practices which support innovating companies to identify and integrate relevant external knowledge. Next to conventional arrangements such as innovation alliances or contract research, internet technology has enabled new forms of distributed problem solving, subsumed under the terms *tournament-based crowdsourcing* (Afuah & Tucci, 2012) or *broadcast search* (Jeppesen & Lakhani, 2010). These new forms are considered to be especially well suited to overcome local search biases and to tap into unobvious knowledge domains.

Prior research on broadcast search platforms has mainly focused on the characteristics of participants and its effect on the contest's outcome (Jeppesen & Lakhani, 2010) as well as on perceptions of the contests after participation (Franke et al. 2013). Within this paper, we examine the barriers researchers face in contributing to such an open call. We focus on the importance of knowledge distance in determining participation in innovation contests. Knowledge distance is closely related to the classical trade-off organizations face during the knowledge transfer process: If knowledge is too far away, it is difficult to transfer; if it is too close, there is little new information (Gilsing et al, 2008).

We use an empirical study in the field of nanotechnology to analyze the relation of knowledge distance and participation likelihood. Using a topic modeling approach, we match researchers in nanotechnology with real real-world contests broadcasted by innovation intermediaries. Participants first report their perception of the contest in a survey; subsequently they are confronted with variants of contractual arrangements in a conjoint study to analyze the drivers of their willingness to participate.

We find that knowledge distance negatively affects participation likelihood, a relation that can only be partially mitigated by contractual design parameters. This finding is interesting from the perspective of innovation intermediaries (or more generally the innovation seekers), as former studies argue that more distant solvers are likely to submit more innovative solutions (Jeppesen and Lakhani, 2010). Investigating further into the effects of the contractual design parameters, we find that these more distant solvers response most positively to an increase in monetary incentives and more retained patenting rights. Hence, this study contributes to our understanding of the optimal distance of solvers in broadcast search and solvers' participation drivers.

The paper is organized as follows: In the next section we describe the theoretical background on broadcast search and knowledge distance. In Section 3 we describe our data gathering process and the employed empirical methodology. Section 4 presents the results of our study which are discussed in Section 5. Section 6 concludes with implications for further research and limitations.

2 Theoretical background

In this section we describe the theoretical background of our research. First, we summarize existing research on broadcast search and identify knowledge distance as the central construct affecting the willingness of scientists to participate in innovation contests. Subsequently we derive hypotheses regarding the effect of this construct on participation likelihood and its interaction with contractual settings.

2.1 Innovation Crowdsourcing

Innovation crowdsourcing describes a search mechanism where a *seeker's* (typically a company) technical problem is announced broadly via web-based platforms to a large and diverse group of potential external *solvers* in form of an open request for proposals (RfP) (Jeppesen & Lakhani, 2010). The idea is to spread the problem as widely as possible to attract solvers even from unobvious knowledge domains and fields of expertise. Potential solvers screen the problem description and self-select whether to invest in solving the problem and to submit a solution proposal. The seeker then selects among all submissions the solutions that meet pre-defined performance criteria best and either awards a pre-defined prize money or negotiates terms of collaboration with the identified solution providers (Spradlin, 2012).

Very often, this process is facilitated by specialized intermediaries who provide broadcast search as a service to connect solution seeking clients with external solvers (Lakhani et al., 2007). Established intermediaries in this domain include NineSigma, InnoCentive, YourEncore, Atizio, or Yet2.com. Their success is greatly dependent on the ability to match seekers with solvers. Hence, most intermediaries maintain a web-based community of pre-registered solvers. In addition, intermediaries support clients in terms of drafting good problem statements, maintaining client anonymity, pre-selecting appropriate solutions, and monitoring fair play to prevent exploitation of solution proposals without the acquiring the underlying intellectual property (Diener and Piller, 2010).

Research on innovation crowdsourcing to assist technical problem solving is still scarce. It has mostly focused on the efficient design of contests and the effectiveness of the mechanism. Terwiesch and Xu (2008) propose a theoretical model that shows that an increase in the solver base results in a trade-off between the overall solution diversity and quality as well as solvers' problem-solving effort. This trade-off can be shifted in favor of the former through performance-contingent rather than fixed-price rewards. Boudreau et al. (2011) empirically study the effects of an increased solver base for innovation contests on the TopCoder platform. They find that benefits of a larger solver base, i.e. higher solution quality, outweighs the costs of fiercer competition and solvers' reduced effort, especially for complex problems. This is in line with Franke et al. (2014) who argue that a promising measure to increase chances of discovering a successful solution is to invite more candidates. However, with the number of invited contestants the required effort for solution screening is likely to increase (Piezunka and Dahlander, 2014) as there will be a wide range of unfitting proposals. In addition, the number of scientists and their available time is finite, hence there exists a natural limit for this way of improving a contest's outcome.

In order to overcome this problem, pre-selecting candidates based on the RfP may be a suitable solution. In a conceptual paper, Afuah & Tucci (2012) derive a number of testable propositions regarding the effect of the characteristics of problems, type of knowledge transfer and crowd characteristics on the effectiveness of the crowdsourcing mechanism. In particular, they argue that a seeker is more likely to crowdsource a problem if the distance to the knowledge needed is large. Jeppesen and Lakhani (2010) study the effect of distance in their study of crowdsourcing at InnoCentive. They define two types of marginality as predictors of innovation success: a technical marginality that indicates a difference in professional background between the solution seeker and the solver as well as a more social marginality which encompasses gender-related biases. For our study, it is particularly interesting that the effect of technical marginality is positive. The likelihood of submitting a winning solution increases with solvers' perceived technological distance between the problem domain and the solvers' field of expertise which is attributed to a changed perspective of researchers. These findings raise the question how solvers with a high distance to the problem react to an RfP.

We can conclude that the outcome of innovation crowdsourcing is driven by the number of participants and their proposal quality. Both these factors are influenced by the distance of the potential solver to the problem. Hence, this necessitates a closer analysis of the distance concept and its quantification.

2.2 Knowledge Distance

Distance can be understood as knowledge heterogeneity that arrives from diverse knowledge resources that each individual exhibits. It has been defined as the distance between different persons in terms of their mental perception function and ability (Nooteboom et al., 2007; Wuyts et al., 2005). The way this distance is structured arises from past behavior and experiences of persons and is therewith unequal for each individual. Due to different backgrounds, people interpret, comprehend and judge the world in various ways (Nooteboom et al., 2007). Resulting from cognitive inequality, the capability of solving one specific problem is diverging from one solver to another and explains why different scientist have divers distances to divers topics (Nooteboom et al., 2007; Wuyts et al., 2005).

Alternatively distance can be defined in terms of technological knowledge among potential partners of a knowledge exchange (Nooteboom et al., 2007). Then it is a construct of how large the knowledge bases and fields of expertise deflect among organizations and individual (Hartig, 2011) and aims at the interspace between organizations in terms of technological assets (Benner & Waldfogel, 2008). Technological distance is a determining factor when transferring knowledge. If the distance in terms of technological knowledge between the communicating instances becomes too large, a mutual understanding between those is precluded and the transfer is likely to fail (Nooteboom et al., 2007). If the technological distance is too close, the technological familiarity takes "out the innovative steam" and dramatically decreases the likelihood of novelty creation, which is the actual purpose of knowledge transfer (Gilsing et al., 2008). This view stresses the mutual learning aspect of collaborations. Mowery et al. (1996) on the other hand argue that firms are often not learning from one another in collaborations but are rather accessing or acquiring specific information. This perspective, which has also been confirmed in empirical studies by Grant and Baden-Fuller (2004) and Nielsen and Nielsen (2009), argues that larger distances are useful as the new knowledge only has to be integrated, which avoids the cost of learning (Balconi et al., 2013). In the context of broadcast search, this knowledge assessing view reflects the perspective of the seeker. Thus, it can be used to explain why seekers aim at getting proposals from distant solvers. However, the solvers’ incentives and benefits are neglected in analysis by Balconi et al.

While prior research on innovation platforms considered the positive effects of distance in performance, the effect of distance on participation likelihood has so far been neglected. Intuitively, a perceived distance to a subject decreases the probability of participation as scientists are likely to spend most of their attention on their specialty. Prior research has confirmed this intuition: in research on decision theory under uncertainty it has been shown that unfamiliarity reduces the likelihood of acting on an opportunity, there exists a bias towards the status quo (Samuelson & Zeckhauser, 1988). Constant et al. (1996) find that higher expertise leads to more contributions in online discussion. This is confirmed by Wasko and Faraj (2000) who show that individuals tend to respond more frequently in crowdsourcing situations when they feel to have sufficient expertise in the respective field. Haas et al. (2015) find additional evidence that individuals allocate more attention to a problem that is closer to their field. The costs and benefits of participation can serve as an explanation for such behavior. In addition to monetary rewards, solving RFPs is also expected to enhance reputation or to encourage future reciprocity (Chiu et al., 2006; Constant et al, 1996; Wasko and Faraj, 2000). These benefits are more likely and higher in expectation if the contest is closer to the participants' interests, thus expected benefits are higher. On the other hand, similar fields increase the chances that participants have the necessary absorptive capacity to understand the RFP (Cohen and Levinthal, 1990). As Kotha et al. (2013) argue, overlapping expertise fosters mutual knowledge and reduces communication costs. Therefore, it is easier for potential solvers to understand and make sense of a problem, capture its special characteristics and dependencies, and finally identify and submit a solution (Thomas et al, 2001; Tsai, 2001). Boudreau et al. (2011) add that greater knowledge distance can be interpreted as being less well informed, which is regarded in risk and decision theory as an indicator for greater uncertainty. It is thus likely that potential participants discount their chances of succeeding in the crowdsourcing contest on the basis of "ambiguity aversion" (Fox and Tversky, 1995). Thus, costs for participation are higher in expectation if knowledge distance is higher. Piezunka and Dahlander (2014) find that inviting many solvers to contribute overburdens the inviting company which then resorts to focus on input that is not distant. So even if distant solvers participate there is a chance that the seeker may be biased against distant knowledge. Additionally, the probable lack of prior experience with innovation platforms may introduce additional bias. This implies that scientists unfamiliar with innovation platforms and/or a given technical problem from an innovation platform, are less likely to participate in a contest than scientists experienced with platforms and/or the sort of technical problem. The impact of knowledge (dis)-similarity has also been researched in the field of alliance formation: companies active in similar technological contexts are more likely to cooperate (Mowery et al. 1998, Rothaermel and Boeker 2008) as common knowledge stocks increase absorptive capacity, enabling firms to assimilate knowledge at lower cost (Lane and Lubatkin 1998). The concept of absorptive capacity as explanatory mechanism for a negative effect of distance on participation has been studied by Haas et al. (2015) in the context of company internal forums for problem solving. Consequently, we formulate our first hypothesis:

Hypothesis 1: The larger the knowledge distance of a scientist towards an RFP, the less likely is s/he willing to participate.

Next to the knowledge distance, it is likely that contractual details influence participation behavior (Franke et al., 2013). While the expected direct effect of contractual details is obvious in many cases (e.g., increased participation for higher monetary payment), the interaction of knowledge distance and contractual details is interesting (as these details can be influenced by the seeker / intermediary). In particular, it is interesting to know what measures an intermediary can employ to attract more distant solvers. The various business models of innovation intermediaries hint to the fact that different sets of incentives or platform characteristics are used to attract different kinds of solvers. We can group the effect of contractual details in two categories: Safeguards and effects on costs and benefits of the solver. We expect more distant solvers to require additional contractual safeguards as their distance is likely to result in higher uncertainty. Hence, given ambiguity aversion, distant solvers would be less likely to contribute given the same set of safeguards. Since distant solvers are less likely to benefit from potential spillovers such as reputational effects or industry connections when they participate in broadcast search additional compensation may be required to entice their participation.

3 Data and Methods

3.1 Identification of researchers and deriving a distance measure

The distance between a scientists' experience and the technical problem is an important aspect in the design of our study. For the purpose of our study it would prove useful to be able to measure distance as a first step to pre-select respondents. In contrast to interpersonal distance a more technological distance may be quantified by careful analysis of the texts describing the technical problem on the one hand and the texts written by the potential solver on the other. A manual approach to this problem, ideally by experts in the respective fields, may yield high quality results but seems impractical given the large amount of data. Instead we use machine learning algorithms to compare texts.

Our research focusses on RFPs and researchers in the field of nanotechnology. While this limits the scope of our study to a certain extent, nanotechnology is a general purpose technology and therefore possible distance values are not too restricted. As a first step we gathered information on RFPs published by innovation intermediaries related to nanotechnology by searching the web. We found some 4.700 RFPs from two of the leading broadcast search platforms: NineSigma and Innocentive. RFPs related to nanotechnology were identified by a simple search for the keyword "nano*" (i.e. words containing "nano").

The results were checked using a more complex search based on Arora et al. (2012) to rule out false positives (such as “nanoliter”, which contains the letters “nano” but does not necessarily imply that the article is about nanotechnology), resulting in 110 nanotech RFPs. These were manually checked for suitability. RFPs that, despite the keyword-based searches, were not clearly related to nanotechnology were removed (in some instances nanotechnology was only mentioned very briefly among possible approaches to a viable solution). We also removed RFPs that differed significantly in text length (i.e. data available to automatic processing), time required for solution as well as required team size. The resulting set of 38 RFPs was processed to remove irrelevant information (e.g. specifics on the submission process, formatting) as well as information relating to the treatments of the planned conjoint analysis (e.g. firm identity). We noticed that inactive RFPs (those that have been withdrawn or where a winner has been awarded) contained less information compared to active RFPs, with inactive RFPs representing the majority of RFPs in the downloaded dataset. However, the information contained in inactive RFPs corresponds well to the information contained in a publication abstract. While we lose information in comparison to an analysis based on full RFP descriptions and full papers the smaller amount of data significantly eases pre-processing. Once pre-processed, the RFPs were integrated into an online survey based on the estimated distance between RFP topic and scientists’ field of expertise. The data source for estimating this distance were publications relating to nanotechnology from 2000 to 2011 downloaded from the Web of Science. The bibliographic information from the Web of Science articles was searched for author e-mail addresses. To increase the expected response rate we only kept e-mails from researchers who published in the years 2010 or 2011. Starting with this set of e-mails, author names were disambiguated using a custom Python script: Word similarity metrics based on name components (last name, first name and initials), location information and co-author information were taken into account to find all nanotechnology publications in the dataset for each of the authors identified by searching for e-mail addresses. As a result, we obtained data on approximately 24,000 scientists. To estimate the proximity of a scientist's prior work to an RFP we used Latent Dirichlet Allocation (LDA), a generative model for text data also known as topic model (Blei et al., 2003). Topic models are generative statistical models that return probability distributions of groups of words that tend to occur together in texts (topics). As a topic model translates texts into vector space the model can be used to compare two texts with similarity measures. To make sure that our model accounts for the variance in scientific and technical texts we used a large number of scientific paper abstracts as well as patent abstracts from the field of nanotechnology published in the years 2000-2012 (approximately 850.000). After pre-processing the abstracts using the Python library NLTK (Bird et al. 2009) (removing irrelevant information such as copyright data, stemming the remaining text and removing stop-words) a number of models were calculated by varying the topic parameter. We compared two implementations of LDA: Gensim and Mallet (McCallum, 2002; Rehurek & Sojka, 2010). Model perplexity and subjective tests of similarity scores obtained with various models were used to select a model estimated with 250 topics in Mallet. We proceeded to calculate the similarity of each paper for each author to the RFPs in our sample. The similarity measure employed is cosine similarity. The average of similarity scores for one author's papers to all RFPs as well as the highest similarity was used to determine the proximity of each author's knowledge stock to the various technical challenges described in the RFPs. For each author three RFPs were selected, two that are conceptually close (highest and second-highest similarity score between author's papers and one RFP) and one that is more distant to the author's work (RFP where similarity score is close to the mean similarity of all author-RFP pairings). For the online survey either the closest or the medium distance RFP was randomly allocated to candidates. The second-closest RFP was kept as a reserve; in the event that a respondent decides not to complete the survey after reading the initial RFP the scientist is given the option to complete the study with the second-best RFP instead.

Since prior art conceptualizes knowledge disparity as distance we convert the topic model cosine similarity measure to a knowledge distance measure using the approach described in Goldberg et al. (2015) of using an exponential link between negative distance and similarity.

3.2 Survey

We invited the scientists identified with Web of Science articles to a survey in which they were confronted with the RFP selected as described in section 3.1. In the questionnaire to the RFP we retrieve some general information on researchers. Table 1 shows variables obtained either from the survey or by analysis of bibliographic data.

--- insert table 1 about here ---

Potential solvers were asked to report their experience (in years) in academia and industry respectively. Solvers were also asked to report how many patents they (co-) invented. To measure the degree of involvement with industry we used the industrial involvement index developed by Bozeman and Gaughan (2007) and normalize it to the range 0-1. Own Broadcast Experience is a dummy variable that takes the value 1 if potential solvers already had some experience with broadcast search platforms in the past.

Subsequently researchers participated in a conjoint experiment where the following variables were modified across a set of 5 scenarios: the seeker type was set to either small company, large company or governmental organization. The identity, i.e. the name of the company or organization, is either revealed or withheld. Seeker location could be either

Asia, Europe or the US. Incentives and barriers consisted of one variable for the required technical maturity, a variable for the timing of IP disclosure by the solver as well as one variable for retained publication and patenting rights respectively. Finally, four levels of financial rewards for submitting a winning solution were defined, ranging from US$ 10,000 to US$ 75,000. Details on these variables are shown in table 2. Following each scenario, respondents were asked to choose for one of two contractual designs differing in the variables described, a none-option was included.

--- insert table 2 about here ---

In order to account for the possibility of different RFPs containing varying amounts of information despite pre-processing we control for the length of the RFP description in words. We also control for the RFP source (Ninesigma or Innocentive). The RFP similarity was obtained using the topic modeling approach described above.

The location of the solver was estimated from the top-level domain of author e-mail addresses obtained from bibliographic data. Bibliographic data was also used to determine the citation count for each author. We further defined a variable indicative of the breadth of a scientist's knowledge by comparing a scientist's publication abstracts to all RFPs in our dataset. The average similarity of one scientist's abstracts to the topics obtained by topic modelling, relative to the average similarity of all scientists' publications to these topics, was used to decide whether a scientist is familiar with a given topic. We then summed the familiar topics for each author to obtain a breadth measure ranging from 0 to 250. We finally asked respondents to choose between two contracts with the additional option of not participating in either. Based on this question we form our dependent variable participation.

In total, we received 249 responses to the survey and the conjoint analysis with five contractual experiments per respondent.

3.3 Selection bias

Bias may be introduced at two points in our study: after receiving an invitation e-mail a potential responder may choose to open the survey. Subsequently the potential responder chooses to complete the survey or not. It is possible that these two decisions are functions of the characteristics of the potential respondent. Hence the remainder of completed surveys used for further analysis may differ from the invited (random) population systematically. We correct for this issue using a modified two-step Heckman correction. According to Heckman (1979), selection effects can be regarded as instances of truncated data. Including an inverted Mills ratio calculated from an initial probit regression in a subsequent linear regression corrects the introduced bias. The Heckman correction has since been extended for cases of double selection (Mohany 2001), i.e. two subsequent selection effects. In this case the correction involves calculating two Mills ratios from a bivariate probit regression to correct the bias in both selection stages.

3.4 Latent class estimation

In order to analyze participants' responses in the conjoint study, we use latent class regression, an extension of the logit model. Logit models are a type of generalized linear model that estimate the utility of a choice as linear function of parameters that is linked to the categorical dependent variable through the logit function. The model can extended for the case of more than two outcome categories (in our case respondents can opt for one of two contractual arrangement or a none-option) with a multinomial logit model. If in addition to choice-variant attributes choice-invariant variables are to be included a conditional logit model is employed. The latent class model is a conditional logit model that allows correcting for unobserved preference heterogeneity with latent classes: observations are grouped along similar utility parameter estimates. For a logit model with k parameters each latent class adds another k parameters, hence there is a risk of overfitting. We use the Bayesian Information Criteria (BIC) to calculate a trade-off between the number of additional classes and the gain in log-likelihood. The BIC penalizes the use of additional parameters more strongly than the Akaike Information Criterion (AIC), thereby allowing us to optimize the trade-off between more complex model specification and model (over)-fit.

4 Results

Table 3 shows the results from the selection model. We account for the selection process of scientists into our sample and potential selection biases in three stages: *1st stage*: 24,374 invited scientists decide whether or not to inspect the assigned RfP. *2nd stage*: 569 scientist who inspected the assigned RfP decide whether or not to evaluate the contractual details of the assigned RfP. *3rd stage*: 229 scientists decide whether or not to accept certain contracts with respect to the assigned RfP. In the statistical analysis of this double selection process (cf. Mohanty 2001), a scientist's average distance to all 35 RfPs in our study serves as an exclusion restriction from stage 1 to stage 2, whereas RfP length and source serve as exclusion restrictions from stage 2 to stage 3. The bivariate probit selection model covering the first and second stage shows that the first stage decision to inspect the RfP is less likely for scientists with higher number of citations after controlling for other background variables. Thus, academic achievement in terms of citations has a negative effect on the interest in innovation crowdsourcing. In the second stage, conditional upon inspecting the RfP, a further interest in judging the contractual details of the assigned RfP is less likely for scientists with a greater knowledge distance towards the RfP. This lends initial support to our overall conjecture that knowledge distance reduces scientists' attention paid towards crowdsourced innovation problems. The significant correlation of error terms rho reveals a significant selection effect such that unobserved factors positively (negatively) affecting the decision to inspect the RfP in the first stage negatively (positively) affect the decision to further evaluate the contractual details in the second stage.

--- insert table 3 about here ---

Table 4 shows descriptive statistics for variables used in model estimation, across the three stages (invited population, partial completion of survey, completed survey). Cross correlation does not appear to be a concern. Noticeable are some outliers as regards the number of patents, citation count and years of experience in industry and academia.

--- insert table 4 about here ---

Table 5 shows a minimal BIC for a specification with two latent classes, compared to those with 1 or 3 classes. Based on BIC, two latent classes clearly yield the best model fit. To make the model even more parsimonious, we test for the restriction whether the control variables that affect the baseline adoption utility do not vary across the two classes (model 4). We accept this restriction because it improves model fit in terms of BIC. In model 5, we include our focal variable knowledge distance as an active covariate to predict class membership of scientists. Since the constant utility also varies with latent classes, this specification allows knowledge distance to exert both a direct effect on baseline adoption utility and a moderating effect on contract preferences. This specification yields a lower BIC, albeit with only little improvement. Based on this specification, we test a further restriction of a linear effect of prize money in model 6, which needs to be rejected. Instead, the specification of a monotonic increasing effect of award money in model 7 gives a more parsimonious model with better approximation to the data.

--- insert table 5 about here ---

Table 6 shows covariates of the baseline adoption utility. Inverse Mills Ratios from the bi-variate probit were included to correct for the double selection effect. As both lambda parameters are significant we can reject the null hypothesis that no selection effect occurs. In particular, this means that the 229 sampled scientists positively select themselves into the adoption of a crowdsourcing contracts in the third stage compared to a random sample from the 24,374 invited scientists due to unobserved factors affecting the first stage decision to inspect the RfP itself. And the 229 sampled scientists negatively select themselves into the adoption of a crowdsourcing contract in the third stage compared to a random sample from the 569 scientists who inspected the RfP due to unobserved factors affecting the second stage decision to evaluate the contractual details of the RfP. The latter effect could be interpreted that scientists who would benefit from entering into a crowdsourcing contract are deterred from thinking about contractual details in the first place.

Conditional upon selection into the contract decision stage and controlling for (unobserved) factors affecting the double selection into this third stage, the sampled scientists' number of citations do not exert any further (negative) effect on baseline adoption of innovation crowdsourcing. Instead, previous experience with innovation crowdsourcing as well as knowledge breadth in the problem domain have a positive impact on the baseline utility of accepting a crowdsourcing contract, whereas the number of patents in the problem domain as a proxy for scientists' own commercialization potential has a negative effect.

Our focal variable knowledge distance is highly significant in predicting membership of the second class, which comprises over 70% of our sample. Scientists in this distant class have a significantly lower baseline adoption utility, thus, again lending support to our overall conjecture that knowledge distance reduces scientists' attention paid towards crowdsourced innovation problems. Furthermore, they have on average a lower responsiveness to all contracting attributes except for award money. The second important difference is that they benefit much more from being granted the right to apply for an own patent with a non-exclusive licensing.

--- insert table 6 about here ---

Table 7 shows averaged parameters for the choice-variant attributes. Except for the seeker type all attribute parameters are significantly different from zero. A Wald test for equality reveals that there are significant differences between one parameter across the two classes for seeker identity, IP disclosure, retained patenting rights and award money. Table 7 also shows the relative importance of attributes for the two classes with the distant class attributing more importance to financial incentives as well as publication and patenting rights.

--- insert table 7 about here ---

Table 8 shows detailed parameter estimates and standard deviations for choice attributes for both latent classes. As the parameters of the non-linear logit models cannot be easily interpreted we also included a willingness to pay measure to indicate the relative importance of an attribute value relative to the financial incentive. For class 1 the contractual incentives / barriers are significant and affect participation in the expected way. While higher required technical maturity reduces participation likelihood, an increase in the retained publication or patenting rights or in monetary incentives increases participation likelihood. The WTP indicates the non-linear relation between levels of the categorical variables: retaining publication rights without any restriction seems to be of much higher utility than either time or content restrictions. Compared to class 1 the more distant class seems to be more difficult to motivate for participation with contractual settings. Several incentive levels that motivate class 1 solvers are not significant for class 2 solvers. If they are significant, the WTP is lower. Interestingly, more distant solvers exhibit a strong preference for non-exclusive patent licensing. More distant solvers also appear to prefer seekers who reveal their identity. This may indicate that uncertainty plays a role in participation likelihood for this class of solvers and that seekers can reduce this uncertainty by being more transparent.

The difference in baseline adoption utility shows that more distant solvers are less likely to approve of the contractual settings, which corresponds well to the previous results of only a few contractual parameters being useful in enticing their participation.

--- insert table 8 about here ---

5 Discussion

According to Nooteboom et al. (2007) the positive effect of distant knowledge arising from its novelty is discounted by increasing demands on absorptive capacity required to assimilate distant knowledge. In the context of broadcast search platforms problems can be accurately described and a narrow solution space defined. This should enable seekers to avoid the negative effects related to distant knowledge as provided solution proposals are likely to adhere to the pre-defined solution space. In this case the novelty value of distant knowledge is merely discounted by the decreasing likelihood of distant solver’s contribution. However, it is also possible that seekers leave the solution space open in order to attract more distant solvers. In this case absorptive capacity is likely to play a role as proposed solutions to a defined problem stemming from unfamiliar contexts still require more effort on the part of the seeker to understand and as seekers tend to ignore distant solutions when confronted with resource conflicts in the filtering stage (Piezunka and Dahlander 2014). In this case the novelty value of distant knowledge is discounted by increasing demands on absorptive capacity as well as lower likelihood of distant solver participation. In either case the negative effect of distance on participation likelihood needs to be taken into account. Our study shows that knowledge distance conceptualized as function of RfP description and texts produced by potential solvers transformed to a vector space can be used to predict participation likelihood and reveals interactions with contractual design. Given our research design it is important to correct for selection bias in order to correctly estimate the effect of distance on participation. We have corrected for a selection effect that occurs when inviting potential solvers to the survey and in a second step, when the solvers decide on whether to complete the survey. This selection effect offered first insights into the link between knowledge distance and participation with many solvers opting out of the survey when they were more distant to the problem.

Prior literature has shown that more distant solvers tend to contribute more valuable solutions, which is a central (implicit) benefit of the concept of broadcast search. Our findings suggest that the promise of broadcast search cannot be realized if the effect of knowledge distance on participation is not accounted for. In a latent class regression we estimate the effect of various choice-variant and invariant parameters on participation. Individual-level variables are only of limited use when attempting to determine who participates. The academic background does not appear to be a significant predictor of participation at this point. However, the selection model has shown a strong negative effect of the citation count on the decision to participate in the survey. Prior experience with broadcast search platforms affects participation likelihood, suggesting that there are either learning effects or a reduction in uncertainty. This implies that innovation intermediaries need to overcome a hurdle when inviting new solvers and should invest into their existing solver network.

Analysis of contractual parameters using a conjoint study shows that more distant solvers do not react as well to incentives. Intermediaries can only partially mitigate this effect by being more transparent or by offering higher rewards. Distant solvers appear to be more sensitive to opportunity costs as evidenced by the class 2 parameters for financial incentives and non-exclusive patent licensing. The distant solvers' preference for higher rewards and greater freedom with regard to contractual aspects can be explained by relatively higher anticipated costs of understanding and contributing to the problem (Haas et al. 2015). Also, the direct benefit to a distant solver is expected to be lower: a close solver who contributes solutions to a company active in a similar context can expect to benefit from newly gained personal connections and learning effects that may apply to the solver's research. Awareness of these opportunity costs may explain distant solver's preference for higher financial rewards and relaxed contractual barriers.

For the seeking company this implies that increasing incentives will most likely attract more solvers at knowledge distance levels that are not optimal. Aside from the expense related to the incentives, additional effort would be required to filter the submissions. Prior research has shown that under these conditions seekers are likely to be overburdened and resort to discarding submissions that are more costly to evaluate due to the higher distance to their own knowledge domain (Pizunka and Dahlander 2014). The dilemma of distant solvers being less willing to participate even though they may have the knowledge required for optimal solutions may necessitate modifications to the concept of broadcast search.

6 Conclusion

From a managerial perspective, the results of our paper are helpful for intermediaries and other hosts of idea contests. We show that contests can be designed more effectively (in terms of attracting more and better suited participants) if the host is aware of the trade-off implied by knowledge distance. We have validated topic modeling as tool for objectively measuring knowledge distance and preselecting survey respondents. Our results show that publication abstracts serve as a useful predictor for relatedness of researchers to certain problems. This can be useful for technology transfer offices or innovation platforms: Intermediaries may want to confront the distance dilemma by pre-selecting potential respondents with appropriate knowledge distance. Limiting the population of possible participants in this way, it may be possible to attract solvers at the optimal knowledge distance with high incentive levels while limiting expenses and detriments related to evaluation of a flood of submissions with low probability of success due to sub-optimal knowledge distance. Furthermore, intermediaries may consider a transparent design of the intermediation process to reduce uncertainty in distant solvers.

From a theoretical perspective, we contribute to the understanding of participation motives in idea contests. Our research design that studies potential participants in idea contests allows us to distinguish between different types of participants and their respective motives and barriers to participation.

There are some limitations to our study: our topic model is based only on the abstracts of publications and the publicly available information of RFPs. A more advanced model could use full papers and more detailed descriptions of the technical problems. This might lead to a more accurate measure of distance between researchers and RFPs and thus allow for better match-making. Ideally one would compare scientific abstracts from all disciplines to all types of RFPs to assure coverage of the entire range of distance values between scientific papers and technical problems. We focused our attention on one branch of science, enabling us to download a limited number of RFPs and papers. However, with this restriction comes a possible bias in data selection as far as the possible range of knowledge distances is concerned: by excluding solvers from disciplines not related to the problem described in the RfP the possible knowledge distance is limited with an upper limit to the possible distance. In order to keep this problem to a minimum, we focus on a general purpose technology (nanotechnology,) which limits the scope of the study without infringing too much on possible distance values.

By presenting only one RFP to a potential solver we simplify from the real situation of potential solvers having to choose from a set of RFPs. However, with regard to the effect of distance on participation this is likely to lead only to an under-estimation of the effect as an increased number of choices tends to shift solver attention to closer problems (Haas et al. 2015).

Further research may attempt to investigate possible non-linear relations between distance and participation likelihood: if a full range of possible distance values is taken into account by expanding the dataset to solvers from completely unrelated disciplines the negative effect of distance on participation may be reinforced.

References

Afuah, A., Tucci, C.L., 2012. Crowdsourcing As a Solution to Distant Search. Academy of Management Review 37, 355–375. doi:10.5465/amr.2010.0146

Arora, S.K., Porter, A.L., Youtie, J., Shapira, P., 2013. Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs. Scientometrics 95, 351–370. doi:10.1007/s11192-012-0903-6

Balconi, M., Lorenzi, V., Saviotti, P.P., Zucchella, A., 2013. Cognitive distance in research collaborations. Università di Pavia, DEM Working Paper Series 51, 1–36.

Benner, M., Waldfogel, J., 2007. Close to You? Bias and Precision in Patent-Based Measures of Technological Proximity (No. w13322). National Bureau of Economic Research, Cambridge, MA.

Bird, S., Klein, E., Loper, E., 2009. Natural language processing with Python, 1st ed. ed. O'Reilly, Beijing ; Cambridge [Mass.].

Blei, D.M., Ng, A.Y., Jordan, M.I., Lafferty, J., 2003. Latent dirichlet allocation. Journal of Machine Learning Research 3, 2003.

Boudreau, K.J., Lacetera, N., Lakhani, K.R., 2011. Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis. Management Science 57, 843–863. doi:10.1287/mnsc.1110.1322

Bozeman, B., Gaughan, M., 2007. Impacts of grants and contracts on academic researchers' interactions with industry. Research Policy 36, 694–707. doi:10.1016/j.respol.2007.01.007

Chesbrough, H.W., 2003. Open innovation: the new imperative for creating and profiting from technology. Harvard Business School Press, Boston, Mass.

Chiu, C., Wang, E.T.G., Shih, F., Fan, Y., 2011a. Understanding knowledge sharing in virtual communities: An integration of expectancy disconfirmation and justice theories. Online Information Review 35, 134–153. doi:10.1108/14684521111113623

Chiu, C., Wang, E.T.G., Shih, F., Fan, Y., 2011b. Understanding knowledge sharing in virtual communities: An integration of expectancy disconfirmation and justice theories. Online Information Review 35, 134–153. doi:10.1108/14684521111113623

Constant, D., Sproull, L., Kiesler, S., 1996. The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. Organization Science 7, 119–135. doi:10.1287/orsc.7.2.119

Dahlander, L., Gann, D.M., 2010. How open is innovation? Research Policy 39, 699–709. doi:10.1016/j.respol.2010.01.013

Diener, K., Piller, F., 2010. The Market for Open Innovation: Increasing the efficiency and effectiveness of the innovation process. RWTH Aachen University.

Flanagin, A.J., 2007. Commercial markets as communication markets: uncertainty reduction through mediated information exchange in online auctions. New Media & Society 9, 401–423. doi:10.1177/1461444807076966

Fox, C.R., Tversky, A., 1995. Ambiguity Aversion and Comparative Ignorance. The Quarterly Journal of Economics 110, 585–603. doi:10.2307/2946693

Franke, N., Keinz, P., Klausberger, K., 2013. “Does This Sound Like a Fair Deal?”: Antecedents and Consequences of Fairness Expectations in the Individual’s Decision to Participate in Firm Innovation. Organization Science 24, 1495–1516. doi:10.1287/orsc.1120.0794

Franke, N., Lettl, C., Roiser, S., Tuertscher, P., 2014. “Does God Play Dice?” - Randomness vs. Deterministic Explanations of Crowdsourcing Success. Academy of Management Proceedings 2014, 15164–15164. doi:10.5465/AMBPP.2014.235

Grant, R.M., Baden-Fuller, C., 2004. A Knowledge Accessing Theory of Strategic Alliances. Journal of Management Studies 41, 61–84. doi:10.1111/j.1467-6486.2004.00421.x

Guggenbichler, J.P., Stickler, G.B., 1975. [Past and present aspects of diarrheal disease in childhood. Clinical study and treatment (author’s transl)]. Infection 3, 189–201.

Haas, M.R., Criscuolo, P., George, G., 2015. Which Problems to Solve? Online Knowledge Sharing and Attention Allocation in Organizations. Academy of Management Journal 58, 680–711. doi:10.5465/amj.2013.0263

Hannan, M.T., Goldberg, A., Kovács, B., 2015. What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. American sociological review.

Hartig, J., 2011. Learning and innovation @ a distance an empirical investigation into the benefits and liabilities of different forms of distance on interactive learning and novelty creation in german biotechnology SMEs. Gabler Verlag, Wiesbaden.

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47, 153. doi:10.2307/1912352

Jeppesen, L.B., Lakhani, K.R., 2010. Marginality and Problem-Solving Effectiveness in Broadcast Search. Organization Science 21, 1016–1033. doi:10.1287/orsc.1090.0491

Kotha, R.R., George, G., Srikanth, K., 2012. Bridging the Mutual Knowledge Gap: Coordination and the Commercialization of University Science. SSRN Electronic Journal. doi:10.2139/ssrn.1534049

Lakhani, K.R., Jeppesen, L.B., Lohse, P.A., Panetta, J.A., CBS, C.B.S., 2007. The Value of Openness in Scientific Problem Solving.

Lane, P.J., Lubatkin, M., 1998. Relative absorptive capacity and interorganizational learning. Strategic Management Journal 19, 461–477. doi:10.1002/(SICI)1097-0266(199805)19:5<461::AID-SMJ953>3.0.CO;2-L

McCallum, A.K., 2002. MALLET: A Machine Learning for Language Toolkit.

McLure Wasko, M., Faraj, S., 2000. “It is what one does”: why people participate and help others in electronic communities of practice. The Journal of Strategic Information Systems 9, 155–173. doi:10.1016/S0963-8687(00)00045-7

Mitsuhashi, H., Greve, H.R., 2009. A Matching Theory of Alliance Formation and Organizational Success: Complementarity and Compatibility. Academy of Management Journal 52, 975–995. doi:10.5465/AMJ.2009.44634482

Mohanty, M.S., 2001. Testing for the specification of the wage equation: double selection approach or single selection approach. Applied Economics Letters 8, 525–529. doi:10.1080/135048500011957

Mowery, D.C., Oxley, J., Silverman, B., 1998. Technological overlap and interfirm cooperation: implications for the resource-based view of the firm. Research Policy 27, 507–523.

Nielsen, B.B., Nielsen, S., 2009. Learning and Innovation in International Strategic Alliances: An Empirical Test of the Role of Trust and Tacitness. Journal of Management Studies 46, 1031–1056. doi:10.1111/j.1467-6486.2009.00840.x

Nooteboom, B., Gilsing, V.A., Vanhaverbeke, W., Duysters, G., van den Oord, A., 2006. Network Embeddedness and the Exploration of Novel Technologies: Technological Distance, Betweenness Centrality and Density. SSRN Electronic Journal. doi:10.2139/ssrn.903743

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., van den Oord, A., 2007. Optimal cognitive distance and absorptive capacity. Research Policy 36, 1016–1034. doi:10.1016/j.respol.2007.04.003

Piezunka, H., Dahlander, L., 2015. Distant Search, Narrow Attention: How Crowding Alters Organizations’ Filtering of Suggestions in Crowdsourcing. Academy of Management Journal 58, 856–880. doi:10.5465/amj.2012.0458

Rehurek, R., Sojka, P., 2010. Software Framework for Topic Modelling with Large Corpora, in: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. ELRA, Valletta, Malta, pp. 45–50.

Rosenkopf, L., Almeida, P., 2003. Overcoming Local Search Through Alliances and Mobility. Management Science 49, 751–766. doi:10.1287/mnsc.49.6.751.16026

Rosenkopf, L., Nerkar, A., 2001. Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. Strategic Management Journal 22, 287–306. doi:10.1002/smj.160

Rothaermel, F.T., Boeker, W., 2008. Old technology meets new technology: complementarities, similarities, and alliance formation. Strategic Management Journal 29, 47–77. doi:10.1002/smj.634

Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision making. Journal of Risk and Uncertainty 1, 7–59. doi:10.1007/BF00055564

Spradlin, D., 2012. Are you solving the right problem? Asking the right questions is crucial. Harvard Business Review 90, 84–101.

Stuart, T.E., Podolny, J.M., 2007. Local search and the evolution of technological capabilities. Strategic Management Journal 17, 21–38. doi:10.1002/smj.4250171004

Terwiesch, C., Xu, Y., 2008. Innovation Contests, Open Innovation, and Multiagent Problem Solving. Management Science 54, 1529–1543. doi:10.1287/mnsc.1080.0884

Thomas, J.B., Sussman, S.W., Henderson, J.C., 2001. Understanding "Strategic Learning": Linking Organizational Learning, Knowledge Management, and Sensemaking. Organization Science 12, 331–345. doi:10.1287/orsc.12.3.331.10105

Tsai, W., 2001. KNOWLEDGE TRANSFER IN INTRAORGANIZATIONAL NETWORKS: EFFECTS OF NETWORK POSITION AND ABSORPTIVE CAPACITY ON BUSINESS UNIT INNOVATION AND PERFORMANCE. Academy of Management Journal 44, 996–1004. doi:10.2307/3069443

West, J., Salter, A., Vanhaverbeke, W., Chesbrough, H., 2014. Open innovation: The next decade. Research Policy 43, 805–811. doi:10.1016/j.respol.2014.03.001

Wuyts, S., Colombo, M.G., Dutta, S., Nooteboom, B., 2005. Empirical tests of optimal cognitive distance. Journal of Economic Behavior & Organization 58, 277–302. doi:10.1016/j.jebo.2004.03.019

Tables

Table 1: description of variables

Variable	Source	Description
Experience in Academia (years)	Survey	Self reported number of years spent in academia
Experience in Industry (years)	Survey	Self reported number of years spent in industry
Number of Patents	Survey	Self reported number of (co-) invented patents
Industrial Involvement	Survey	Self reported number of 10 possible interaction channels used in the past 3 years
Own Broadcast Experience	Survey	Dummy variable equal to 1 if respondent has participated in broadcast search before
RFP length	RFPs	Length in words of the RFP description
RFP source	RFPs	Dummy variable euql to 1 if RFP originated from NineSigma (Innocentive=0)
RFP similarity	RFPs	Similarity measure between respondent's papers and RFP (see text)
Location (Asia, Europe, US, other)	Bibliographic data (WoS)	Dummies indicating respondents nationality based on e-mail address top level domain
Number of Citations	Bibliographic data (WoS)	Count of citations to all papers (co-) authored by the respondent
Knowledge Breadth	Bibliographic data (WoS)	Number of topics covered by the respondent's publications (see text)

Table 2: variables of conjoint experiment

Attribute	Base Level	2nd Level	3rd Level	4th Level
Seeker Type	Public / Governmental (base)	SME	Large Corporation	
Seeker Location	Different Continent	Same Continent	Same Country	
Seeker Identity	Undisclosed	Disclosed		
IP Disclosure	Immediately in 1st Step of Submission	Only in 2nd Step after Negotiation		
Required Solution Maturity	Theoretical Proof	Reduction-to-Practice	Prototype	
Retained Publication Rights	Complete Ban	With Content Restrictions	With Time Delay	Without Restrictions
Retained Patent Rights	Complete Ban	Seeker Patent with Solver Inventorship	Solver Patent with Exclusive Licensing to Seeker	Solver Patent with Non-Exclusive Licensing to Seeker
Award Money	$10,000	$25,000	$50,000	$75,000

Table 3: selection model

Independent variables	**Bivariate Probit Selection Model** Parameter		(S.E.)
Second Stage DV: Evaluate Contract Details (y/n)			
Constant	1.369		(0.872)
Knowledge Distance to Focal RfP	-0.159	**	(0.068)
RfP Source (1=NineSigma)	0.067		(0.182)
RfP Length (words)	0.001		(0.001)
Experience in Academia (years)	0.022		(0.017)
Solver Location - Europe	-0.223		(0.179)
Solver Location - USA	0.168		(0.220)
Solver Location - Asia	-0.374		(0.285)
Solver Location - Other (base)			
First Stage DV: Inspect RfP (y/n)			
Constant	-1.516	***	(0.249)
Average Distance to all RfPs	-0.047		(0.047)
Knowledge Breadth	0.002		(0.002)
Ln(Number of Citations)	-0.110	***	(0.021)
Experience in Academia (years)	0.007		(0.007)
Solver Location - Europe	0.168	**	(0.074)
Solver Location - USA	-0.224	***	(0.084)
Solver Location - Asia	-0.224	***	(0.074)
Solver Location - Other (base)			
Disturbance Correlation Rho	-0.609	*	(0.348)
No of Obs. in First Stage	24,375		
No of Obs. in Second Stage	569		
Parameters (k)	17		
Log likelihood (2)	-3,084		
Log likelihood (k)	-2,994		
Chi-square	179	***	
McFadden R^2 (adj.)	0.023		

Two-tailed t -tests; * < 0.1, ** p < 0.05, *** p < 0.01

Table 4: descriptive statistics

Variable		Source	N	Mean	Std.	Min	Max	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1st Stage - Invited Scientists: N=24,375																				
(1)	Solver Location - Other (base)	WoS	24,375	0.1	0.2	0.0	1.0	-0.223	-0.126	-0.178	-0.088	-0.06	-0.06	0.0724						
(2)	Solver Location - Asia	WoS	24,375	0.4	0.5	0.0	1.0		-0.418	-0.592	-0.109	-0.06	-0.176	0.1553						
(3)	Solver Location - USA	WoS	24,375	0.2	0.4	0.0	1.0			-0.333	0.0391	0.0574	0.2008	-0.22						
(4)	Solver Location - Europe	WoS	24,375	0.3	0.5	0.0	1.0				0.1284	0.0464	0.0484	-0.017						
(5)	Experience in Academia (years)	WoS	24,375	8.2	2.9	0.0	12.0					0.2343	0.3431	-0.21						
(6)	Knowledge Breadth	WoS	24,375	62.5	10.6	18.0	129.0						0.1499	-0.436						
(7)	Ln(Number of Citations)	WoS	24,375	4.5	1.0	0.0	9.3							-0.468						
(8)	Average Distance to all RfPs	RfPs/WoS	24,375	2.6	0.5	1.3	4.8													
2nd Stage - Inspected RfP: N=569																				
(1)	Solver Location - Other (base)	WoS	569	0.1	0.3	0.0	1.0	-0.18	-0.10	-0.29	-0.16	-0.06	-0.07	0.05	0.00	0.06	-0.02	0.00	0.00	
(2)	Solver Location - Asia	WoS	569	0.3	0.5	0.0	1.0		-0.23	-0.67	-0.17	-0.16	-0.08	0.17	-0.07	0.04	0.03	0.23	0.00	
(3)	Solver Location - USA	WoS	569	0.1	0.3	0.0	1.0			-0.38	0.01	0.04	0.10	-0.22	-0.03	-0.03	-0.17	0.14	0.00	
(4)	Solver Location - Europe	WoS	569	0.5	0.5	0.0	1.0				0.23	0.15	0.04	-0.04	0.08	-0.05	0.09	-0.30	0.00	
(5)	Experience in Academia (years)	WoS	569	8.3	2.9	2.0	12.0					0.30	0.36	-0.22	0.03	0.02	-0.04	-0.05	0.00	
(6)	Knowledge Breadth	WoS	569	63.1	10.4	32.0	95.0						0.17	-0.46	0.03	-0.05	-0.07	-0.05	0.02	
(7)	Ln(Number of Citations)	WoS	569	4.3	1.0	1.4	7.1							-0.36	-0.06	-0.01	-0.16	0.14	0.01	
(8)	Average Distance to all RfPs	RfPs/WoS	569	2.7	0.5	1.5	4.3								-0.05	0.09	0.31	-0.01	-0.02	
(9)	RfP Length (words)	RfPs/WoS	569	141.2	66.5	74.0	439.0									-0.31	0.02	-0.04	0.00	
(10)	RfP Source (1=NineSigma)	RfPs/WoS	569	0.9	0.3	0.0	1.0										0.03	0.02	0.00	
(11)	Knowledge Distance to Focal RfP	RfPs/WoS	569	1.8	0.9	0.2	3.7											-0.04	0.00	
(12)	Inverse Mills Ratio (Lambda) 1 Stage	Estimate	569	2.3	0.5	1.4	3.6												0.94	
(13)	Inverse Mills Ratio (Lambda) 2 Stage	Estimate	569	0.0	0.9	-1.3	1.8													
3rd Stage - Completed Contract Decisions: N=229																				
(1)	Solver Location - Other (base)	WoS	229	0.1	0.3	0.0	1.0	-0.15	-0.13	-0.35	-0.18	-0.05	-0.09	0.12	-0.02	0.10	0.05	-0.10	-0.21	-0.02
(2)	Solver Location - Asia	WoS	229	0.2	0.4	0.0	1.0		-0.21	-0.55	-0.11	-0.15	-0.04	0.00	-0.07	0.15	0.03	0.79	0.74	-0.04
(3)	Solver Location - USA	WoS	229	0.2	0.4	0.0	1.0			-0.49	0.00	0.07	0.14	-0.17	-0.03	0.05	-0.15	0.16	-0.41	0.15
(4)	Solver Location - Europe	WoS	229	0.6	0.5	0.0	1.0				0.19	0.09	-0.02	0.06	0.09	-0.21	0.06	-0.69	-0.17	-0.07
(5)	Experience in Academia (years)	WoS	229	8.6	2.8	2.0	12.0					0.30	0.40	-0.20	0.05	-0.01	-0.11	-0.22	-0.32	0.04
(6)	Knowledge Breadth	WoS	229	64.1	10.6	37.0	90.0						0.20	-0.47	0.04	-0.07	-0.03	-0.24	-0.24	0.16
(7)	Ln(Number of Citations)	WoS	229	4.3	1.0	1.6	6.7							-0.40	-0.05	-0.05	-0.22	0.29	-0.05	0.14
(8)	Average Distance to all RfPs	RfPs/WoS	229	2.6	0.5	1.6	3.8								-0.09	0.13	0.34	0.03	0.22	-0.23
(9)	RfP Length (words)	RfPs/WoS	229	144.0	70.4	74.0	439.0									-0.44	0.10	-0.14	-0.11	-0.05
(10)	RfP Source (1=NineSigma)	RfPs/WoS	229	0.9	0.3	0.0	1.0										0.00	0.14	0.04	0.10
(11)	Knowledge Distance to Focal RfP	RfPs/WoS	229	1.7	0.8	0.2	3.5											0.19	0.54	-0.09
(12)	Inverse Mills Ratio (Lambda) 1 Stage	Estimate	229	2.9	0.3	2.4	3.6												0.73	0.01
(13)	Inverse Mills Ratio (Lambda) 2 Stage	Estimate	229	1.1	0.2	0.7	1.8													-0.11
(14)	Own Crowdsourcing Experience	Survey	229	0.2	0.4	0.0	1.0													
(15)	Industrial Involvement	Survey	229	3.7	2.4	1.0	10.0													
(16)	Ln(Number of Patents)	Survey	229	1.4	1.2	0.0	5.4													

Table 5: latent class model fit

Model	Specification	Npar	LL	BIC(LL)	McFadden R^2
(1)	1 latent class	29	-1,107.96	2,373.51	0.111
(2)	2 latent classes	59	-984.43	2,289.44	0.210
(3)	3 latent classes	89	-910.43	2,304.45	0.269
(4)	2 latent classes & restricted covariates of baseline adoption	48	-995.15	2,251.12	0.201
(5)	... & including knowledge distance as a covariate of class membership	49	-992.09	2,250.44	0.204
(6)	... & restricted award money attribute to me metric	45	-1,005.35	2,255.21	0.193
(7)	... & restricted award money attribute to be monotonic increasing	47	-993.86	2,243.10	0.203

Table 6: latent class model part 1

	Covariates of Baseline Adoption Utility		
Independent variables	Parameter		(S.E.)
Own Crowdsourcing Experience	0.594	*	(0.313)
Industrial Involvement	-0.075		(0.053)
Ln(Number of Patents)	-0.257	**	(0.114)
Ln(Number of Citations)	-0.619		(0.377)
Knowledge Breadth	0.030	*	(0.017)
Experience in Academia (years)	0.038		(0.054)
Solver Location - Europe	0.964		(1.090)
Solver Location - USA	-3.151	***	(0.737)
Solver Location - Other	-1.057		(0.692)
Solver Location - Asia (base)	0.000		
Inverse Mills Ratio (Lambda) 2 Stage	-7.424	***	(2.865)
Inverse Mills Ratio (Lambda) 1 Stage	8.319	**	(3.768)
	Covariates of Class 2 Membership		
Independent variables	Parameter		(S.E.)
Constant	-0.203		(0.397)
Knowledge Distance to Focal RfP	0.673	***	(0.240)

Two-tailed t -tests; * < 0.1, ** p < 0.05, *** p < 0.01

Table 7: latent class model part 2

RfP - Contract Attributes	Attribute Significance			Class Differences			Contract Preferences: Class 1: "close"		Class 2: "distant"	
	Wald(0)	df		Wald(=)	df		Abs. Import.	Rel. Import.	Abs. Import.	Rel. Import.
Seeker Type	3.4	4		2.7	2		0.61	8%	0.05	1%
Seeker Location	11.5	4	**	7.7	2	**	0.94	12%	0.10	3%
Seeker Identity	6.3	2	**	0.5	1		0.42	5%	0.21	5%
IP Disclosure	11.9	2	***	11.9	1	***	0.99	13%	0.15	4%
Required Solution Maturity	29.0	4	***	1.0	2		0.89	11%	0.53	13%
Retained Publication Rights	39.8	6	***	2.4	3		1.34	17%	0.82	20%
Retained Patent Rights	36.7	6	***	14.1	3	***	1.44	18%	0.82	20%
Award Money	100.1	4	***	36.8	3	***	1.18	15%	1.43	35%

Two-tailed t -tests; * < 0.1, ** p < 0.05, *** p < 0.01

Table 8: latent class model part 3

RfP - Contract Attributes	Contract Preferences: Class 1 "close"; size=29.37% Parameter	(S.E.)	Marginal	(S.E.)	WTP	Class 2 "distant"; size=70.63% Parameter	(S.E.)	Marginal	(S.E.)	WTP
Seeker Type										
Large Corporation	0.102	(0.143)	0.023	(0.032)	$ 5,594	0.053	(0.144)	0.012	(0.032)	$ 2,388
SME	0.607	(0.380)	0.135	(0.084)	$ 33,327	0.030	(0.359)	0.007	(0.080)	$ 1,373
Public / Government (base)										
Seeker Location										
Same Country	-0.027	(0.130)	-0.006	(0.029)	$ -1,484	0.104	(0.150)	0.023	(0.033)	$ 4,721
Same Continent	0.913 ***	(0.296)	0.203 ***	(0.066)	$ 50,164	0.066	(0.348)	0.015	(0.077)	$ 3,004
Different Continent (base)										
Seeker Identity										
Disclosed	0.421	(0.283)	0.093	(0.063)	$ 23,106	0.207 *	(0.109)	0.046 *	(0.024)	$ 9,383
Undisclosed (base)										
IP Disclosure										
Only in 2nd Step	0.993 ***	(0.302)	0.221 ***	(0.067)	$ 54,538	-0.151	(0.113)	-0.034	(0.025)	$ -6,851
Immediately (base)										
Required Solution Maturity										
Prototype	-0.819 ***	(0.133)	-0.182 ***	(0.029)	$ -44,998	-0.535 ***	(0.143)	-0.119 ***	(0.032)	$ -24,227
Reduction-to-Practice	-0.886 ***	(0.349)	-0.197 ***	(0.078)	$ -48,696	-0.514	(0.338)	-0.114	(0.075)	$ -23,275
Theoretical Proof (base)										
Retained Publication Rights										
Without Restrictions	1.336 ***	(0.435)	0.297 ***	(0.097)	$ 73,413	0.820 ***	(0.162)	0.182 ***	(0.036)	$ 37,135
With Time Delay	0.977 ***	(0.151)	0.217 ***	(0.034)	$ 53,708	0.620	(0.433)	0.138	(0.096)	$ 28,091
With Content Restrictions	1.084 ***	(0.435)	0.241 ***	(0.097)	$ 59,587	0.384 **	(0.172)	0.085 **	(0.038)	$ 17,417
Complete Ban (base)										
Retained Patent Rights										
Non-Exclusive Licensing	0.376	(0.445)	0.084	(0.099)	$ 20,683	0.817 ***	(0.180)	0.182 ***	(0.040)	$ 37,022
Exclusive Licensing	1.287 ***	(0.149)	0.286 ***	(0.033)	$ 70,693	0.451	(0.497)	0.100	(0.110)	$ 20,443
Inventorship	1.440 ***	(0.432)	0.320 ***	(0.096)	$ 79,100	0.238	(0.161)	0.053	(0.036)	$ 10,783
Complete Ban (base)										
Award Money										
$75,000	1.183 ***	(0.296)	0.263 ***	(0.066)		1.435 ***	(0.168)	0.319 ***	(0.037)	
$50,000	1.183 ***	(0.169)	0.263 ***	(0.038)		1.115 ***	(0.296)	0.248 ***	(0.066)	
$25,000	0.000	(0.000)	0.000	(0.000)		0.871	(0.170)	0.194	(0.038)	
$10,000 (base)										
Constant										
Baseline Adoption Utility	16.733 **	(7.885)				13.169 *	(7.853)			

Two-tailed t -tests; * < 0.1, ** p < 0.05, *** p < 0.01