07-050

The Value of Openness in Scientific Problem Solving

Karim R. Lakhani
Lars Bo Jeppesen
Peter A. Lohse
Jill A. Panetta

The Value of Openness in Scientific Problem Solving

Karim R. Lakhani[1], Lars Bo Jeppesen[2], Peter A. Lohse[3] & Jill A. Panetta[3]

October 2006

[1] *Harvard Business School, Soldiers Field, Boston, MA, 02163, USA*[*]

[2] *Copenhagen Business School, Kilevej 14A, DK 2000 Frederiksberg, Denmark*[**]

[3]*InnoCentive.com, 35 New England Business Center, Andover, MA 01810-1071, USA*

[*] Corresponding author
Karim R. Lakhani's research was funded by a doctoral fellowship from Canada's Social Science and Humanities Research Council. L.B. Jeppesen was supported by the Research Centre on Biotech Business at Copenhagen Business School.

Abstract

Openness and free information sharing amongst scientists are supposed to be core norms of the scientific community. However, many studies have shown that these norms are not universally followed. Lack of openness and transparency means that scientific problem solving is constrained to a few scientists who work in secret and who typically fail to leverage the entire accumulation of scientific knowledge available.

We present evidence of the efficacy of problem solving when disclosing problem information. The method's application to 166 discrete scientific problems from the research laboratories of 26 firms is illustrated. Problems were disclosed to over 80,000 independent scientists from over 150 countries.

We show that disclosure of problem information to a large group of outside solvers is an effective means of solving scientific problems. The approach solved one-third of a sample of problems that large and well-known R & D-intensive firms had been unsuccessful in solving internally. Problem-solving success was found to be associated with the ability to attract specialized solvers with range of diverse scientific interests. Furthermore, successful solvers solved problems at the boundary or outside of their fields of expertise, indicating a transfer of knowledge from one field to others.

Introduction

Openness and unrestricted information sharing amongst scientists have been identified as institutional norms that are critical to scientific progress [1] and the key distinguishing features that separate the scientific community from other spheres of activity [2]. However, empirical studies of the behavior of scientists have shown that career [3], publication priority [4], intellectual property and financial concerns [5,6] often trump openness to the potential detriment of overall scientific advancement. For example, 47 per cent of academic geneticists who asked other researchers for additional information or materials regarding published research reported that at least one of their requests had been denied in the preceding three years [5]. Another study showed that only 14 per cent of experimental biologists were willing to talk openly about their current research with other colleagues [7]. The result of this lack of openness is that scientific problem solving activity is constrained and fails to adequately leverage the larger accumulation of knowledge amongst the wider scientific community.

More recently, free and open source software communities have demonstrated that actually practicing the norms of openness and information sharing in a peer-production setting can result in the creation of complex technological products that approach, and sometimes rival, the scope and quality of similar products produced by proprietary efforts [8,9]. The robustness of open information sharing and collective production has been demonstrated in fields as diverse as the creation of encyclopedia entries on science-related topics [10], bioinformatics [11] and cultural products [12]. These initiatives have shown that openness and collaboration before, during, and after problem solving efforts can harness the knowledge of many towards creating unique solutions.

Here we show that the opening up of information about difficult and unsolved scientific problems to a large group of unknown "outsiders" can be an effective problem solving strategy. Most problem solvers extensively use prior experience and knowledge in their attempts at solving problems, resulting in a "local search" of the potential solution space [13,14]. Opening up the search process and broadcasting problem information to outsiders can alleviate the negative effects of local search. We call this problem solving approach "broadcast search." The premise of broadcast search is the central insight that knowledge is unequally and widely distributed in society [15] and that the locus of innovation and problem solving shifts to where knowledge is stickiest (i.e. difficult to access or move)[16].

Our study finds that the broadcast of problem information to outside scientists results in a 29.5% resolution rate for scientific problems that had previously remained unsolved inside the R & D laboratories of well-known science-driven firms. Problem solving success was associated with the ability to attract specialized scientists with diverse scientific interests. Furthermore, successful solvers created solutions to problems that were on the boundary or outside of their fields of expertise, showing that openness in science can trigger the transfer and transformation of knowledge from one scientific field to other. We also found that solvers mainly relied on information from previously developed solutions when attempting to solve broadcast problems, indicating a relatively efficient knowledge transfer mechanism. Finally, successful solvers were motivated to engage in problem solving effort by either intrinsic motivations or financial reward.

Setting for Studying Openness in Science via Broadcast Search

In this paper, we present evidence of the efficacy of broadcast search by illustrating its application to 166 discrete scientific problems from the research laboratories of 26 firms from 10 different countries between June 2001 and January 2005. The firms spanned

diverse industries, including agrochemicals, biotechnology, chemicals, consumer products, and pharmaceuticals. Most firms had initially tried to solve the problem within their own laboratories, with some logging several years of effort. The problems were posted on InnoCentive.com's (IC) website, whose business model is centered on broadcasting science problems. IC broadcasts scientific challenges to over 80,000 independent scientists from over 150 countries. Each scientific problem statement posted includes the problem's background and the solution requirements, as well as deliverables which outside solvers are expected to provide. Firms offer a pre-set monetary award for the "best" solution from all the potential solutions received. There is no advance compensation for effort in developing the solution, and the seeker firm may choose to award one or more prizes or none at all. Solution requirements for the problems are either "reduction to practice" (RTP) submissions, i.e., requiring experimentally validated solutions, such as actual chemical or biological agents or experimental protocols, or "paper" submissions, i.e., rationalized theoretical solutions codified through writing. (Further details on the broadcasting process and intellectual property issues are in the appendix)

Of the 166 problems posted, 58% required developing RTP solutions. In the remaining 42% of the cases a "paper solution" was sufficient. All problems offered a substantial financial award (mean: $29,689; range: US$2,000–$105,000). Solutions had to be delivered within a limited time (mean: 166 days; range: 14–554 days).

We conducted two types of analyses of the application of broadcast search to scientific problems. First, we analyzed the determinants of successful problem resolution by examining the problem characteristics and the types of outside scientists attracted to creating solutions. Information on problem characteristics and the types of scientists attracted was obtained from IC's databases. Second, we analyzed what determined

whether an outside scientist created a winning solution by examining his or her motivation and fields of expertise and the problem-solving process used. Information on outside scientists and their problem-solving processes was obtained via an online, web-based survey of individuals who had submitted solutions to problems and from IC's database. The survey was sent to 993 outside scientists and yielded a relatively high response rate of 35% (n=357) [17].

Results: The efficacy of Broadcast Search

Table 1 shows the overall performance of broadcast search-based scientific problem solving: 49 of the 166 problems were solved using this approach, yielding a 29.5% resolution rate (The appendix contains descriptions of all the problems). On average, 240 (sd: 195, range: 19-1058) individuals examined each detailed problem statement and 10 (sd: 14, range: 0-103) individuals submitted solutions for evaluation. In 71% of the solved cases, only one award was made, to a single solver who provided a workable solution. In the remaining 29% of the solved cases, multiple awards were given to multiple solvers (range: 2–5). Overall, 75 solution awards were given out. Our data also show very few repeat winning solvers, with 87.5% of winning solvers winning just once and 8% winning twice. Two contract research labs won three and four times, with different individuals from the labs leading the problem-solving efforts.

INSERT TABLE 1

What Explains Which Scientific Problems Get Solved?

Table 2 shows the logit regression results of the likelihood of a problem being solved as a function of its characteristics (solution requirement [RTP or theoretical], award size, time window to solve problem) and the characteristics of the scientist base that each problem

attracted (total number of would-be problem solvers, number of solution submissions, heterogeneity of interests of scientists and generalist/specialist orientation of scientists). The strongest and most significant effect relates to the presence of heterogeneous scientific interests amongst scientists submitting solutions. At registration time with IC, would-be problem solvers indicate their scientific interests from 56 options – they can select as many or as few as they prefer. We find that, the more heterogeneous the scientific interests attracted to the solver base by a problem, the more likely the problem is to be solved.

Most organizations have limited access to such a range of heterogeneous problem solving perspectives and algorithms. The case for the need of a pharmaceutical firm to find clinically meaningful biomarker useful for identifying a specific patient population exemplifies this well. The scientific team inside the firm had expended significant time and resources to obtain a solution, however, their internal efforts were not successful as they had followed limited alternative paths. The broadcast of this problem triggered interest from 739 solvers from over 20 countries representing over 15 distinct fields of expertise. The problem ultimately received 30 very different solution proposals of which the wining solution was developed by a scientist from Argentina with a background in molecular biology. In another case, an aerospace physicist, a small agribusiness owner, a transdermal drug delivery specialist and an industrial scientist all submitted winning solutions to the same scientific problem: the identification of a polymer delivery system.

Figure 1 shows that, controlling for all other variables, a one-standard-deviation increase from the mean in the number of scientific interests in the solver base increases the probability of successful problem resolution by 39%.

INSERT TABLE 2 and Figure 1

We also find that the average number of scientific interests per solver per problem is significantly and negatively correlated with solvability. This implies that problems that attract solvers who indicate relatively fewer scientific interests, i.e. more specialized, are more likely to be solved. It is interesting to note that the effect of number of submissions per problem is non-significant in our analyses. We speculate that this implies that diversity in scientific interests prevails over sheer number of solutions from similar fields.

Table 2 also shows that the number of days a particular problem is open for resolution is negatively and significantly correlated with problem solvability. The number of days a problem is open is an indication of problem complexity as assessed by the seeker firm. Since broadcast search is a non-traditional method of problem solving, we can expect some learning effects in participating seeker laboratories. Scientists inside firms may learn over time how to select and/or articulate problems for resolution by outsiders. We measured seeker learning by counting the number of previous problems a firm had broadcast with IC. Our results show a marginally positive effect of seeker learning.

Solver Profile and Solution Creation Process

Our web-based survey revealed that would-be problem solvers were well-educated, with a majority (65.8%) holding a Ph.D. Solvers reported spending, on average, 39.9 hours (sd: 86.7, range: 0.1 - 800) developing solutions; winning solvers reported spending more than twice as much time solving problems as non-winning solvers (winning solvers: 74.1 hours, non-winning solvers: 35.7 hours, p=0.009).

Only 10.6% of our respondents reported working in teams to solve the problem, with 7.5% of winners (n=3) and 11.4% of non-winners (n=36) indicating a team effort. Average team size was 2.8 members (s.d.: 1.6), with no significant difference in team

size for winning versus non-winning solvers. A vast majority of solvers (79.6%) also reported that they did not consult others (excluding team members, if any) in the development of their solutions, with 83.3% of winners and 73.8% of non-winners reporting no consultation with others.

To investigate the origins of the solutions being provided, we asked solvers to what degree their submissions built on pre-existing solutions from their own work and/or the work of other individuals. Overall we found that 72.5% of winning solvers stated that their submissions were partially or fully based on previously developed solutions, with 55% relying on their own prior work and 60% relying on the previously developed work of others. More than half the winning solvers (55%) also reported that they made major modifications to previously developed solutions during their submission process. This indicates that broadcast search leverages pre-existing knowledge and the creative (re)combination and transformation of knowledge in the solution generation process. (Details on the response patterns to this question are in the appendix)

Who Becomes A Successful Solver?

We studied the probability of a problem solver developing a winning solution as a function of their expertise, specialist vs. generalist orientation (number of scientific interests) and their motivations to participate in the problem solving effort. Table 3 shows the standardized coefficients of the relevant variables in a logit regression. In our survey, we asked the solvers to assess the distance between the problem and their own field of expertise. We found a positive and significant correlation between the self-assessed distance between the problem field and the solver's expertise and the probability of being a winning solver. The further the focal problem was from the solvers' field of expertise, the more likely they were to solve it.

At a first glance this finding appears to be puzzling. However the case of understanding anomalous research findings in a drug discovery program illustrates the mechanism. A firm's research and development laboratory did not understand the toxicological significance of a particular pathology that they had observed in a study. They consulted without success top toxicologists inside and outside the firm. They then broadcast their problem via IC and it was solved, within weeks, by a scientist with a Ph.D. in protein crystallography using methods common in her field. This particular solver would normally not be exposed to toxicology problems or solve such problems on a routine basis; however, in this case, she successfully applied common knowledge from crystallography to toxicology.

Figure 2 shows, controlling for all other variables, that there was a 10% increase in the probability of being a winning solver if the broadcast problem was assessed to be completely outside their field of expertise. Consistent with our finding about specialization (Table 2), we found a marginally significant negative correlation between the number of scientific interests expressed and the probability of being a winning solver. Thus, being more specialized (expressing fewer scientific interests) resulted in a higher probability of creating a winning solution.

INSERT TABLE 3 and Figure 2

Do motivations to participate in broadcast search-based problem solving impact whether a solver will create a winning solution or not? Questions regarding motivations to participate were derived from an examination of existing economics [18] and psychology [19,20] literatures. The literature review suggested that even though winning the award money was the most obvious reason to participate, social and work-related motivations like career and professional reputation concerns, and peer and work pressure to submit a

solution should not be ignored. Alternatively, solvers may have participated for the challenge and enjoyment of scientific problem solving; thus, intrinsic motivations need to be considered as well. Being the first to solve a scientific challenge and beat others is also a strong motivational driver for scientists [3]. Solvers might have been motivated to participate because they had free time/capacity or were simply bored in their current jobs. We asked our respondents to rate 16 items on various motivations for creating a solution and found that 10 of the motivation items loaded onto two separable factors that could be labeled intrinsic motivations and social/career motivations.

As Table 3 shows, the probability of being a winning solver is significantly and positively correlated with both a desire to win the award money and intrinsic motivations like enjoying problem solving and cracking a tough problem. Even though there was a substantial monetary prize for creating the best solution, the effect of intrinsic motivation is stronger and more significant. Table 3 also shows that reporting having free time to actually participate in the problem solving effort significantly and positively correlates with being a winning solver. Participating due to career and social motivations or to beat others to solving the problem was negatively correlated with winning.

Discussion

We have demonstrated in this paper that openness regarding current scientific problems via the broadcasting of problem information to a diverse community of solvers can yield effective solution rates. We do not yet have an empirical basis for comparing this outcome with the effectiveness of traditional problem solving activities within academic or commercial laboratories for similar discrete problems. However, recall that many of the R & D laboratories posting these problems had been unsuccessful in creating solutions to these problems, thus implying a noteworthy outcome.

Our most counter-intuitive finding was the positive and significant impact of the self-assessed distance between the problem and the solver's field of expertise on the probability of creating a winning solution. This finding implies that the farther the solvers assessed the problem as being from their own field of expertise, the more likely they were to create a winning submission. We reason that the significance of this effect may be due to the ability of "outsiders" from relatively distant fields to see problems with fresh eyes and apply solutions that are novel to the problem domain but well known and understood by them.

This is consistent with the findings of studies of idea generation in science showing that "outsiders" of a given scientific community are a likely source of new ideas and innovation [21]. The history of science has shown that innovative solutions to difficult scientific problems can arise when knowledge from one scientific discipline is applied to another [22]. More recently, Zhou et al. [23] reported dramatic reductions in computation time (up to 100 million times faster) when "tried and true" methods from material science were imported into synthetic biology through a collaboration between materials scientists, biologists and physicists. Openness and access to information about problems between fields thus appears to be important for scientific progress and is systematically achieved through problem broadcasting and openness.

The degree of openness in our investigation was relatively narrow. Outside solvers worked independently and did not share their knowledge and potential solutions with others who were also attempting to solve the problem. Furthermore, the final best solution was not revealed to others who had created unsuccessful submissions to the problems. However, it may be advantageous to bring diverse problem solvers together and encourage them to collaborate on solutions that leverage multiple knowledge domains. Mathematical modeling and computer simulations have indicated that groups

of diverse problem solvers have the potential to outperform groups of high-ability but homogenous problem solvers [24,25]. Empirical evidence from software writing contests has shown that inclusion of random diverse collaborators in problem solving can increase computation performance by a factor of 10 to 100 [26]. It is reasonable to think that an open-source-like setting with transparency, access, and collaboration throughout the scientific problem-solving process has the potential to deliver even higher problem resolution rates.

The relative effectiveness of openness via broadcast search also implies that a systematic inclusion of radically diverse perspectives and heuristics in scientific problem-solving attempts may offer advantages over within-field attempts at problem solving which may be yielding "normal science" results [27]. This implies that scientific research not only needs to be open ex-post, i.e., upon publication of results, but also ex-ante, i.e., during scientific problem solving, allowing for various perspectives during the solution development process. However, achieving this level of openness and "outside" engagement in scientific problem solving may be a significant challenge. Many organizations engaged in research might be reluctant to reveal problem information to outsiders for fear of revealing proprietary research programs and activities. Furthermore, institutional norms like publication priority, promotions, grants, prizes and tenure typically reward individual or small team accomplishment. Thus, achieving true openness and collaboration will require change in the mindsets of both scientists and lab leadership. However, as our results suggest, opening up the scientific problem solving process can yield innovative technical solutions, increase the probability of success in science programs and ultimately boost research productivity.

Methods

Statistical Methods

We used logit regression models to determine the size and strength of relationship between dependent and independent variables. A logit model regression model is appropriate when the outcome variable is binary and is categorical (i.e. the problem was solved or not solved (table 2); solver had a winning solution or not a winning solution (table 3) [28]. The logit model is non-linear with an assumption that errors are normally distributed with a variance of the errors equal to $\Pi^2/3$ [29]. Regressions were computed using robust estimates for the standard errors thus allowing the estimates of the standard errors to be "robust" to failure to meet assumptions of normality and homogeneity of variance of the residuals. All of the analyses were conducted on Stata Version 8.

Data Sources and Variable Construction

Regression 1 – Which Problems Got Solved (Table 3)

The data for the regression correlating which problems got solved was obtained via access to IC's database. IC provided us with all salient information about each of the problems including solution requirements (RTP vs. paper), scientific discipline, seeker firm (anonymized), award value, days a problem was open for submission and size of IC's solver network over time. In addition, IC also provided us with anonymized information about the scientific interests of the solvers who submitted solutions. At registration time with IC, solvers are asked about their scientific interests from a list of 56 options spanning chemistry and applied sciences and the life sciences. The scientific interest information helped us to understand the types of solvers that were being attracted to the various problems and to analyze the intellectual diversity of the solver base.

Dependent variable:

Was Problem Solved?: Value = 1 if a solution reward was given out. Value = 0 if no solution award is given out.

Independent variables

RTP Problem Type: Value = 1 if the solution requirement for the problem was a reduction to practice submission. Value = 0 if the solution requirement was a paper submission,

Award Value: Actual value in US dollars for the award money for the problem being successfully solved.

Days Problem Open: The time window in days between the broadcast of the problem and the submission deadline.

Previous problems posted by seeker firm: The total number of previous problems broadcasted by the seeker firm on IC. Summed from 30 days prior to the post of the current problem.

Solver base size: Total number of registered users on IC website at the time of the posting of the problem.

Number of submissions: Number of submissions received at the end of the time window of a problem.

Distinct scientific interests attracted: At registration time with IC, solvers were asked about their scientific interests from a list of 56 options spanning Chemistry and Applied

Sciences and the Life Sciences. This variable consists of counting the total number of distinct (unique) scientific interests from the solvers who submitted a solution to the problem. Double counts of same the scientific interests by different solvers were eliminated. The higher the number the more unique scientific interests represented in solving the problem.

Generalist orientation of the solver: This variable consists of first summing the raw count of scientific interests indicated by the solvers who submitted a solution to the problem and then dividing this sum by the total number of solvers who submitted a solution. This thus creates the average number of scientific interests per solver per problem. The higher the number the larger the average number of interests per solver per problem and the more generalist an orientation of the solver community that is creating a solution.

The appendix contains the correlations table for the variables and the complete regression table.

Regression 2 – Who Becomes a Winning Solver (Table 4)

We wanted to understand how solvers came up with a solution and the determinants of a solver being able to successfully create a "winning" solution. Information on solvers and the problem solving process was obtained via an online, web-based survey of individuals who had submitted solutions to problems (A copy of the survey instrument is in the supplementary appendix). To test the reliability of the survey we conducted a pilot test survey with two current IC solvers and three individuals with similar backgrounds, such as a PhD in a scientific discipline. The resulting survey was administered in cooperation with IC and took about 20 minutes to complete.

Each solver received a customized email from IC's Chief Scientific Officer. The email asked the solvers to respond to the survey by reminding them of a specific problem for which they had attempted to create a solution along with the date of their submission to IC. Solvers who had created submissions to multiple problems were asked about their most recent submission. Those who had been successful in at least one attempt were asked to respond to the survey with regard to their most recent winning submission. Most solvers also had the ability to review the detailed problem statement and their submission on their personal account space on IC's website. The survey was sent to 993 individuals and yielded a relatively high response rate of 35.9 percent (n = 357) [17]. In all, 68 percent of the winning solvers and 34 percent of the non-winning solvers responded to our survey. We checked for non-response bias in our survey by comparing award values, days a problem was open, solvers' scientific interests and problem types for respondents and non-respondents. The comparisons yielded no significant differences, indicating that the survey sample adequately represented the IC solver population.

Dependent variable

Who becomes a winning solver? This data was available from the InnoCentive Database per problem. For each individual who responded to our survey we had information if they had won the prize award. Variable =1 if solver won an award for their submission. Variable = 0 if solver did not win an award for their submission.

Independent Variables:

Interest Count: Number of scientific interests indicated by solver when first registering with IC – from a list of 56 options.

Problem distance from field of expertise: Based on the answer to the following survey question: Is the particular challenge: "1 – inside your field of expertise, 4 – at the boundary of your field of expertise, 7 – outside your field of expertise". Respondents could choose any value between 1 and 7.

Motivations: The web survey asked the following question about motivations to participate in solving a challenge for IC: "*There are many reasons for participating in an InnoCentive Challenge. Tell us how true the following statements are for you. Please answer all items. I submitted a solution: (1-Not true at all, 4-Somewhat true, 7-Very true).*" Table 2 in the supplementary information appendix contains the specific items used for motivations and provides a factor analysis to group the items. The two factors, intrinsic motivation and social and work-related motivation, were developed from multiple items were constructed by first standardizing (transforming them so that mean = 0 and variance = 1) each of the items and then added and averaged and then further standardized.

RTP Problem Type: Value = 1 if the solution requirement for the problem was a reduction to practice submission. Value = 0 if the solution requirement was a paper submission.

Time to develop solution: Time in days as reported by solvers required to create a solution.

The appendix contains the correlations table for the variables and the complete regression table.

Duality of Interest: Two of the secondary authors: Lohse & Panetta are employees of the firm where the data for this study are obtained. The research design, data collection and analysis were done by the first authors (Lakhani and Jeppesen) who are affiliated with academic institutions.

References

1. Merton RK ([1942] 1973) The sociology of science: Theoretical and empirical investigation. Chicago: University of Chicago Press.
2. Mulkay M (1975) Three models of scientific development. Sociological Review 23: 509-526.
3. Stephan PE, Levin SG (1992) Striking the mother lode in science: the importance of age, place, and time. New York: Oxford University Press. xiii, 194 p.
4. Hagstrom WO (1974) Competition in Science. American Sociological Review 39: 1-18.
5. Campbell EG, Clarridge BR, Gokhale NN, Birenbaum L, Hilgartner S, et al. (2002) Data withholding in academic genetics - Evidence from a national survey. Jama-Journal of the American Medical Association 287: 473-480.
6. Grushcow JM (2004) Measuring secrecy: A cost of the patent system revealed. Journal of Legal Studies 33: 59-84.
7. Walsh JP, Hong W (2003) Secrecy is increasing in step with competition. Nature 422: 801-802.
8. von Hippel E (2005) Democratizing Innovation. Cambridge, MA: MIT Press.
9. Feller J, Fitzgerald B, Hissam S, Lakhani KR, editors (2005) Perspectives on Free and Open Source Software. Cambridge: MIT Press.
10. Giles J (2005) Internet encyclopedias go head to head. Nature: 900-901.
11. Benkler Y (2004) Commons-based strategies and the problems of patents. Science 305: 1110-1111.
12. Lessig L (2001) The Future of Ideas: The Fate of the Commons in a Connected World. New York, NY: Random House.
13. Simon H (1991) Bounded rationality and organizational learning. Organization Science 2: 125-134.
14. Simon HA, Newell A (1962) Computer Simulation of Human Thinking and Problem Solving. Monographs of the Society for Research in Child Behavior 27: 137-150.
15. Hayek FA (1945) The use of knowledge in society. American Economic Review 35: 519-530.
16. von Hippel E (1994) 'Sticky information' and the locus of problem solving: Implications for innovation. Management Science 40: 429-439.
17. Sheehan K (2001) E-mail Survey Response Rates: A Review. Journal of Computer Mediated Communication 6.
18. Frey B (1997) Not just for the money: an economic theory of personal motivation. Brookfield. VT: Edward Elgar Publishing Company.
19. Amabile TM (1996) Creativity in context. Boulder, CO: Westview Press.
20. Deci EL, Koestner R, Ryan RM (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125: 627-688.
21. Chubin DE (1976) The Conceptualization of Scientific Specialties. The Sociological Quarterly 17: 448-476.

22. Heisenberg W (1962) Physics and philosophy; the revolution in modern science. New York: Harper.
23. Zhou F, Grigoryan G, Lustig SR, Keating AE, Ceder G, et al. (2005) Coarse-Graining Protein Energetics in Sequence Variables. Physical Review Letters 95: 148103.
24. Hong L, Page SE (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. PNAS 101: 16385-16389.
25. Hong L, Page SE (2001) Problem Solving by Heterogeneous Agents. Journal of Economic Theory 97: 123-163.
26. Gulley N (2004) In praise of tweaking: a wiki-like programming contest. ACM Interactions 11: 18-23.
27. Kuhn T (1970) The structure of scientific revolutions. Chicago, Ill: University of Chicago Press.
28. Greene WH (2000) Econometric Analysis. Upper Saddle River, NJ: Prentice- Hall Inc.
29. Long JS, Freese J (2001) Regression Models for Categorical Dependent Variables Using Stata. College Station, TX: Stata Press.

Figure 1 - Marginal Impact of Increasing Intellectual Heterogeneity in Solver Base

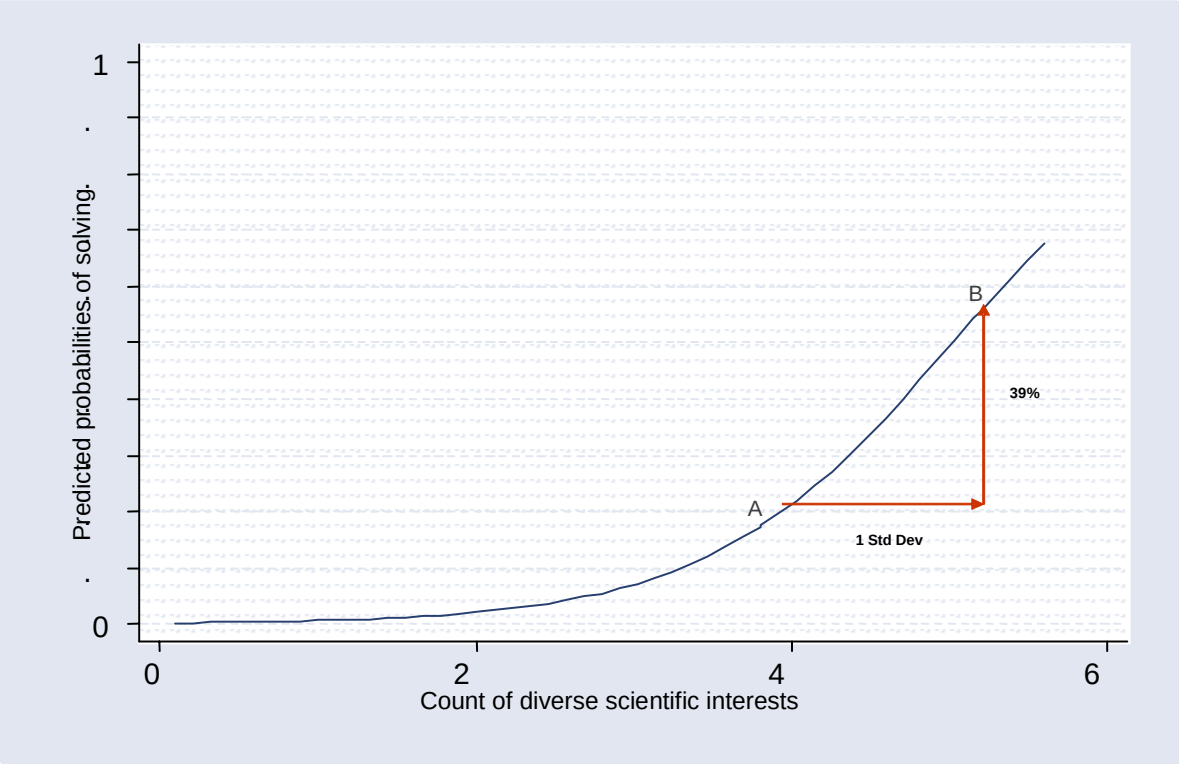
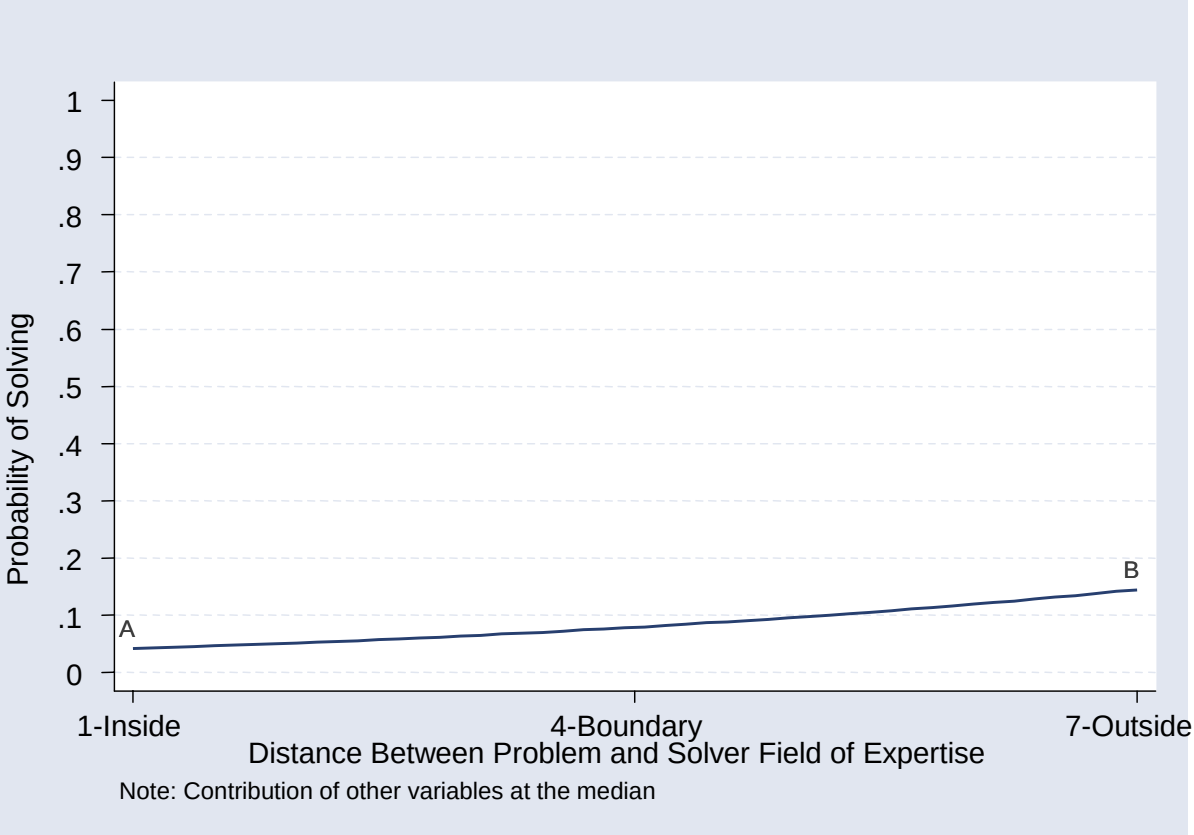


Figure 2 – Marginal Impact of Problem Distance and Probability of Creating a Winning Solution



Tables

Table 1 Overall Performance of Broadcast Search by Scientific Disciplines

Discipline of Problems Posted	Number of Problems	Solution Requirements: Theoretical |Reduction to Practice (%)	Average Award Value (USD$)	Average Number of People Expressing Interest	Average Number of Submissions	Number of Problems Resolved	Solving Rate (%)
Life Sciences							
Biochemistry	11	27|73	33181	269	5.7	0	0.0
Molecular Biology	7	43|57	15000	116	3	2	28.6
Biology	7	71|29	14571	236	9	5	71.4
Toxicology	3	67|33	12500	80	1	2	66.7
Structural Diversity	2	50|50	14000	228	4	1	50.0
Chemistry and Applied Sciences							
Synthesis	71	30|70	37408	223	9	22	31.0
Formulation	27	66|44	24666	220	10	8	29.6
Analytical	16	50|50	25375	314	13	1	6.3
Polymer	13	54|46	26884	254	8	1	7.7
Materials Science	4	50|50	25000	335	11	3	75.0
Other	5	60|40	22676	464	35	4	80.0
Total	166	42|58	29689	240	10	49	29.5

Table 2 Logit Regression on Problem Being Solved (N=132 Problems)

	Coefficient	Robust Standard Error	P-value
Problem Characteristics			
RTP Problem Type	0.566	0.413	0.171
Award Value	-0.418	0.449	0.352
Days Problem Open	-1.697	0.536	0.002
Seeker Firm Experience			
Previous problems posted by seeker firm	0.626	0.376	0.096
Solver community			
Solver base size	-1.897	1.134	0.094
Number of submissions	0.049	0.333	0.882
Types of Solvers Attracted			
Distinct scientific interests attracted	2.305	0.739	0.002
Generalist orientation of solvers	-1.638	0.628	0.009
Log Pseudolikelihood	-50.59		
Wald's Chi Square	44.29		
Df	19		
Pseudo R Square	0.39		

Controlled for year effects and scientific disciplines of problem

Table 3 - Logit Analyses Predicting Which Solver Submits A Winning Solution (N=295 Respondents)

	Coefficient	Robust Standard Error	P-value
Expertise			
Interest count (at registration)	-0.315	0.172	0.068
Problem distance from field of expertise	0.398	0,197	0.044
Motivations			
Win award money	0.503	0.214	0.019
Social and work related motivations	-0.398	0.221	0.072
Intrinsic motivations	0.668	0.220	0.002
Beating other solvers	-0.400	0.234	0.088
Unsatisfactory job	-0.126	0.265	0.635
Had free time	0.559	0.234	0.017
Control Variables			
RTP Problem Type	0.330	0.446	0.460
Time to develop solution	0.004	0.002	0.012
Log Pseudolikelihood	-85.62		
Wald's Chi Square	32.14		
Df	10		
Pseudo R Square	0.15		

Appendix to Paper

1. Supplementary Discussion - InnoCentive.com and how it works with seekers and solvers.

2. Supplementary Table - List of scientific problems that were broadcasted and Scientific Problems in Analysis

3. Supplementary Notes - Survey sent to solvers

4. Supplementary Table - Factor analysis to determine solver motivation clusters.

5. Supplementary Tables - Correlations for regression analyses.

6. Supplementary Tables - Full versions of Logit regressions.

7. Supplementary Tables - The source of solution information used by winning solvers.

1 – Information on InnoCentive.com

The data for our analysis was obtained in cooperation from InnoCentive.com (an independent venture of the Eli Lilly & Company pharmaceutical firm), whose business model is centered on broadcasting science problems. InnoCentive.com (IC) acts like a knowledge broker between "seeker" firms and over 80,000 independent and globally dispersed "solvers" from over 150 countries. IC's business model is contingent upon attracting seeker firms to post internal research problems on its website and encouraging solvers to examine and submit solutions to those problems for a potential monetary award.

Seeker firms work in consultation with IC's scientific operations staff to articulate their internal problems in a form that can be understood by an external scientific audience. Solution requirements for the problems are either "reduction to practice" (RTP) submissions, i.e. requiring original research data in the form of the actual chemical or biological agent or detailed experimental results, or "paper" submissions, i.e. requiring a theoretical submission with a validated research proposal. Problems are posted on IC's website along with a pre-set monetary award for the "best" solution and a deadline date for submissions. IC then broadcasts the problem to its entire solver base via email and invites them to participate in solving the problem. IC solvers do not work collectively to solve the problem; they do not know who else is working on the problem and how many solutions have been submitted. IC screens all submitted solutions to ensure that the problem requirements have been met and then forwards them to the seekers. Scientists from within the originating R&D laboratory assess the submissions and then inform IC if they have found one that meets their criteria. The seeker firm may chose not to award any prizes or to award multiple prizes.

Seekers and solvers remain anonymous to each other throughout the problem solving process. Care is taken to protect the intellectual property (IP) rights of seekers and solvers. When a problem is broadcasted, solvers initially see an abstract of the problem definition. If they are interested in seeing full details and requirements about the problem they have to first agree to a solver agreement which outlines the general contract terms, confidentiality, and intellectual property transfer clauses for accepted solutions. Solvers that submit solutions give a temporary license to the seeker firm and IC to evaluate their solution. If the solution is deemed acceptable by the seeker firm, the solver then receives the pre-announced award prize and transfers all IP rights to the seeker company. Before the transfer takes place IC contacts the solver's employer to ensure that they release any and all IP claims to the solution[1]. If the solution is not accepted the seeker firm relinquishes any rights to use the information provided in the submission in any future work and any IP remains with the solver. This is enforced by contracts between IC and the seeker firm, which allow IC the right to initiate audits on the output of the seeker firm's research laboratories.

[1] There have been only two cases where the employer of the solver refused to release the IP rights to a solution. We did not consider those two cases in our analyses.

2 – Supplementary Table 1 – List of Scientific Problems in Analysis

Problem Description	Discipline	Country of Originating Lab	Award Value (USD)	Problem Title
Cyclohexaneacetic acid	Synthesis	USA	30000	An efficient synthetic strategy for the listed cyclohexaneacetic acid derivative is required.
Challenge #2068	Synthesis	USA	80000	A novel synthetic route is required for the listed target molecule.
Substituted Piperazine	Synthesis	Belgium	50000	The following piperazine derivative is in need of an efficient synthetic strategy.
Substituted Cyclopentaneaceti c Acid	Synthesis	USA	30000	An efficient synthetic strategy for the following substituted cycopentaneacetic acid is required.
1-Bromo-6-fluoronaphthalene	Synthesis	UK	45000	The following disubstituted naphthalene is in need of an efficient synthetic strategy.
Chiral 2-Methyl-4-piperidone	Synthesis	Belgium	55000	The listed 4-piperidone is in need of an efficient synthetic strategy.
2-Bromo-6-fluoronaphthalene	Synthesis	UK	45000	The specific disubstituted naphthalene is in need of an efficient synthetic strategy.
Substituted indole	Synthesis	USA	65000	The following substituted indole is in need of an efficient synthetic strategy.
Substituted pyridine	Synthesis	USA	50000	Can you effectively synthesize this particular substituted pyridine?
Cyclopentenone	Synthesis	USA	25000	An efficient synthetic strategy for cyclopentenone is desired.
Challlenge # 3097	Synthesis	USA	100000	An efficient synthetic route is required for the following chemical structure.
Novel Synthetic Route	Synthesis	USA	50000	This ethyl ester derivative is in need of an efficient synthetic route.

Challenge # 3103	Synthesis	USA	90000	The following thiabicyclo ethyl ester is in need of an efficient synthetic strategy.
Challenge # 3106	Synthesis	USA	65000	The substituted indole listed is in need of an efficient synthetic strategy.
4-(4-Hydroxyphenyl) butanoic acid	Synthesis	USA	25000	An efficient synthetic strategy for the following butanoic acid derivative is required.
Bicycloketo ethyl ester	Synthesis	USA	80000	Can you efficiently synthesize this bicycloketo ethyl ester?
Substituted thiophene	Synthesis	USA	70000	The following substituted thiophene is in need of an efficient synthetic strategy.
Thiabicyclo ethyl ester.	Synthesis	USA	2000	Please provide a detailed retrosynthetic analysis with literature precedence for the following thiabicyclo ethyl ester.
Efficient Synthetic Strategy	Synthesis	USA	2000	Please provide a detailed retrosynthetic analysis with literature precedence for the following chemical structure.
Novel Synthetic Route	Synthesis	USA	2000	Please provide a detailed retrosynthetic analysis with literature precedence for the following chemical structure.
Deazaguanine ester	Synthesis	USA	90000	An efficient synthetic strategy is required for the deazaguanine ester.
Fmoc-L-Neo-Trp	Synthesis	USA	60000	Can you synthesize this protected unnatural amino-acid in its enantiomerically pure form?
Fmoc-D-2-Me-Trp	Synthesis	USA	75000	Can you synthesize this protected unnatural amino-acid in its enantiomerically pure form?
Fmoc-L-2-Me-Trp - enzymatic	Synthesis	USA	105000	Can you synthesize this protected unnatural amino-acid in its enantiomerically pure form?

(5-aza-benzofuran-7-yl) acetic amide	Synthesis	USA	75000	The following substituted aza-benzofuran is in need of an efficient synthetic strategy.
7-Formyl-Indole	Synthesis	USA	75000	The following substituted indole is in need of an efficient synthetic strategy.
Challenge # 40964	Synthesis	USA	65000	Can you devise the "best" synthetic method for the above transformation?
D-glucopyranose	Synthesis	USA	40000	The following substituted chiral azasugar is in need of an efficient synthetic strategy.
D-xylopyranose	Synthesis	USA	40000	The following substituted chiral azasugar is in need of an efficient synthetic strategy.
Substituted isoquinoline	Synthesis	USA	20000	The following substituted isoquinoline is in need of an efficient synthetic strategy.
Chiral Hexose-nucleoside	Synthesis	USA	60000	Can you synthesize the following chiral hexose-nucleoside?
4-AZIDO CHIRAL HEXOSE	Synthesis	USA	50000	Can you synthesize the following 4-azido chiral hexose-nucleoside?
4-hydroxypyrimidine	Synthesis	USA	15000	Can you effectively synthesize a 2-protected-2,6-diamino-4-hydroxypyrimidine?
Preserved Parenteral Suspension Placebo	Formulation	USA	100000	Can you formulate a simple, stable and safe injectable suspension placebo that has no pharmacological and biological activity?
Regio- and stereocontrolled tricyclic alcohols	Synthesis	USA	5000	Please provide a flexible retrosynthetic analysis that will allow convenient access to defined regio- and stereochemical isomers of this tricyclic alcohol.
Malononitrile - stable label	Synthesis	USA	15000	Can you synthesize uniformly labeled malononitrile?
4-nitroacetophenone - stable label	Synthesis	USA	25000	Can you synthesize 4-nitroacetophenone-[ring-13C6]?
Surfactant Analysis	Analytical	USA	40000	Can you develop and validate a robust

				analytical method to measure low-levels of a surfactant in a liquid formulation matrix?
cis-PTAP	Synthesis	USA	40000	An efficient synthetic strategy for the title compound is required.
Sulfoethoxylates	Synthesis	USA	2000	An economical synthesis route to sulfoethoxylates is required.
Oxidation of parrifins	Synthesis	USA	2000	An efficient synthesis strategy for conversion of long chain paraffins to near terminal long chain alcohols is required.
Branched alcohols	Synthesis	USA	2000	Devise the best synthesis strategy for conversion of tallow oil or other low cost oils to branched alcohols.
BTCA	Synthesis	USA	2000	An efficient synthesis to 1,2,3,4-butanetetracarboxylic acid (BTCA) is required.
Protein crosslinks	Medicinal Chemistry	USA	3000	Can you develop a novel paper proposal of a molecule to spontaneously break protein crosslinks?
Filtration of a Formulation	Formulation	USA	3000	An in-process test to confirm when sodium carboxymethylcellulose dissolution is complete and will filter properly is needed.
Properities of CMC	Formulation	USA	3000	Why does the viscosity of a liquid formulation containing sodium carboxymethylcellulose decrease during heat sterilization and why does this property continues to decrease on storage at 40 degrees Centigrade?
Stimulus to Elicit Urination by Untrained Rats of Either Sex	Biology	USA	2000	An economical, reproducible, noninvasive, stimulus to elicit on-demand urination by untrained rats of either sex is needed?
Paracrystalline Arrays	Toxicology	USA	5000	Please provide a novel mechanistic approach to define the pathogenesis

				and toxicologic significance of paracrystalline arrays in mitochondria.
Megamitochondria	Toxicology	USA	2500	Please provide a novel mechanistic approach to define the pathogenesis and toxicologic significance of megamitochondria in hepatocytes.
1-Azabicyclo [3.2.2] nonan-3-one	Synthesis	USA	70000	The following 1-Azabicyclo [3.2.2] nonan-3-one is in need of an efficient synthetic strategy.
1-Azabicyclo [3.2.1] octan-3-one	Synthesis	USA	65000	The following 1-Azabicyclo [3.2.1] octan-3-one is in need of an efficient synthetic strategy.
Vacuum Blood Collection System	Biochemistry	USA	10000	Can you manufacture a rapid, inexpensive, reproducible small vacuum blood collection system (tubes) with and without an anticoagulant?
N-Boc-7-azabicyclo [2.2.1] heptene	Synthesis	USA	60000	The following N-Boc-7-azabicyclo [2.2.1] heptene is in need of an efficient synthetic strategy.
Yeast molecular genetics (1)	Molecular Biology	USA	2000	A solver(s) is needed to create knockout strains of a S.cerevisiae strain. The solver must provide the knockout strain(s) and appropriate evidence of success.
Yeast molecular genetics (2)	Molecular Biology	USA	3000	A solver(s) is needed to create knockout strains of a S.cerevisiae strain. The solver must provide the knockout strain(s) and appropriate evidence of success.
Alkyl phenyl alkanols	Synthesis	USA	2000	Seeking novel synthesis of alkyl phenyl alkanols. Compounds have been reported in the literature.
Low Surface Energy Particles for Reduction of Friction	Material Science	USA	5000	The identification of small, cost effective, low surface energy, spherical particles for

				deposition on surfaces to reduce friction is needed.
Efficient synthetic route	Synthesis	USA	50000	The following molecule is in need of an efficient synthetic strategy.
Efficient synthetic route	Synthesis	USA	50000	The following molecule is in need of an efficient synthetic strategy.
A-MOE	Synthesis	USA	50000	Devise and execute a novel synthetic strategy that allows for the efficient synthesis of this compound.
G-MOE	Synthesis	USA	50000	Devise and execute a novel synthetic strategy that allows for the efficient synthesis of this compound.
In vitro Bone Formation Assay	Biology	USA	5000	Please provide a proposal for the development of a novel in vitro bone anabolic assay that can predict an in vivo bone formation response.
Chitosan Life Sciencespolymer	Synthesis	Belgium	75000	A cost effective synthetic or biosynthetic route to chitosan biopolymer is needed.
Picolinic acid (Derivative 2)	Synthesis	USA	25000	A regioselective synthetic route to a picolinic acid derivative is needed.
Picolinic acid (Derivative 1)	Synthesis	USA	25000	A regioselective synthetic route to a picolinic acid derivative is needed.
Yeast molecular genetics	Biochemistry	USA	5000	Create a knockout strain of S.cerevisiae strain. The scientist must provide the final and all intermediate strains created, as well as evidence that the genes were correctly knocked out.
Crosslinking Polysaccharides and Polycarboxylic acids	Synthesis	USA	50000	An efficient catalyst to esterify polysaccharides with polycarboxylic acids is needed.
Crosslinking Polysaccharides and Polycarboxylic acids	Synthesis	USA	3000	An efficient catalyst to esterify polysaccharides with polycarboxylic acids is needed.

Polymer analysis in surfactant matrices	Polymer	USA	5000	Fast and effective methods for polymer identification and analyses in high surfactant matrices are needed.
Trifluoro-lactate Derivative	Synthesis	UK	7000	Devise and execute an efficient synthetic pathway for the compound shown above.
Pyrrolo-pyrimidine	Synthesis	UK	10000	Devise and execute an efficient synthetic pathway for one of the compounds shown above.
Seeking Small Molecules Libraries (I)	Chemical Diversity	USA	18380	The Seeker is seeking to purchase quantities of heterocyclic molecules with MW < 650.
Analytical Method for Active Ingredient	Analytical	USA	20000	A simple, reliable, robust and reproducible analytical method for determining the concentration of an active ingredient in various product formulation matrices is needed.
Procedure to Develop Artificial Human Fluid	Formulation	Italy	15000	A procedure to develop artificial human fluid that can reliably simulate the corresponding real human fluid is needed.
Stabilization of liquid formulation	Formulation	Germany	5000	Stabilization of highly acidic liquid formulation is needed.
Purification of silicone based solvents	Technology	USA	10000	New recovery methods for purifying silicone based solvents is needed.
Life Scienceslogical Targets for Inflammation	Biology	USA	5000	Can you suggest five biological targets for INFLAMMATION?
Life Scienceslogical Targets for AntiLife Sciencestics	Biology	USA	5000	Can you suggest five biological targets for BROAD SPECTRUM ANTIBIOTICS?
Life Scienceslogical Targets for Obesity	Biology	USA	5000	Can you suggest five biological targets for OBESITY?
Life Scienceslogical Targets for Insulin-	Biology	Germany	30000	Please provide data and evidence to support the identity of the common

Releasing Compounds				efficacy target of 3 compounds that will be provided by the seeker.
Substituted Propionic Acid	Synthesis	USA	30000	An efficient synthetic route is required for the substituted propionic acid.
Amino Indanol	Synthesis	USA	30000	An efficient synthetic route is required for the following indanol derivative.
Incomplete Release of Active Ingredient	Formulation	USA	5000	A research proposal to understand the incomplete release of an active ingredient from a micro-encapsulated formulation is needed.
Novel colorant materials #1	Formulation	USA	5000	A novel material that can mimic the characteristics of the displayed reflectance spectrum is needed.
Novel colorant materials #2	Formulation	USA	15000	A novel material that can mimic the characteristics of the displayed reflectance spectrum is needed.
Elasticity improvement in textiles	Analytical	Germany	5000	A substance that has high binding affinity to textile fibers and can improve the elasticity of textiles is needed.
Synthesis of dipalmityl- or distearyl-diketene	Synthesis	Germany	10000	Can you find a suitable catalyst system to catalyze a reaction that produces distearyl-diketene or dipalmityl-diketene?
Synthesize hexamethylene-1	Synthesis	Germany	10000	Can you find a suitable catalyst system to catalyze a reaction that produces hexamethylene-1,6-diisocyanate?
TMBA (3	Synthesis	Italy	5000	A proposal for a cost-efficient synthesis of 3,4,5-tri-methoxy benzoic acid (TMBA) is needed.
Gallic Acid	Synthesis	Italy	5000	A proposal for a cost-efficient synthesis of gallic acid is needed.
DNA Extraction Method	Molecular Biology	USA	10000	A cost-effective DNA extraction method is required.

Burst Release Formulation	Formulation	USA	5000	A formulated product for burst release is needed.
Plant Selectable Marker	Molecular Biology	USA	35000	A combination of a gene and a chemical (or combination of chemicals) which together allow efficient selection of transformed plant cells in vitro and transformed plants in vivo is needed.
Plastid Selectable Marker	Molecular Biology	USA	10000	A combination of a gene and a chemical (or combination of chemicals) which will allow selection of plastids and cells containing the gene by application of the chemical during growth in vitro is needed.
Crosslinking Reaction for Polymers	Polymer	Germany	12000	A proposal identifying reactive groups for the desired crosslinking reaction of emulsion polymers is needed.
Calcium carbonate nanoparticles in water	Nanotechnology	Germany	35000	A suspension of "aggregate free" calcium carbonate nanoparticles in water is needed.
Immortalized Preadipocyte Cell Line	Molecular Biology	USA	35000	The Seeker is seeking an immortalized preadipocyte cell line.
Microbial strain for the production of an amino acid	Biochemistry	Germany	40000	We are seeking a microbial strain for the production of an amino acid.
Compounding Method	Polymer	Germany	7500	An improved and inexpensive compounding method is needed.
Compounds forming hydrogen-bonds	Polymer	Germany	5000	Compounds that form intermolecular and intramolecular hydrogen bonds are needed.
Cerium containing organic solution	Formulation	Germany	10000	An organic solution containing a Ce(III) or Ce(IV) compound is needed.
New Chem and Applied Sciencesical routes to a substituted	Synthesis	France	25000	A novel and cost-effective synthetic route to a substituted benzaldehyde is required.

benzaldehyde				
Novel “Green” additives	Material Science	France	50000	Identification of a novel additive with "greener"properties is needed.
Diagnostic test for Interstitial Cystitis.	Analytical	USA	40000	A sensitive and specific diagnostic test for interstitial cystitis is required.
Particle Size Reduction	Formulation	USA	55000	Methodology to reduce the particle size for a given material is needed.
Controlled Encapsulation and Release of Electrolyte	Formulation	USA	30000	A novel material that can encapsulate a saturated salt solution but release electrolyte upon dilution is needed.
pH Modification	Analytical	USA	30000	A material or combination that can produce pH-increase upon dilution is needed.
Full-Length cDNA Isolation	Molecular Biology	USA	10000	A method to isolate a full-length cDNA based on the 3' EST sequence is needed.
DNA inverted repeat analysis	Biochemistry	USA	20000	A method to detect inverted repeats in random transgenic DNA inserts is needed.
Food-Grade Polymer	Polymer	USA	35000	A food-grade polymer suitable for use as a delivery vehicle is required.
Decrease of Cr (VI) concentration	Formulation	USA	15000	A proposal identifying a novel method for decreasing the concentration level of Cr(VI) is needed.
Bubbling Action	Formulation	Germany	10000	Proposals for chemicals that create bubbling action are needed.
Formulation for a proLife Sciencestic powder	Formulation	USA	45000	Formulation for a probiotic powder that can be applied to food is needed.
Stable form of tetrasodium pyrophosphate	Formulation	USA	35000	A stable form of tetrasodium pyrophosphate is needed.
Synthesis of an acrylic acid polymer (2)	Polymer	Germany	15000	A paper proposal for a novel strategy for the synthesis of an acrylic acid polymer is needed.

Synthesis of an acrylic acid polymer (1)	Polymer	Germany	60000	Reduction to practice of a novel strategy for the synthesis of an acrylic acid polymer is needed.
Lowering of CO levels	Material Science	USA	35000	Identification of a media to lower CO level is required.
Substituted Benzenes	Synthesis	USA	35000	A novel catalytic process for formylation of substituted benzenes is needed.
Hedonics of Oral Chem and Applied Sciencesesthesis	Biochemistry	USA	55000	A novel method to study the hedonics of oral chemesthesis in non-human animals.
Synthesis of 2	Synthesis	USA	15000	A cost-effective synthesis of the title compound is required.
Porous carbohydrate resin	Polymer	USA	40000	An efficient method to produce a porous carbohydrate resin is needed.
Gel-forming polymer	Formulation	USA	40000	A gel-forming polymer to make water-based gels is required.
Water vapor barrier glue	Polymer	Germany	35000	Identification of a material that can function as a glue and provide a water vapor barrier is required.
Lactose Polymerization	Polymer	Denmark	25000	An enzyme capable of polymerization of lactose is needed.
2- Specific lipase of microbial origin	Biochemistry	Denmark	50000	A 2- Specific lipase of microbial origin is needed
Non toxic inhibitor for lipases	Medicinal Chemistry	Denmark	25000	A low cost, non toxic, reversible, and strong inhibitor for Lipase is needed.
Platelet Aggregometry Device	Technology\ Knowledge Aggrigation	USA	25000	Information on a device for measuring platelet aggregation is needed.
Additive to alter surface properties	Analytical	USA	40000	A hydrophobic additive for the alteration of surface properties is needed.
Enzyme Stabilizer	Formulation	USA	65000	An enzyme stabilizer at high pH is required.
Flash point elevation	Formulation	USA	45000	A formulation for an aqueous ethanolic solution with a flash

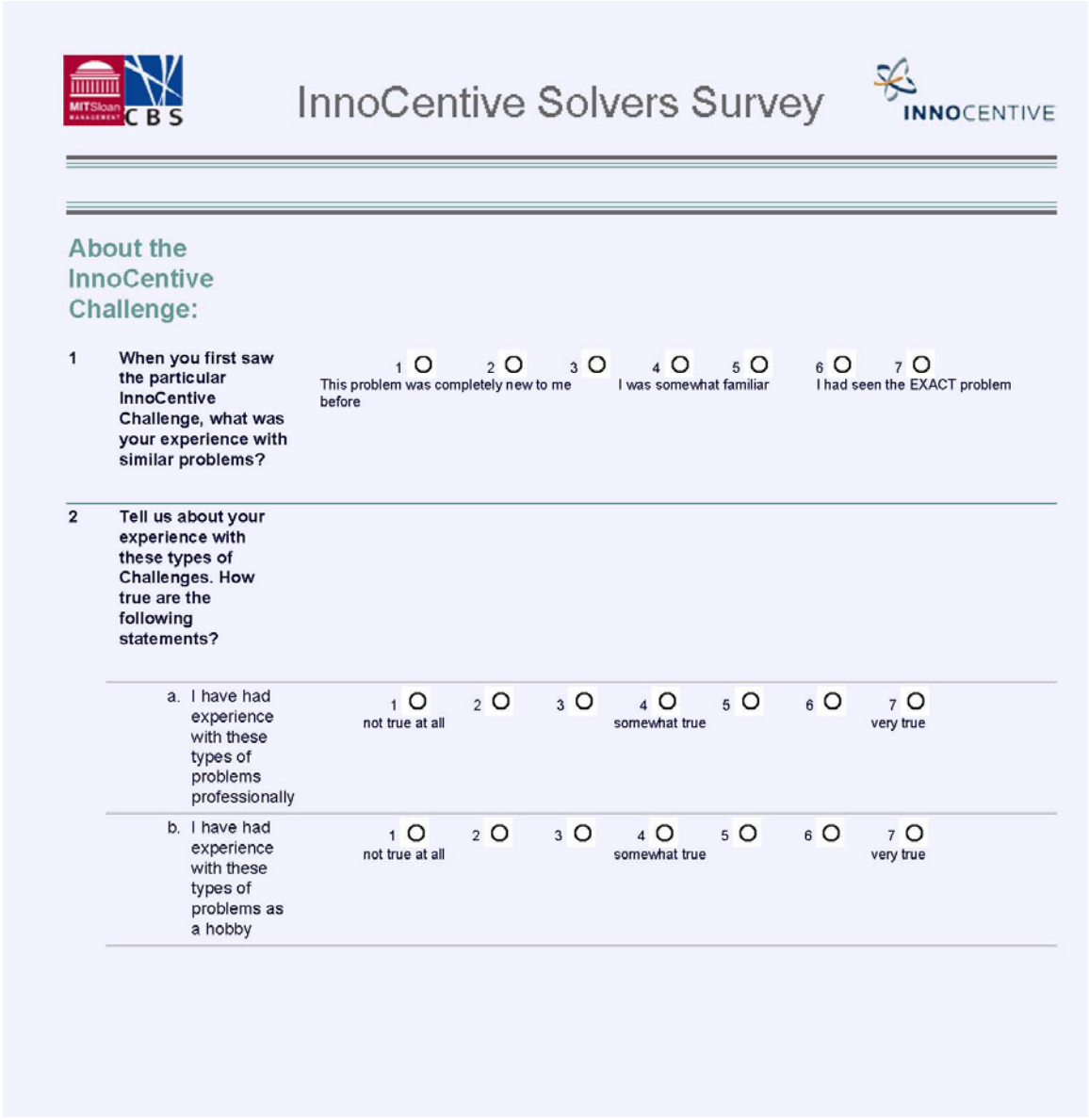
				point > 60 degrees C is needed.
Ethanol absorbents	Formulation	USA	30000	Materials that absorb or sequester ethanol are needed.
Retort stable form of Vitamin C	Formulation	USA	15000	A novel method for producing retort-stable Vitamin C is required.
Water Absorbent Material	Polymer	USA	50000	A superabsorbent material that absorbs 50-100 times its weight in the presence of high levels of salt is desired.
Method for peptide bond synthesis	Synthesis	USA	15000	A method for the preparation of peptidic compounds is needed.
Supplier for MgO	Formulation	USA	5000	Identification of a commercial source of high surface area magnesium oxide is needed.
High-throughput format for a Life Scienceslogical assay	Biology	USA	50000	High-throughput format for a biological assay is needed.
Selective removal of a protecting group	Analytical	Germany	12000	A proposal describing a method for selective removal of a formate protecting group is needed.
Method for Addition of a Salt	Technology/ Chem Eng	USA	25000	Identification of a method for addition of a phosphate based salt is needed.
Non-fluorinated oil and water repellent	Formulation	UK	45000	Identification of novel, non-fluorinated technologies/materials for oil and water repellancy is needed.
Improving Solution Appearance with Novel Dyes	Formulation	USA	30000	The Seeker desires a dye molecule that changes from transparent to a blue hue upon dilution in aqueous solution.
Visual Modification of an aqueous dispersion	Formulation	USA	20000	An efficient method for visual modification of an aqueous dispersion is

				needed.
Life Sciencesfilm Indicator	Analytical	USA	25000	A non-toxic procedure for biofilm detection is required.
Water Vapor Permeability	Analytical	USA	30000	Identification of a material that can provide desired water vapor permeability is required.
Synthesis of 3-difluoromethyl-1-methyl-4-pyrazole carboxylic acid	Synthesis	USA	10000	A proposal for a cost-effective large scale synthesis of 3-difluoromethyl-1-methyl-4-pyrazole carboxylic acid is needed.
Preservative Degradation	Formulation	USA	10000	Degradation Mechanism of Preservatives in Water-based Formulations is needed.
Alternate material to cyclododecane	Material Science	UK	10000	Identification of an alternative material to cyclododecane is needed.
Iminium ion synthesis from tertiary amines	Synthesis	USA	10000	A proposal seeking a novel synthetic route to substituted iminium ions using tertiary amines as reactants is needed.
New Phase Change Materials	Analytical	Germany	10000	New Phase Change Materials with high melting enthalpy and improved properties are needed.
Separation of tolualdehyde-acid adducts	Synthesis	USA	15000	A method for the separation of tolualdehyde-acid adducts is needed.
New applications for silane-functionalized polyolefins	Polymer	USA	15000	New applications for silane-functionalized polyolefins are needed.
Photo and Chem and Applied Sciencesical Passivation of Titanium Dioxide Nanopart	Analytical	USA	10000	Identification of a novel surface modification methodology for TiO_2 to affect specific material properties is needed.
Thiophene formation	Synthesis	USA	10000	A logical mechanism for thiophene formation in

				the Fluid Catalytic Cracking process is needed.
Metals removal from heavy petroleum fractions	Analytical	USA	10000	A theoretical proposal for a method using novel chemistry, catalysis, bio-agents or adsorbents is needed to remove metals from heavy petroleum fractions.
Film-forming polymer	Polymer	USA	45000	Identification of a material to form a film on a cellulosic substrate is needed.
Gametogenesis Inhibitor	Analytical	Canada	100000	An inhibitor to disrupt normal gametogenesis is needed.
Efficient synthesis of a Resorcinol Derivative	Synthesis	USA	15000	A theoretical proposal for an efficient synthesis route for a resorcinol derivative is needed.
Method to avoid skin sensitization	Biochemistry	USA	15000	A method to avoid contact sensitization from use of a transdermal patch is needed.
Seeking anti-nitration additive	Analytical	UK	15000	An anti-nitration additive that meets specific requirements is needed.
Seeking ion channel enzyme inhibitors	Biochemistry	USA	100000	Ion channel enzyme inhibitors are needed.
Reduce viscosity of a salt formulation	Analytical	USA	4000	An additive that can reduce the viscosity of a high-strength salt formulation is needed.
Chlorine Removal	Analytical	UK	15000	A process of chlorine removal is required.
Seeking formulation development partners from China and India	Formulation	USA	5000	The Seeker is seeking China- and India-based Solvers with proven capabilities in formulation development.
Detection of DNA sequences	Biochemistry	USA	10000	A theoretical proposal for rapid detection of DNA sequences without the use of PCR technology is needed.
Analytical Assay for Phytate	Biochemistry	USA	30000	An analytical assay for phytate is needed.
Inositol phosphate	Biochemistry	USA	30000	A method for inositol

derivatization				derivatization for quantitative analysis is needed.
Insect mutant line	Toxicology	USA	30000	An insect mutant line resistant to an insecticide is needed.

3 – Solver Survey



InnoCentive Solvers Survey

INNOCENTIVE

About the InnoCentive Challenge:

1 **When you first saw the particular InnoCentive Challenge, what was your experience with similar problems?**

1 ○ This problem was completely new to me before
2 ○
3 ○
4 ○ I was somewhat familiar
5 ○
6 ○
7 ○ I had seen the EXACT problem

2 **Tell us about your experience with these types of Challenges. How true are the following statements?**

a. I have had experience with these types of problems professionally

1 ○ not true at all
2 ○
3 ○
4 ○ somewhat true
5 ○
6 ○
7 ○ very true

b. I have had experience with these types of problems as a hobby

1 ○ not true at all
2 ○
3 ○
4 ○ somewhat true
5 ○
6 ○
7 ○ very true

	c. I have had experience with these types of problems as a student	1 ○ not true at all 2 ○ 3 ○ 4 ○ somewhat true 5 ○ 6 ○ 7 ○ very true
3	**Please tell us if this Challenge was**	1 ○ Inside your field of expertise 2 ○ 3 ○ At the boundary of your field of expertise 4 ○ 5 ○ 6 ○ 7 ○ Outside your field of expertise

About your Submission to InnoCentive:

4	**What was your situation when you encountered this Challenge?**	-----select a response from this pulldown menu-----
5	**Sometimes solutions build on previous work. Was your submission to this Challenge based on:**	
	a. A solution you had already developed in your own work with:	1 ○ No modifications 2 ○ 3 ○ 4 ○ Minor Modifications 5 ○ 6 ○ 7 ○ Major modifications ○ NA – This was not based on any of my previous work
	b. An existing solution you knew about that could solve the Challenge with:	1 ○ No modifications 2 ○ 3 ○ 4 ○ Minor Modifications 5 ○ 6 ○ 7 ○ Major modifications ○ NA – This was not based on anyone else's work
6	**How much time did it take you to develop your submission (please estimate hours of effort)**	☐ hours

7	**How much money (not including your own labor) did you spend in developing your submission (e. g.: Money spent on Reagents, Equipment etc)**	---------select currency--------
8	**Tell us about any additional resources you needed as you worked on this Challenge. How true are the following statements? I had to acquire access to:**	
	a. More laboratory equipment than I normally have access to	1 not true at all 2 3 4 somewhat true 5 6 7 very true
	b. More software than I normally have access to	1 not true at all 2 3 4 somewhat true 5 6 7 very true
	c. More library and literature resources than I normally have access to	1 not true at all 2 3 4 somewhat true 5 6 7 very true
	d. More specialty databases than I normally have access to	1 not true at all 2 3 4 somewhat true 5 6 7 very true

9	**Did you solve the Challenge as an individual or as a team?**	individual: ○ team: ○
	9a If you solved the Challenge as a team – how many people were on your team?	☐
10	**How many other people did you consult with in your problem solving effort (excluding those that were on your team in question 9a)?**	☐
11	**In your estimation, how many others could have developed a submission similar to yours?**	----------select a response from this pulldown menu----------

Your reasons for participating in the InnoCentive Challenge:

12	**There are many reasons for participating in an InnoCentive Challenge. Tell us how true the following statements are for you. Please answer all items. I submitted a solution:**	
	a. To win the award money	1 ○ not true at all 2 ○ 3 ○ 4 ○ somewhat true 5 ○ 6 ○ 7 ○ very true

b. Because others I know have participated before	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
c. Because someone suggested I participate in solving this Challenge	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
d. Because my boss asked me to work on it	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
e. To try to beat other InnoCentive solvers	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
f. Because my work/job at the time was not satisfying	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
g. Because InnoCentive told me about this Challenge	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
h. Because I enjoy solving these types of Challenges	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
i. To enhance my skills	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
j. To enhance my career prospects	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true
k. To impress my colleagues	1 ○ not true at all	2 ○	3 ○	4 ○ somewhat true	5 ○	6 ○	7 ○ very true

l.	Because I already knew how to get the solution	1 O not true at all 2 O 3 O 4 O somewhat true 5 O 6 O 7 O very true	
m.	For the intellectual challenge of solving this Challenge	1 O not true at all 2 O 3 O 4 O somewhat true 5 O 6 O 7 O very true	
n.	To learn about these types of Challenges	1 O not true at all 2 O 3 O 4 O somewhat true 5 O 6 O 7 O very true	
o.	To gain scientific recognition	1 O not true at all 2 O 3 O 4 O somewhat true 5 O 6 O 7 O very true	
p.	Because I had free time available	1 O not true at all 2 O 3 O 4 O somewhat true 5 O 6 O 7 O very true	
q.	Other		
13	**Would you have attempted to solve this Challenge if there was no financial reward offered?**	1 O Most definitely Not 2 O 3 O 4 O Maybe 5 O 6 O 7 O Most definitely Yes	
14	**Will you attempt to solve an InnoCentive Challenge in the future?**	1 O Most definitely Not 2 O 3 O 4 O Maybe 5 O 6 O 7 O Most definitely Yes	
15	**How satisfied were you with your experience with InnoCentive**	1 O Highly Satisfied 2 O 3 O Neither satisfied or dissatisfied 4 O 5 O 6 O 7 O Highly dissatisfied	
16	**Any thing else you may want to tell us about your experience with InnoCentive?**		

Your Background:

17	What is your Gender?	male: ○ female: ○
18	What is the highest academic qualification you have received?	----------select----------
19	What year did you receive your highest degree?	
20	What is the name of the institution where you got your highest academic qualification?	
21	What is the field in which you have received your highest qualification?	
22	What city were you living in at the time of your submission?	
23	What was your occupation at the time of the submission of the Challenge (including student or retired)?	
24	May Innocentive contact you via email if the MIT-CBS Research Team has any further questions on this topic?	Yes: ○ No: ○

Submit Survey

4 - Factor analysis to determine solver motivation clusters

Supplementary Table 2 - Motivations to Participate in Broadcast Search Problem Solving, Scores and Factor Loadings

Question: There are many reasons for participating in an InnoCentive Challenge. Tell us how true the following statements are for you. Please answer all items. I submitted a solution: (1-Not true at all, 4-Somewhat true, 7-Very true)

	Non-winning Solvers		**Winning Solvers**			**Factor 1 - Loadings**	**Factor 2 - Loadings**
	Mean	S.D	Mean	S.D	Significance	Intrinsic Motivation	Social and Work-Related Motivation
To learn about these types of Challenges	4.31	2.21	3.85	2.08	-	0.62	
Because I enjoy solving these types of Challenges	5.84	1.49	6.45	1.01	*	0.68	
For the intellectual challenge of solving this Challenge	5.09	1.99	6.03	1.62	**	0.71	
To enhance my skills	4.78	2.06	5.20	1.98	-	0.75	
To gain scientific recognition	3.41	2.28	3.21	2.12	-		0.5
Because someone suggested I participate in solving this Challenge	1.50	1.21	1.20	0.79	-		0.5
To enhance my career prospects	3.36	2.32	3.03	2.16	-		0.52
Because others I know have participated before	1.61	1.39	1.43	1.26	-		0.55
Because my boss asked me to work on it	1.13	0.60	1.05	0.22	-		0.57
To impress my colleagues	2.18	1.81	2.10	1.89	-		0.62
Because I had free time available	3.26	1.92	3.98	2.41	*		
Because I already knew how to get the solution	3.75	2.05	3.70	2.15	-		
Because InnoCentive told me about this Challenge	3.80	2.34	2.80	2.33	*		
Because my work/job at the time was not satisfying	2.19	1.79	2.13	1.96	-		
To try to beat other InnoCentive solvers	3.15	2.21	2.58	2.17	-		
To win the award money	5.44	1.83	5.73	1.38	-		
Eigenvalue						3.1	1.76
Percentage of variance explained (two factor solution)						0.69	0.31
Cronbach Alpha						0.78	0.71

Factor analysis. Varimax Rotation. Horst Correction. Loadings <= 0.4 not retained. Stata version 8

* p <0.05. ** p <0.01

5 – Table of Correlations for Regression Analyses

Supplementary Table 3 - Correlations Between Variables Predicting Problem Being Solved (N = 132 Problems)

	1	2	3	4	5	6	7	8
Problem Characteristics								
1 RTP Problem Type	1.00							
2 Award Value	0.69***	1.00						
3 Days Problem Open	0.38***	0.43***	1.00					
Seeker Firm Experience								
4 Previous problems posted by seeker firm	0.07	0.03	0.24**	1.00				
Solver community								
5 Solver base size	-0.16†	-0.18*	-0.18*	-0.18*	1.00			
6 Number of submissions	-0.31***	-0.2*	-0.12	-0.04	0.13	1.00		
Types of Solvers Attracted								
7 Distinct scientific interests attracted	-0.22**	-0.14	-0.16†	-0.12	0.57***	0.39***	1.00	
8 Generalist orientation of solvers	0.07	0.10	-0.25***	-0.15†	0.41***	-0.13	0.59***	1.00

† significant at 10%; * significant at 5%; ** significant at 1%; *** significant at 0.1%

Supplementary Table 4 - Correlations Between Variables Predicting Solver Being A Winner (N = 295 Respondents)

		1	2	3	4	5	6	7	8	9	10
Control Variables											
1	RTP Problem Type	1									
2	Time to develop solution	0.09	1								
Motivations											
3	Money	0	-0.06	1							
4	Extrinsic motivation	0.01	0.07	-0.11†	1						
5	Intrinsic motivation	-0.05	0.05	-0.21*	0.32***	1					
6	Beating other solvers	0	-0.07	-0.02	0.27***	0.26**	1				
7	Unsatisfactory job	0.07	0.01	-0.01	0.18***	-0.04	0.06	1			
8	Had free time	0.01	-0.13*	-0.08	0.05	0.08	0.03	0.25**	1		
Expertise											
9	Interest Count (at registration)	-0.03	-0.06	0.12**	-0.04	0.04	0.05	-0.02	0.03	1	
10	Problem mapping with field of expertise	-0.01	0.11†	-0.03	0.05	0.09	-0.01	0.11†	-0.01	0.13*	1

† significant at 10%; * significant at 5%; ** significant at 1%; *** significant at 0.1%

6 - Statistical models: Full versions of logit analyses

Supplementary Table 5 (Full version of Table 2 in original paper): Logit Regression on Problem Being Solved (N=132 Problems)

	Model 1	Model 2	Model 3	Model 4	Model 5
Control Variables					
RTP Solution Type	-0.025	-0.045	0.29	0.45	0.566
	(0.349)	(0.374)	(0.394)	(0.430)	(0.413)
Award Value	-0.46	-0.44	-0.452	-0.614	-0.418
	(0.430)	(0.407)	(0.433)	(0.439)	(0.449)
Days Problem Open	**-0.699***	**-0.941***	**-1.079****	**-1.313****	**-1.697****
	(0.341)	(0.396)	(0.398)	(0.427)	(0.536)
Seeker Firm Experience					
Previous problems posted by seeker firm		0.55	0.523	**0.604†**	**0.626†**
		(0.338)	(0.360)	(0.347)	(0.376)
Solver community					
Solver base size (Log)			-0.458	-1.329	**-1.897†**
			(1.106)	(1.099)	(1.134)
Number of submissions			**1.155****	**0.803***	0.049
			(0.399)	(0.363)	(0.333)
Types of Solvers Attracted					
Distinct scientific interests attracted				**0.993***	**2.305****
				(0.399)	(0.739)
Generalist orientation of solvers					**-1.638****
					(0.628)
Log Pseudolikelihood	-65.25	-63.62	-58.51	-54.51	-50.59
Wald's Chi Square	34.24***	32.34***	49.14***	46.89***	44.29***
Df	14	15	17	18	19
Pseudo R Square	0.22	0.24	0.29	0.34	0.39

Robust standard errors in parentheses. Includes controls for years and scientific disciplines of the problem (not shown for space reasons).

†significant at 10%; * significant at 5%; ** significant at 1%; *** significant at 0.1%

Supplementary Table 6 (Full version of Table 3 in original paper): Logit Analyses Predicting Which Solver Submits A Winning Solution (N=295 Respondents)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Control Variables						
RTP Solution Type	0.18	0.183	0.238	0.237	0.292	0.33
	(0.410)	(0.411)	(0.422)	(0.423)	(0.438)	(0.446)
Time to develop solution	**0.003***	**0.004****	**0.004****	**0.004****	**0.004***	**0.004***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Motivations						
Win award money		0.307	**0.386***	**0.426†**	**0.470***	**0.503***
		(0.187)	(0.185)	(0.241)	(0.220)	(0.214)
Career | Social motivations		-0.258	**-0.463***	**-0.464***	**-0.371†**	**-0.398†**
		(0.188)	(0.220)	(0.220)	(0.214)	(0.221)
Intrinsic motivations			**0.566****	**0.599****	**0.625****	**0.668****
			-0.182	-0.218	-0.203	-0.22
Beating other solvers					**-0.417†**	**-0.400†**
					(0.228)	(0.234)
Unsatisfactory job					-0.074	-0.126
					(0.264)	(0.265)
Had free time					**0.513***	**0.559***
					(0.237)	(0.234)
Expertise						
Generalist orientation of solver						**-0.315†**
						(0.172)
Problem distance from field of expertise						**0.398***
						(0.197)
Log Pseudolikelihood	-98.544	-96.706	-93.276	-93.251	-88.174	-85.627
Wald's Chi Square	6.36*	11.81*	22.40***	22.05***	26.59***	32.14***
Df	2	4	5	6	8	10
Pseudo R Square	0.0270	0.0451	0.0790	0.0793	0.1294	0.1545

Robust standard errors in parentheses

†significant at 10%; * significant at 5%; ** significant at 1%; *** significant at 0.1%

7 - **Tables and figures source of solution information used by winning solvers**

Supplementary Table 7A and 7B – Source of Solution Information Used by Winning Solvers in Their Submission
(N = 40)

Table 7A

Information from a solution previously developed by solver with:	Percent
No Modification	5.0
Minor Modification	17.5
Major Modification	32.5
Did Not Use Previously Developed Solution	45.0
Total	100.0

Table 7B

Information from a solution previously developed by someone else with:	Percent
No Modification	0.0
Minor Modification	12.5
Major Modification	47.5
Did Not Use Previously Developed Solution	40.0
Total	100.0

Supplementary Table 7C Source and amount of modification of prior solutions in creating present solution by Winning Solvers

	Solution Previously Developed by Solver				
Solution Information from Somewhere Else	No Modifications	Minor Modifications	Major Modifications	Did Not Use Previous	Total
Minor Modifications	0	10	0	2.5	13
Major Modifications	2.5	5	25	15	48
Did Not Use Previous	2.5	2.5	7.5	27.5	40
Total	5	17.5	32.5	45	100

Pearson chi2(6) = 21.534

Pr = 0.001
n = 40 solvers