HARVARD | BUSINESS | SCHOOL

"Fit": Field Experimental Evidence on Sorting, Incentives and Creative Worker Performance

Kevin J. Boudreau
Karim R. Lakhani

Working Paper

11-107

"FIT":
FIELD EXPERIMENTAL EVIDENCE ON SORTING, INCENTIVES AND CREATIVE WORKER PERFORMANCE

Kevin J. Boudreau, London Business School, kboudreau@london.edu

Karim R. Lakhani, Harvard Business School, k@hbs.edu

Abstract:

We present the results of a 10-day field experiment in which over 500 elite software developers prepared solutions to the same computational algorithmic problem. Participants were divided into two groups with identical skills distributions and exposed to the same competitive institutional setting. The "sorted" group was composed of individuals who preferred the competitive regime instead of a team-based outside option. The "unsorted" group had population-average preferences for working in the regime or the outside option. We find this sorting on this basis of institutional preferences doubled effort and the performance of solutions—controlling for skills, monetary incentives and institutional details.

Keywords: Sorting, Incentives, Institutions, Tournaments, Contests, Intrinsic Motivations, Creative Workers, Field Experiment

Acknowledgements: We are grateful to members of NASA's Space Life Sciences Directorate, including Jeff Davis, Elizabeth Richard, Jennifer Fogarty and Bara Reyna, for assistance in identifying an appropriate computational engineering challenge. The TopCoder team, including Jack Hughes, Rob Hughes, Mike Lydon, Ira Heffan, Jessie Ford and Lars Backstrom, provided invaluable assistance in carrying out all aspects of the experiment. This research particularly benefitted from thoughtful comments by Kenneth Arrow, Pierre Azoulay, Iain Cockburn, Peter Coles, Daniel Elfenbein, Silke Januszewski Forbes, Shane Greenstein, Nicola Lacetera, Mara Lederman, Joshua Lerner, Muriel Niederle, Gary Pisano, Al Roth, Sandra Slaughter, Scott Stern, Catherine Tucker, Eric von Hippel, Heidi Williams and D. J. Wu. Eric Lonstein provided outstanding project management and research assistance. All errors are our own. Kevin Boudreau would like to acknowledge the financial support of the LBS M-Lab. Karim Lakhani would like to acknowledge the generous support of the HBS Division of Research and Faculty Development. A Google Faculty Research Grant supported both authors.

1 Introduction

The standard economic approach to eliciting effort from workers is to offer pay-for-performance incentive schemes that reward workers for their observable outputs (Harris and Raviv 1978; Hölmstrom 1979). However, when it comes to creative workers engaged in novel problem solving and innovation-related tasks, high-powered incentives are often problematic. The encouragement of risky, uncertain search for solutions and repeated trial-and-error experimentation, essential for effectiveness in creative work, may be harmfully dampened if payments are directly contingent on every outcome (Hölmstrom 1989; Azoulay, Graff Zivin, and Manso 2009; Manso 2011). It is also often difficult to find sensible observables on which contracts can be written when the essential task to be accomplished is creative, non-routine and subtle in nature, with long gestational periods and uncertain outcomes (Hölmstrom 1989; Aghion and Tirole 1994). Further complicating the governance of creative problem solvers is the important role of intrinsic motivations in driving effort and innovative performance. High-powered, pay-for-performance incentives can "crowd-out" intrinsic motivations (Lepper and Greene 1978; Frey 1994; Kreps 1997; Benabou and Tirole 2003; Ariely, Gneezy, Lowenstein, and Mazar 2009), and creative workers might be particularly susceptible to these effects.

An added challenge of designing incentive schemes is that, beyond governing the conduct of workers, the particular institutional details of incentive schemes will attract different kinds of workers and generate sorting, thereby affecting the composition of the workforce (Salop and Salop 1976, Rosen 1986, Lazear 2000, Besley and Ghatak 2004, Dohmen and Falk 2011). Consistent with rational optimizing behavior, numerous empirical studies across both field and experimental settings find that higher-powered, contingent incentive schemes tend to attract higher-skilled workers (e.g.: Lazear 2000; Dohmen and Falk 2011). However, apart from sorting on skills, emerging evidence, particularly from creative settings, suggests that workers may simply sort on the basis of their intrinsic preferences for one regime or another—that is, their intrinsic "fit" with available institutions. For example, academic science attracts (and socializes) individuals who deeply value autonomy and connections to a broader community (Dasgupta and

David 1994; Aghion, Dewatripoint, and Stein 2008). Scientific workers transitioning from academia to the private sector are even willing to sacrifice monetary remuneration to do their work in institutions that assure these characteristics (Stern 2004). Studies in a range of other creative contexts like architecture (Brain 1991, Nasar 1999), arts (including classical music, drama, film, fashion, music, etc.) (Throsby 1994, Caves 2000, Frey 2000), law (Weisbrod 1983), the non-profit and public sectors (Besley and Ghatak 2004), and software development (Beecham, Baddoo, Hall, and Robinson 2008) also point to workers being greatly attuned to the institutional details of their work environment and the nature of the organizations for which they work.

In this article, we present the results of a field experiment that was designed to measure the performance implications of sorting on the basis of workers' intrinsic preferences for—or "fit" with—the institutional regime in which they work. We designed and executed a 10-day field experiment in which our subjects, more than 500 elite software developers from around the world, generated solutions to a complex algorithmic engineering problem (from NASA's Space Life Sciences Directorate) while working in a competitive contest regime on the TopCoder innovation contest platform. The TopCoder setting enabled us to obtain fine-grained and precise skill, effort and objective problem solving performance measures.

In keeping with the TopCoder context, participants competed in virtual on-line "rooms" of 20 direct competitors, with each room and each member of each room competing to solve the same problem, whose precise details were only revealed after the experiment and contest started. Each competitor could observe detailed profiles of other competitors, after being assigned to a room and the contest started. The top 5 competitors in each group were awarded prizes and remaining competitors in each room received zero cash awards. Following the event, an objective score of performance for each competitor was publicly posted.

We devised a sorting experimental approach in which we compared sets of participants who were exposed to the very same institutional regime, as above, and who possessed an identical distribution of raw problem-solving skills. However, one ("sorted") set of participants uniformly preferred to work in the competitive contest regime, rather than an outside option regime (that involved working in a team). The other, equally-skilled ("unsorted") group simply held the population-average preferences for working in the contest regime or the outside option. Further, to allow us to calibrate the size of sorting effects in relation to the application of extrinsic monetary incentives, roughly half of the rooms of 20 participants competed for zero cash prizes (allowing us to compare rooms with $1000 prizes versus $0 prizes).

The thrust of our analysis involves comparing the activity and effort exerted by the subjects in the sorted and unsorted groups and the resulting problem-solving performance. Our main finding is that sorting on the basis of intrinsic institutional preferences nearly doubled our measures of problem solving performance compared to the unsorted control group, holding constant raw skill, extrinsic incentives and features of the institutional context. The magnitude of the sorting effect was not only absolutely large, it was statistically indistinguishable in magnitude from the effect of competing for a $1,000 rather than $0 cash prize for each group of 20 competitors. We found no interactions between sorting on intrinsic incentives and cash incentives on absolute levels of performance and effort measures. The effect of sorting on institutional preference in this context (holding skill, monetary incentives and institutional details constant) can be attributed to how much workers chose to work. Sorted workers chose to work more hours on the problem, and the hours worked explained all statistically meaningful variation attributable to the sorting effect.

The application of formal cash incentives worked in a different way than the sorting effect. The absolute effect of cash incentives was significantly greater for higher-skilled participants than for lower-skilled participants, consistent with high-skilled workers having a greater chance of winning cash prizes. In addition, the cash incentive acted most acutely by increasing the fraction of subjects who worked more than the minimum amount of time, with a greater impact on the highest-skilled workers. By contrast, we found that the absolute effect of sorting on our measures of effort and performance were the same across the skills distribution. Sorting based on intrinsic institutional preferences acted both through increasing the proportion of workers who exerted more than the minimum level of effort, and by increasing the level of effort exerted by individuals who worked more than the minimum level. Thus, sorting based on institutional preferences for the regime appears to have created more general effects across different sorts of workers.

Therefore, as opposed to sorting on skill, which might be understood to primarily create compositional differences in the work force from one context to another, sorting on intrinsic institutional preferences led to (large) behavioral differences that had a real and meaningful impact on performance. We should have expected these behavioral differences to have been greater than those that might have been created by varying monetary incentives, given that the observed effects were the same as having a formal cash incentive *at all*, or not (rather than just incremental variation in the incentive). More broadly, the evidence affirms the important dual role of institutions, perhaps particularly for creative workers; the institutional arrangement served to both independently motivate and sort workers. Further, the sorting of workers, in effect, is not separable from the question of motivation. Thus, these results underline the crucial importance of the "fit" of workers. Here, we focused on fit in relation to the rules of the game *per se*. We might expect the social identity and other richer features of "living" organizations to further affect this relationship.

Methodologically, we implement a novel approach to a sorting experiment, a design that untangles the role of institutional fit, holding skills constant while varying cash incentives. We hence extend the sorting experiments approach pioneered in a series of lab experiments (e.g., Camerer and Lovallo 1999; Bohnet and Kübler 2005; Cadsby, Song, and Tappon 2007; Dohmen and Falk 2011; Lazear, Malmendier, and Weber 2011). We also bring this tradition to a field setting in which real workers exerted cognitive effort on a real, cognitively demanding problem that required a creative solution.

The remainder of our paper is organized as follows. The next section reviews the relevant literature on sorting and self-selection. Section 3 discusses in detail the sorting field experiment approach used in our study. Section 4 describes our data. Section 5 presents the main results and various robustness tests of our findings. We conclude in Section 6.

2 Related Literature on Sorting Effects and Worker Productivity

Prior work on institutions and incentive regimes has begun to tease out the role of self-selection and sorting of workers. Salop and Salop (1976) identified the importance of worker self-selection into incentive schemes that rewarded fast or slow turnover.

Jovanovic (1979) showed that worker turnover in the economy is driven by individuals trying to find a match and fit between their own productivity and the nature and type of work in a firm. Workers try to find and stay in jobs in which they are going to be relatively highly productive and self-select out of situations in which they have low productivity. Rosen (1986) developed a theory of equalizing difference by emphasizing that the different tastes and preferences of workers results in diversity of employment choices and wages. Hence, Weisbrod (1983) has argued that the large (up to 40%) wage differential between lawyers specializing in public interest litigation compared to other types of traditional law practice can be accounted for by individual taste for public service and notoriety, while controlling for differences in age, law school quality and academic performance. Besley and Ghatak (2004), in the context of mission-oriented organizations (non-profits and public administration), reasoned that matching between mission preferences of agents and organizations results in efficiency and economizing on the need for high-powered incentives.

Empirical work on the importance of institutional fit to worker performance has attempted to differentiate the "treatment" effect of inducing new behaviors of given workers by changing the rules of the game from the self-selection and sorting effect. The main message from a range of studies involving manual labor is that self-sorting into a variable-pay incentive scheme is driven by higher worker skill resulting in improved productivity. For example, Lazear (2000), using field data from a car windshield installation firm, showed that for manual labor changes in the incentive scheme from hourly wages to piece rate improved firm productivity by 44%. He found that half of the productivity improvement could be attributed to the new incentive system and the other half to changes in the workforce, whereby higher skilled workers were attracted to work at the firm and enjoy the benefits of piece rate. Bandiera, Guiso, Prat, and Sadun (2010), in the context of management, find that family-owned firms as compared to widely-owned firms offer contracts that are less sensitive to performance which in turn attract less talented managers who work less, earn less and are generally dissatisfied. Further natural and field experiments in the context of tree planting and garment workers have provided additional support for this skill-based sorting effect (Shearer 2004; Franceschelli, Galiani and Gulmez 2010; Shi 2010).

Closely related to our work is Dohmen and Falk's (2011) study showing sorting effects in a laboratory setting in which subjects multiplied single and double-digit numbers. The authors found that more highly skilled workers prefer contingent-incentive schemes like tournaments or piece-rates, which also drove higher levels of output and effort, to fixed payments. In a similar vein, laboratory experiments by Cadsby, Song and Tappon (2007) and Ericksson, Teyssier and Villeval (2009) demonstrate that higher productivity workers prefer to work under contingent-payment schemes, and that this results in improved performance outcomes.

Laboratory experiments in behavioral economics that have shown individuals to have differential tastes for institutional regimes further suggest the significance of sorting and self-selection for a variety of related outcomes. The salience and impact of sorting have been studied in prisoner's dilemma games (Bohnet and Kübler 2005), bargaining games (Oberholzer-Gee and Eichenberger 2008; Lazear, Malmendier, and Weber 2011), the gift-exchange game (Eriksson and Villeval 2008), assessment of overconfidence (Larkin and Leider 2010) and market-entry games (Camerer and Lovallo 1999).

3 Design of the Sorting Experiment

In the remainder of the paper, we present the design and results of a field experiment for estimating how workers' preferences for working within given institutional contexts influences problem-solving effort and outcomes. The essential idea is to compare a "sorted" group, the members of which uniformly prefer to work within a given regime, to an "unsorted" group that simply reflects the population average distribution of preferences. Thus, rather than attempt to expose identical groups to different treatments, the usual experimental approach, this sorting experiment does just the opposite: it exposes groups that systematically differ in a particular way (while being held identical in terms of skills distribution) to identical treatments. The main activity consisted of sorted and unsorted participants competing to solve an algorithmic computational-engineering problem over the course of ten days.

3.1 Field Setting

The inherent objective in a sorting experiment such as ours is to demonstrate the importance of different types of participants on outcomes, in particular, the interaction between institutional preferences and institutional environment. It follows that the types of participants in question and the institutional context have some empirical relevance. For this reason, we pursued a field rather than laboratory setting. At the same time, the estimation of sorting effects here places especially high requirements for observing relevant microeconomic variables within a controlled environment.

We conducted the experiment on TopCoder.com, an on-line platform on which elite programmers from around the world who sign up as members compete against each other in a regular stream of contests that involve solving software development and computational-algorithmic problems for a variety of firms. Winners receive cash prizes, typically on the order of several hundred dollars.[1] TopCoder insists on maintaining high fidelity records on all contests and participants. Thus, when members compete directly against one another, TopCoder selects winners through an objective, computationally-based scoring criterion with no performance ambiguity and all results of the contests are publicly displayed. Furthermore, after each contest, each participant is given a precise and public ranking and skill rating for that particular problem type. The TopCoder rating is based on the long-established "Elo" system used to evaluate, rate and rank chess grandmasters (Mass and Wagenmakers 2005) and other competitive contexts like the US College Bowl systems, National Scrabble Association members and the European Go Federation. The Elo rating creates a relative performance metric based on the performance of all other participants and an individual's current and past performance. Thus, at any given time, participants have a clear idea of their ranking and rating within the entire population of TopCoder participants (See Appendix 1 for a view of the public profile and ratings of a competitor). TopCoder adds further differentiation to the rating system by color-coding rating ranges to allow for easier identification and sense of achievement for participants (red being the highest rated band). Interviews with TopCoder executives and competitors indicate that the TopCoder skills rating is often used as a credible signal in personnel hiring decisions by information technology (IT) intensive firms. Organizations like Google, Facebook and the US National Security Agency often encourage job applicants to obtain a TopCoder rating in order to be

[1] See Boudreau, Lacetera and Lakhani (2011) for an extensive description of the context.

considered seriously for an open position.

In the experimental set-up, we sought to follow the routine characteristics of the usual TopCoder contest as much as possible. Just as TopCoder often does, we divided participants into 20-person groups that would compete directly with one another to solve a real problem in virtual competition "rooms." Participants use the TopCoder "arena" interface to program their solutions and observe the competitive field in their rooms. Information on direct competitors is updated in a side window of the interface; the problem-solving screen is in the center. (See Appendix 2 for screen shots.) The side window lists the 19 other competitors' unique TopCoder "handles" (pseudonyms) and numerical skill ratings. The side window also displays the best scores for submitted solutions to that point. Clicking on any name reveals a complete history of the participant on the platform (as shown in Appendix 1).

Over the course of contest, individual competitors could submit as many times as they liked. Each solution submitted to the system was near-instantaneously subjected to a barrage of automated tests to register a score. Therefore, submissions provided a means of receiving feedback on interim solutions. Final scores of each participant were based on the highest score attained by the individual, almost always for the last submission by the individual. Cash prizes were awarded on the basis of the rank order of scores attained in the room. First place in a group of direct competitors received $500, second place $200, third place $125, fourth place $100 and fifth place $75. Thus, five of twenty competitors in each room received prizes.

The TopCoder regime therefore represents an institution with a distinctly competitive character in which individuals compete autonomously, have their performance and skills objectively measured and shared publicly and are rewarded based explicitly on performance and subsequent ranking. Both the intensely competitive and autonomous characteristics of this context are salient to findings regarding the software developer labor market. Decades of descriptive and survey-based research has consistently reported considerable heterogeneity in psychological and behavioral orientations of software workers (Beecham et al. 2008), a large subset of whom prefer to be autonomous "loners" (Schneiderman 1980) who crave individual rewards and recognition (Couger and Zawacki 1980).

3.2 The Problem to be Solved by All Participants

The problem to be solved by each participant was to optimize the contents of the “Space Flight Medical Kit” for NASA’s Integrated Medical Model. This is a computational-engineering problem that involves developing a robust algorithm for determining what components (consumable (e.g., medicines) and non-consumable (e.g., heart defibrillator) resources) to include in the space medical kit included in each of NASA's space missions. There exist algorithms that have been developed by NASA staff; NASA's goal in participating in this experiment was to increase the sophistication and effectiveness over a wider range of applications (including missions to the International Space Station) in which mission length would increase greatly. (The winning algorithm from this experiment is now in use for all NASA missions.) The solution had to take into account that mass and volume are restricted in space flight, and that the kit’s resources needed to be sufficient to accommodate both expected and unexpected medical contingencies encountered while in space lest the mission have to be aborted and an afflicted astronaut returned to earth.[2] The contents of the kit also had to be attuned to the characteristics of the space flight and crew and nature of the mission. The challenge was thus to develop an algorithm that addressed mission characteristics while trading off mass and volume against sufficient resources to minimize the likelihood of medical evacuation. (See Appendix 3 for a full problem statement and the scoring approach.)

NASA also worked with TopCoder to develop a precise scoring function that would provide an objective performance metric for the code submissions from our subjects. The automatic scoring was based on an already established simulated set of 200,000 mission scenarios involving various medical contingencies that may occur during space flight. Each code submission in the experiment was subjected to a random set of 10,000 scenarios on the basis of which the performance of the algorithm and score were determined. The problem, being relatively focused, was to be solved as an integral project capable of being divided into a set of subroutines and call programs. The solution to this problem is not a matter of “software development” as might be casually thought of, but rather a non-trivial sort of algorithmic problem that participants in TopCoder

[2] The health and safety of NASA Astronauts is covered under the general safety regulations of the Occupational Health and Safety Administration.

tournaments frequently encounter.

3.3 Eliciting Preferences, Sorting and Matching Procedures

The central point of the design of this sorting experiment is to compare a "sorted" group of participants (in which participants uniformly have a preference to work within the competitive TopCoder regime) to an "unsorted" group (in which participants possess the population-average distribution of preferences). A key challenge here, however, is that individuals' institutional preferences may be correlated with raw problem-solving skills. Indeed, past studies have found evidence that higher-skilled workers tend to have a greater likelihood of preferring competitive environments and high-powered incentive schemes (e.g., Lazear 2000; Dohmen and Falk 2011). In our analysis, however, we are interested in how individuals' preferences per se influence outcomes, not how preferences might be correlated with skill levels. One way to account for skills when drawing comparisons between sorted and unsorted groups is simply to exploit TopCoder's skill rating measures, applying these measures as controls when making econometric comparisons. Our experimental design is intended to further deal with observable as well as possibly unobservable characteristics by means of an assignment procedure that involves a combination of matching and randomization. Figure 1 summarizes the assignment procedure.

<Insert Figure 1 Illustration of the Assignment Procedure>

As a first step towards assuring that sorted and unsorted groups will have identical skills distributions, we rank order all participants according to their TopCoder skill rating. From this rank-ordered list, we create, from top to bottom, successive "ordered pairs," or sets of two consecutive participants in terms of skill level. We then split the overall population into two equally sized groups of participants with identical skills distributions by randomly assigning members of each ordered pair to one group or the other (i.e., group "A" and group "B" in Figure 1). To construct the sorted group, we secretly asked members of one of these groups about their preferences for working in the TopCoder regime. We followed past experimental work involving sorting (Dohmen and Falk 2011; Ericksson, Teyssier and Villeval 2009; Lazear, Malmendier, and Weber 2011**)** by presenting alternative choices and asking half of our participants to choose. However, we diverged from past work by attempting to elicit our subjects' preferences without implying that a statement of preference would necessarily lead to an assignment of their choice. This was accomplished by asking participants to state their preference for a regime on a likert scale under three different hypothetical scenarios. The ordering of the likert scale choices was randomly reversed to prevent any sort of order preference or recency bias. (See Appendix 4 for the instrument used.) Our aim was to minimize any altered behavior that might result from the solicitation of preferences. (See Section 5.2 on Hawthorne effects for further discussion of this point.)

To elicit participant preferences for the TopCoder competitive regime, we presented members of group A with an alternative concept of competing on a "team" as the outside option instead of working autonomously.[3] In the team option, rather than compete among 20 individuals, participants would cooperate with four other individuals on a team competing against four other teams. Total cash prizes and expected payoffs would remain the same as in the usual competitive regime, but would be divided among team members. Table 1 contrasts the TopCoder competitive regime and outside option of working on a team. To construct an unsorted group with the same skills distribution as the sorted group, but with institutional preferences that reflect the population average distribution of preferences, we simply assigned individuals in group B the same institutional regime as the one preferred by their matched alter (in the ordered pair) in Group A. (Group A participants who preferred the outside option and their matched ordered pair alters in Group B therefore drop out of the sample and analysis.) Figure 2 shows that our assignment procedure achieved a near identical skills distribution between the sorted and unsorted group.

<Insert Figure 2 Kernel Density Skills Distribution for Sorted and Unsorted Groups>

After constructing these larger pools of sorted and unsorted participants from

[3] The option of working on a team or autonomously is consistent with the broad organization of software development tasks in the economy (see, for example, Mowery 1996; Cusumano 2004). However, it should be noted that TopCoder does not offer the team option in its competitions.

groups A and B, the sorted groups were randomly assigned to virtual "rooms" of 20 participants who would be direct competitors. Among these rooms of 20 sorted competitors it was randomly determined which would compete for $1,000 (rather than $0) in cash prizes. To construct rooms of unsorted competitors to which these would be compared, the ordered pairs of these sorted participants were then assigned to "mirror" rooms to enable us to examine them under conditions of identical prizes and identical distribution of skills of competitors.

4 Data Set

Following the assignment procedure described in Section 3, the sample includes 516 observations (individual participants). Of the original 1,040 individuals who participated in the overall event, half (520) were asked their preferences for the TopCoder competitive regime versus the cooperative outside option. Of these, 264 (50.8%) stated that they preferred the competitive TopCoder regime over the cooperative outside option. These 264 participants were randomly assigned to fill up 13 virtual "rooms" (independent groups) of 20 individuals. Of the 13 rooms, 6 competed for a cash prize of $1,000 and the remainder did not. The ordered pairs of these assignees (who were not asked their preferences for the different regimes) were assigned to 13 rooms that mirrored the first 13, again with 6 rooms that competed for a cash prize. The number of observations (i.e., individuals), 516, is not a perfect multiple of 20 (participants per room), as we dropped observations for which the algorithm skill rating was not available.[4]

As anticipated in Section 3, the fraction preferring the competitive TopCoder regime among the 520 individuals who were asked their preferences was positively correlated with skill level. Figure 3 presents a flexible non-parametric regression to illustrate the proportion that preferred the competitive regime at different levels of TopCoder's skill rating.

<Insert Figure 3 Proportion Preferring Competitive TopCoder Regime Over the Outside Option (Cooperative Regime), by Skill Level>

[4] The equal treatment of individuals without algorithm ratings in the experiment was a requirement set forth by TopCoder.

With regard to our research objective of measuring the effects of sorting, it should also be noted that the sample is itself drawn from the pool of TopCoder members. Therefore, the subsequent analysis of sorted and unsorted groups should be interpreted as somehow analogous to "treating on the treated." Thus, we might speculate that any sorting effects we observe here could be small in relation to differences among more diverse groups. Our main dependent variable relates to problem-solving performance. A measure of the quality of each algorithm/solution was calculated with an automated test suite that assessed the performance of the submitted algorithm against a barrage of tests and contingencies, as described in Section 3. The final score assigned to an individual competitor (*ProblemSolvingScore*) was the best for all submissions by a given participant, typically the final submission. Overall, 38% of the sample participants (195) made submissions. Non-submissions received zero points. This led to a bimodal distribution in the sense that this 38% was relatively uniformly distributed up to a maximum score of 8,957; another 62% of observations spiked at a score of zero. This sort of bimodality is also reflected in measures of effort and activity, as described below.

Apart from problem-solving performance, we collected measures of the effort and actions of participants. The measure *NumSubmissions* is an observational measure related to level of activity. It provides a count of the total number of submissions made by a participant over the course of the 10-day experiment. Submitting code in this fashion was virtually costless and resulted in near instantaneous feedback. This is a direct indication of the intensity of development effort, all else being equal, as code submission reflected code testing and evaluation. We also collected a more directly interpretable measure of effort: the number of hours worked over the course of the ten days by each participant. The variable *HoursWorked* was a self-reported estimate of the precise number of hours worked over the course of the ten days. This was collected by means of a mandatory survey that was completed electronically immediately following the experiment. The survey was mandatory in the sense that it needed to be completed prior to learning final results, rankings and winners. Further, receipt of a commemorative t-shirt (including the individual's name on the roster of participants) was conditional on having completed the survey. Where we did not immediately receive a response, we followed up with personalized emails and phone calls to get near complete coverage. The data on hours worked suggest that there is a broad distinction between those who devoted just less than one hour to this exercise and a continuum of hours worked, if greater than this amount. As an indication of the close relationship between the observational code submission measure and survey-based measure, the proportions of observations with non-zero levels are almost identical, 38% versus 39%.

Observations are also coded in terms of whether they correspond to the sorted group with an indicator variable, *SortedonPreference*, and a $1,000 cash prize (rather than no cash prize whatsoever), *CashPrize*. Our measure of raw problem-solving ability, *SkillRating*, for each pariticpant, is based on TopCoder's rating system. We use, specifically, the rating calculated, just prior to the experiment, for what TopCoder terms "Algorithm" matches, software solutions to abstract and challenging problems akin to the problem in the experiment. Tables 2 and 3 provide variable definitions and summary statistics.

<Insert Table 2 Variable Definitions>

<Insert Table 3 Summary Statistics>

5 Results

5.1 Comparison of Simple Means

Given the design of the experiment, a comparison of mean outcomes should, in principle, provide meaningful comparisons. Therefore, we begin by simply comparing *ProblemSolvingScore* across different groups. The mean *ProblemSolvingScore* attained across participants during the 10-day experiment was a score of 1,736, with considerable variation (standard deviation = 2,802). The most important result of this article can be noted by comparing the mean scores of the sorted and unsorted groups: the average problem-solving performance of sorted groups is almost twice as high (an increase of 83%) as the unsorted groups with equal skills, an average score of 2,244 versus 1,228.

Table 4 provides further details by breaking-down outcomes by both sorted and unsorted groups and those that competed for $1,000 prizes or none.

Several additional patterns are immediately apparent. First, the large effect of sorting on institutional preferences exists both with and without the cash prize (1,682 – 758 = 924 point difference without the cash prize; 2,976 – 2,070 = 906 point difference with the cash prize). Similar sorting effects can be seen in the case of activity and effort measures. Whether with or without cash prizes, participants in the group that was sorted on institutional preferences made 1.8 (i.e., 2.58 – .78 or 5.38 – 3.55) more submissions, on average. The cases of cash prize and no cash prize was slightly more substantively (although not statistically) different in the case of the number of hours worked: in the case of no cash prize, sorted participants worked 6.7 more hours (i.e., 10.16 - 3.48), on average; in the case of cash prizes, sorted participants worked 10.7 hours more (i.e., 21.42 - 10.70), on average.

<Insert Table 4 Comparison of Mean Outcomes, Stratified by Treatment>

5.2 Regression Analysis, Robustness and Interpretation

Although the earlier comparisons' means should provide meaningful results, analyzing the data within a regression framework enables us to more explicitly assess key assumptions of the design and more deeply interpret patterns in the data. Baseline OLS regression results, with robust standard errors, are reported in Table 5.

If the assignment procedure was effective and left no systematic differences across treatments, the estimates should be unchanged when we include skill controls. (The specifications here are also reviewed, as they provide a basis for later regression models.) For ease of comparison, model (5-1) begins by reporting the two-way correlation of *ProblemSolvingScore* on *SortedonPreference*. This effectively recasts the earlier descriptive statistics in a regression framework; the coefficient on *SortedonPreference*, 1,016, is simply the difference between mean performance in the sorted and unsorted groups (i.e., the difference between 2,244 and 1,228, as above). Model (5-2) re-estimates the coefficient on *SortedonPreference* with *SkillRating*, now included as an explicit control variable. The estimated coefficient is virtually unchanged. (The estimated constant term dramatically changes and becomes statistically indistinguishable from zero, given the importance of *SkillRating* in explaining performance outcomes.)

To account for possible non-linearities in the relationship between skill and performance, we replace the linear control for skills with a series of dummies for different bands of skill levels. These correspond to the different bands of skill levels (i.e., *SkillRating* in the following bands: <900, 900-1,200, 1,201-1,500, 1,501-2,200, >2,200) TopCoder uses to distinguish different classes of competitors. As reported in model (5-3), the coefficient on *SortedonPreference* is again virtually unchanged. Model (5-4) provides the most stringent skill control by re-estimating the sorting effect by directly comparing the differences between the ordered pairs. (Recall that these pairs were based on matching individuals with effectively identical skills ratings and then randomly assigning one to the sorted group and the other to the unsorted group.) This estimate of the effect of sorting based on "ordered pairs differences" is almost identical to the earlier estimates, again estimating a roughly 1,000-point average effect of sorting.

<Table 5 Baseline OLS Regression Results>

Given the assignment procedure, the assignment to rooms with cash prizes should also be uncorrelated with skills or sorting. Indeed, including *CashPrize*, as in model (5-5), again leaves the coefficient on *SortedonPreference* statistically unchanged. (Note that the coefficient on *CashPrize* cannot be estimated with ordered pair differences, given that ordered pairs were subjected to the same cash prize treatment. Therefore, the earlier-described dummies for different ranges of skill ratings were used instead.) At least as important, the inclusion of *CashPrize* provides another tangible indication of the relative importance of the sorting effect, this time in relation to the presence or absence of a formal high-powered incentive. Although the point estimate of the coefficient on *CashPrize* (1324) is larger than that of *SortedonPreference* (1010), the difference is not statistically significant.

The comparison of simple mean outcomes across the sorted and unsorted groups in Section 5.1 suggests a close link between levels of problem-solving activity

(*NumSubmissions*, *HoursWorked*) and performance (*ProblemSolvingScore*). To more explicitly reveal this link, we regress the two measures of activity on *SortedonPreference*. We again exploit the difference between matched pairs, this time using a fixed effect for each matched pair in a robust count (Poisson) framework, as both measures of effort and activity are non-negative integers.[5] Results are presented in Table 6. We begin with *NumSubmissions*, the observational measure of activity, and report results in model (6-1). The coefficient on *SortedonPreference* is estimated to be .65, implying an incidence rate ratio of 1.91. Model (6-2), an analogous model with our self-reported measure of effort and activity, *HoursWorked*, estimates a coefficient of .84. This implies an incidence rate ratio of 2.31. Clearly, *NumSubmissions* has a natural advantage as a measure of effort and activity, given that it is an observational measure rather than self-reported. However, model (6-2) and the *HoursWorked* measure produced a better-fitting model (log-likelihood of -2,298 versus -707) and more directly interpretable result. Further, it is possible that *NumSubmisssions* captures not only effort and activity, but also what might be called "style" of problem-solving (i.e., a heavy testing and iterative, versus more contemplative and deliberate, method). Therefore, *HoursWorked* is taken to be the preferred measure.

The large effect of sorting on the basis of institutional preference on effort and activity is perhaps analogous to large sorting effects found earlier on problem-solving performance. To further investigate the extent to which this boost in effort and activity can account for the boost in problem-solving performance, we again regress *ProblemSolvingScore* on *SortedonPreference*, this time controlling for level of effort and activity. (We also control for raw problem-solving skill, again using the most stringent approach of estimating the effects on the basis of differences across matched pairs.) If effort and activity account for the performance boost, we should see the coefficient on the sorting variable to drop to zero when we control for effort and activity. To roughly control for effort and activity, we simply include our preferred measure *HoursWorked*, along with a quadratic transform of this variable to allow for possible concavity or convexity. As reported in model (6-3), *ProblemSolvingScore* increases with *HoursWorked* in an increasing and concave way, consistent with diminishing marginal

[5] Linear regressions produce similar results.

returns to effort. Crucially, when effort levels are controlled, even in this simple way, the coefficient on *SortedonPreference* becomes statistically indistinguishable from zero. Therefore, evidence points to the sorting effect on problem-solving performance being mostly attributable to a boost in effort and activity.

<Insert Table 6 Results of Effort and Activity Regressions>

In Table 7, we report results in which we attempt to further interpret the precise nature of the sorting effect being measured. The experiment was designed with the intent of measuring the effect of sorting of individuals on the basis of expressed preferences for types of institutions, the "rules of the game" *per se*. That is, individuals did not express preferences on the basis of *who* would be working within these regimes, as they did not know who would be assigned to their groups. Nonetheless, it is still possible that, once assigned to a room, the behavior of others in the room could have affected the actions, incentives and activities of a participant (Bandeira, Barankay, and Rasul 2005). For example, an especially active or challenging competitor might either stimulate or diminish the performance and activity of competitors in the same room. If so, then the earlier-measured sorting effects would not simply reflect a direct relationship between individual workers and the rules of the game under which they function, but also this social interaction. Therefore, we re-estimated the sorting effects, this time controlling for the performance of the other 19 participants in the same room. We begin with a restatement of the results of model (5-3) for ease of comparison. Model (7-1) then adds a control for the average performance achieved, *ProblemSolvingScore*, by other participants in the same room. The model controls for the series of dummies for different skill bands. (Estimating on differences across matched pairs would eliminate most variation, as matched pairs were assigned to "mirror rooms" of pairs.) Adding this variable to reflect possible social interactions reveals *nothing*: the coefficient on this average is statistically zero and the coefficient on *SortedonPreference* is virtually unchanged. Model (7-2), which adds a range of measures of the distribution of the performance of peers in the same room (variance, skew, maximum) also finds no change. We also ran analogous regressions, but with our measure of activity and effort, *HoursWorked*. Again, we found no change in results and no evidence of any sort of social interaction. We therefore conclude the estimated sorting effect does not include social interactions.

<Insert Table 7 Results of Tests for Social Interactions>

Finally, in interpreting the estimated sorting effects, it is important to assess whether the large effects measured here might have resulted from simply being asked their preferences rather than necessarily subsequently assigning individuals according to their preferences. This would represent a Hawthorne effect of sorts. For example, if individuals believed that being asked their preferences for one regime or the other was tantamount to being given the ability to choose their assignment, they might have then had a sense of, say, accountability or commitment to the choice (cf., Dal Bo, Foster, and Putterman 2010). Our most important approach to mitigating this possibility was to design the process for eliciting preferences to avoid any direct implication that preferences would translate into assignments (Appendix 3). Individuals who were asked their preferences might also have had a heightened sense of being observed within an experiment. We attempted to diminish this effect by embedding the experiment within a "usual" TopCoder event, albeit one that assumed an especially high profile as a usual event (i.e., involving NASA, a large prize purse, ample publicity, etc.).

To explicitly estimate the magnitude of any Hawthorne effects, we attempted to compare outcomes of participants with similar preferences who were assigned to the sorted and unsorted treatments. This was possible in a subset of out-of-sample data in which those who described themselves as "indifferent" were uniformly assigned to the cooperative team outside option. We found no statistical difference between either the *ProblemSolvingScore* or *HoursWorked* of indifferent participants in the sorted group and their ordered pairs who were in the unsorted group.

5.3 Synthesizing an Alternative Control Group with Propensity Scores

The estimated magnitude of the sorting effect will depend on the control group to which the sorted group is compared. The earlier regressions compared the sorted group (in which 100% of participants preferred the regime) to an unsorted group in which the preferences should be the same as the population distribution of preferences. Comparing a sorted group to this population average distribution is, of course, a natural and meaningful comparison to make. However, it is also true that high-skill participants are more likely to prefer the competitive regime (Figure 3). Therefore, our earlier estimates can be understood to underweight the effect of sorting among high-skill participants given simply that the unsorted control group is more similar to sorted group among high-skill competitors. In this section, we generate an alternative "skills-neutral" estimate by synthesizing an alternative control group in which the propensity to compete is fixed to the population average across all skill levels.

As a first step, we build a model of individuals' likelihood or propensity to prefer TopCoder's competitive regime over the outside option using data from the half of the original 1,040 participants who were originally asked their preferences, prior to making assignments (cf., Figure 1). In a Logit model, we regress an indicator for a preference for the competitive TopCoder regime on a variety of demographic variables collected by TopCoder for all members (when they signed up for the platform). Although raw problem-solving skill level is clearly an important predictor of institutional preference, it explains only a minority of variation; the intent here is to explain additional variation above and beyond this. (We will later reweight the control group according to these predicted institutional preferences.)

We present here a series of estimates that progressively add the explanatory variables. Results are presented in Table 8. The advantage of showing results with variables progressively added is to illustrate the stability of the model, despite widely varying specifications. Although we do not need to interpret coefficients, only the "fit" of a model, the stability of the model lends support to the notion that the model is somehow meaningful. We begin by simply regressing the preference for the TopCoder regime on skills, as in model (8-1). To allow for non-linearities, we specify *SkillRating* as the earlier-described series of dummies corresponding to the distinct ranges used by TopCoder.[6] Subsequent regressions add responses to a questionnaire TopCoder administers when new members sign up for its platform. Model (8-2) introduces indicator variables that correspond to self-reported reasons members initially joined the platform.

[6] Linear or quadratic terms of *SkillRating*, if added to this model, are insignificant.

Not surprisingly, those motivated by competition ("technology competition") reported systematically higher preference for the competitive TopCoder regime than for the cooperative outside option. Model (8-3), which introduces indicator variables for different age ranges, finds that older participants tend to prefer the competitive regime. Model (8-4), which introduces an indicator that distinguishes professionals from students, finds no statistically significant effect (although it remains consistently positive when including or excluding other variables). Model (8-5) introduces a series of dummies that capture participants' countries of origin. Even in this quite radical re-specification of the model, in which dummies for the 79 countries "soak up" much of the variation, the remaining model coefficients do not radically change, thereby affirming the robustness of this probability or propensity model. In the analysis to follow, we assayed models (8-3), (8-4) and (8-5) as propensity models. We report results using model (8-4), given that it includes a large number of explanatory variables while remaining transparent in terms of the nature of the relationships that are exploited.

Reweighting to Establish a Constant Average Propensity Across all Skill Levels

Model (8-4) is then used to estimate the unobserved propensities of those in the unsorted group, who were not asked their preferences. We do so by substituting these participants' own demographic data into model (8-4). We then reweight the data to shift the mean propensity to competition to be equal to the aggregate population-average across all skill levels.

<Insert Table 8 Logit Model Results of Probability / Propensity to Prefer Competitive Regime>

To reweight the control group in a way that adjusts propensity to compete while holding the skills distribution constant, we first divide the observations of the unsorted control group into ascending *SkillRating* blocks, following TopCoder's established color-coding of skill levels (i.e., <900, 900-1,200, 1,201-1,500, 1,501-2,200, >2,200). We then adjust the within-block relative weights of observations in order that the total within each block fixes the weighted average of propensities to the population average (i.e., the overall likelihood of preferring the competitive regime, among those who were asked, is

50.7%). To emphasize, the re-weighting occurs within blocks with the total weight of each block kept constant.

To describe this process, let each observation of estimated propensity to compete, *P*, be indexed by *j*. Within each skills block, indexed by *i*, there are N_i. We re-weight observations within each block in linear proportion to the propensity level, i.e., $1+ \omega_i \cdot \mathrm{P}_{ij}$. Therefore, the entire weighting scheme reduces to estimating the *ω* parameter for each skills block:

$$\omega_i = \frac{\overline{P} \cdot N_i - \sum_{j=1}^{N_k} P_{ij}}{\sum_{j=1}^{N_k} P_{ij}^2 - \overline{P} \cdot \sum_{j=1}^{N_k} P_{ij}}$$

The overall weight of each re-weighted block within the control group is also kept equal to its original overall weight so as to leave the skills distribution unchanged.[7]

We proceed to re-estimate effects with this alternative (re-weighted) unsorted control group. To most explicitly reveal the effect of re-weighting, Figure 4 plots results of a flexible non-parametric regression of *ProblemSolvingScore* on *SkillRating* for both the sorted and unsorted groups. For the unsorted group, we plot the relationship for both the un-weighted and re-weighted unsorted control group. As should be expected, the simple fact that the unsorted group is more similar to the sorted group at high skill levels suggests that there is a seeming negligible performance difference between the sorted and unsorted group among high-skilled competitors. However, the re-weighted control group shows an roughly equal difference between sorted and unsorted groups across all skill levels.

<Insert Figure 4 Non-Parametric Regression of Problem-Solving Performance, Stratified by Treatment>

[7] To assure that the within-block re-weighting did not systematically bias the skills distribution (by, say, systematically weighting observations on one "side" of each block), we confirmed that the un-weighted and re-weighted unsorted control group possessed statistically identical estimated means, variance and skew of skills. We also explicitly plotted the skills distribution (i.e., kernel density) of un-weighted and re-weighted data and found them to be almost identical and have no indication of any such systematic distortion. Of note, other approaches to re-weighting that would also hold skills constant while fixing the probability of preferring the competitive regime were possible; however, this approach allows us to simultaneously re-weight the skills distribution in a later analysis.

<Insert Table 9 Regression Results with Synthesized Control Group>

We then summarize and further explore these effects within a regression framework. Table 9 begins by re-stating the results of the un-weighted model (5-5). This model not only estimates the sorting effect, but also includes *CashPrize* and therefore allows us to compare this effect. Further, this model controls for skills with the series of dummy variables corresponding to TopCoder skills color bands. This is appropriate here, as it is no longer appropriate to estimate effects on the basis of differences across matched pairs after we re-weight the unsorted control group in subsequent steps. Model (9-1) re-estimates this model with the synthesized control group, which holds propensity to compete even across skill levels. The newly estimated coefficient on *SortedonPreference* is statistically unchanged, but increases substantially by 13% (from 1,010 to 1,140). The estimated response to the cash prize changes far less, leading the re-weighted estimates to be substantially closer. Given our interest in the overall distribution of effects, model (9-2) interacts the main explanatory variables with skill levels.[8] The model confirms earlier estimates of the magnitude of the sorting effect and no interaction effect (at least once the unsorted control group is re-weighted, as can be seen in Figure 4). However, it appears that the cash prize incentive operates quite differently, with higher-skilled participants responding far more than lower-skilled participants to the cash prize. This is consistent with higher-skilled individuals simply having a greater expectation of winning. The direct, un-interacted effect of the cash prize in this model is estimated to be statistically indistinguishable from zero.

5.4 Analysis of the Distribution (Bimodality) of Outcomes

To this point in the analysis, we have focused on estimating average effects of sorting on the basis of workers' institutional preferences (holding other factors constant). However, the earlier description of data (cf., Section 5.1) highlighted that outcomes were bimodal: a fraction of participants worked no more than a minimum amount of time (i.e., *HoursWorked* < 1 hour) and, consequently, received a zero score. Other participants

[8] There is no significant effect in the remaining possible interaction, that between the cash prize and sorting, in any of the remaining analyses (not reported).

worked more than this minimal amount and achieved a relatively smooth distribution of performance outcomes. To better understand and describe this bimodality, the analysis here decomposes the effect of sorting on the decision to exert more than the minimum level of effort[9] from the effect on performance, conditional on having chosen to exert more than the minimum level of effort.

The Decision to Exert More than the Minimum Effort

Figure 5 begins by examining the decision to exert more than the minimum level of effort (*HoursWorked* > 1). Relationships are plotted separately for the sorted and unsorted groups. Again, we present the un-weighted and re-weighted unsorted control group. The flexible, non-parametric regression lines in this figure essentially trace the fraction of participants that chooses to exert more than one hour of work. The patterns would appear to largely mirror the earlier observed patterns related to problem-solving performance (i.e., Figure 4), with systematic differences between the sorted and unsorted groups. The differences in overall performance across treatments would appear to at least, then, largely be due to the fraction of individuals that simply chooses to exert some level of effort above the minimum.

Summarizing these differences in a regression framework enables us to essentially understand the overall (weighted) average effect, while comparing sorting and incentive effects. Linear probability models of choosing to exert effort are reported in the first columns of Table 10. The unsorted control group in these regressions continues to be re-weighted, as before. Model (10-1) regresses an indicator for exerting more than one hour of effort on *SortedonPreference* and *CashPrize*, while controlling for the series of dummies for skill ranges. The estimated effect of sorting on institutional preference is a highly significant 16%, on average. The estimated effect of providing a formal cash incentive in this same model was estimated to be substantially larger at 24%, but the difference between these two coefficients is not statistically significant. As before, we also include interaction terms with skills ratings to provide deeper insight into the generation of the distribution of outcomes. As in the earlier case of overall performance,

[9] If the analyses were simply descriptive, we might instead model the probability of achieving a problem-solving score greater than zero, and then the score, conditional on being greater than zero. However, the chosen approach better reflects the data generation process.

model (9-2), in model (10-2) we find no interaction between skills and sorting and again find a strong positive interaction between skills and the cash incentive (which, when included, erases the significance of cash prizes on their own).

<Insert Figure 5 Non-Parametric Regression of Probability of Working More that Minimum Level of Skills, Stratified by Treatment>

<Insert Table 10 Probability of Working Greater than the Minimum Amount and Problem-Solving Performance, Conditional on Working Greater than the Minimum Level of Effort>

Problem-Solving Performance Conditional on Exerting the Minimum Level of Effort

With the share of high-effort individuals (alternatively, the share who do not try in earnest) clearly an important contributor to the overall sorting effect, it remains to be determined whether the sorting effect appears in performance, conditional on having exerted more than the minimum level of effort. A simple comparison suggests that it does: the average *ProblemSolvingScore* conditional on *HoursWorked* > 1 was 4,596 in the sorted group versus just 4,281 in the unsorted group, a difference of 315 points (7%). To provide more precise estimates, we study this comparison within a regression framework. (We focus here on problem-solving performance, *ProblemSolvingScore*, as the key dependent variable, conditional on exerting effort. However, the results are closely mirrored in our effort measures given the close link between them.)

We follow the same basic re-weighting approach as earlier (cf., Section 5.3). However, an important difference in this case is that here we are analyzing a subset of the sorted group and subset of the unsorted group. Therefore, we must recalculate propensity weights for the unsorted control group (subset), as distinct from the weights in the earlier analysis. The procedure is identical, however, the data to which the procedure is applied differs. A second difference is that these subsets are no longer identically distributed skills distributions. Therefore, although the relative weights of observations within each skills band of the unsorted control group (subset) are re-weighted to fix the propensity to compete to the population average, the bands themselves are re-weighted to set the skills distribution of the unsorted control group (subset) to be the same as the sorted group (subset).

As above, we begin our analysis with a graphical presentation of differences between the sorted and unsorted groups, as in Figure 5. (We do not present patterns conditional on skills in this case, as was done in Figure 4 or Figure 5, as the fewer data points in this subset lead to far less precise estimates, particularly at high skill ratings at which there are already fewer data points in the full sample.) Perhaps the first and plainest pattern that can be discerned from Figure 5 is that all distributions are relatively smooth and flat, not uniform, but "thickly" distributed across different problem-solving scores. Further, the sorted group is clearly distributed "to the right" of both the initial control group distribution and the re-weighted control group. The difference with the re-weighted control group is even greater because in the group of unsorted workers it was the relatively high-skilled workers who would chose to participate. Therefore, the re-weighting entails reducing the weight on these workers in the statistical comparison.

As reported in model (10-3), the estimated average effect of sorting (after re-weighting the control group, controlling for the presence of a cash prize and controlling for individual skills) is 1,301 points, which represents about half a standard deviation in overall variation of *ProblemSolvingScore* (Table 3). The effect of the presence of a cash prize is estimated to be considerably smaller at 413, and the coefficient is statistically insignificant. In the case of problem-solving performance conditional on exerting above the minimum effort, we find no evidence of interactions, as in model (10-2).

6 Summary and Conclusion

In this paper, we report evidence from a 10-day "sorting" field experiment involving more than 500 elite programmers engaged in trying to create a software solution to a real computational engineering problem from NASA. Our aim in this experiment was to investigate how sorting on the basis of institutional preferences of workers affected their effort and problem-solving performance for a creative and cognitively challenging task. To emphasize, our interest here was in relation to workers' preferences for the inherent "rules of the game" to which they would be subjected, as they did their work. Using a novel sorting experiment design, we were able to estimate the effects of sorting on institutional preferences, accounting for skills, extrinsic incentives and institutional details. Central to the design was the use of an outside option regime to gauge institutional preferences, and then a combination of matching and randomized assignment to make relevant comparisons.

Our main finding is simply that the fit-based sorting measured here has significant economic effects, nearly doubling problem-solving performance. This amplified performance is almost entirely explained by the exertion of more effort by the sorted workers. The effect was roughly uniformly across skill levels. We devised a series of supplementary tests to further assure our interpretation of results. Crucially, it should be emphasized that apart from demonstrating sorting beyond just skills, these results demonstrate that sorting not only leads to compositional differences, it leads to *behavioral differences*. Hence, "fit matters" – a great deal.

The average effect of monetary incentives on effort and performance was also large and statistically indistinguishable in magnitude to sorting effects on the kind we measured here. However, beyond this similarity, there were important qualitative differences in the effect of monetary incentives. First the effect was not uniform across skill levels; cash incentives generated a much greater response from higher-skilled participants (consistent with their having a higher probability of winning). Cash prizes also appeared to act mostly by getting participants to work more than the minimum level of effort. While sorting also boosted the fraction of participants working more than the minimum level of effort, it also led to more effort and performance, conditional on having worked more than this minimum level. (We found no evidence of interactions between sorting and monetary incentives.)

The field context of this experiment was largely chosen to allow us to observe real problem-solvers addressing a real cognitively challenging problem, the solution to which would be used in practice. Nonetheless, it remains a question how generalizable we should regard these large effects of assuring fit of workers and their institutional preferences. On one hand, the pool of experimental participants were themselves highly selected (not drawn not from the wider labor market of 3 million software developers (King et al. 2010), but instead from an elite subset of developers who joined TopCoder and then chose to join the experiment). In a sense, we observe "treatments on the treated" or, rather, "sorting among the sorted." In this sense, effects might be regarded as conservative. Estimated effects should be even more conservative if considering sorting of workers across the economy to entirely different industries, to entrepreneurial firms, government bureaucracy, academia and so on. The idea of workers of equal competence working twice (or many more times) as hard is easily imaginable. At the same time, it should be noted that there is arguably at least as much variety of organizations and workers in software as in other industries (Mowery 1996; Cusumano, MacCormack, Kemerer, and Crandall 2003; Cusumano 2004); there is perhaps at least as much scope for sorting of heterogeneous workers in this industry as there are in others.

References

Aghion, P., M. Dewatripoint, and J.C. Stein. 2008. "Academic freedom, private-sector focus, and the process of innovation." *RAND Journal of Economics* 39(3): 617-635.

Aghion, P., and J. Tirole. 1994. "The Management of Innovation." *The Quarterly Journal of Economics* 109(4):1185-1209.

Ariely, D., U. Gneezy, G. Lowenstein, and N. Mazar. 2009. "Large Stakes and Big Mistakes." *Review of Economic Studies* 76(2): 451-469.

Azoulay, P., J.G. Zivin, and G. Manso. 2009. "Incentives and Creativity: Evidence from the Academic Life Sciences." NBER Working Paper #15466.

Bandiera, O., I. Barankay, and I. Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics* 120(3): 917-962.

Bandiera, O., L. Guiso, A. Prat, and R. Sadun. 2010. "Matching Firms, Managers and Incentives." Harvard Business School Working Paper 10-073.

Beecham, S., N. Baddoo, T. Hall, and H. Robinson. 2008. "Motivation in Software Engineering: A systematic literature review." *Information and Software Technology* 50: 860-878.

Benabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies* 70(3): 489-520.

Besley, T., and M. Ghatak. 2004. "Competition and incentives with motivated agents." *American Economic Review* 95(3): 616-636.

Bohnet, I., D. Kübler. 2005. Compensating the cooperators: Is sorting in the prisoner's dilemma possible? *Journal of Economic Behavior and Organization* 56: 61-76.

Boudreau, K.J., N. Lacetera, and K.R. Lakhani. 2011. "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis." *Management Science* forthcoming.

Brain. D. 1991. "Practical knowledge and occupational control: The professionalization of architecture in the United States." *Sociological Forum* 6(2):239-268

Cadsby, B., C., F. Song, and F. Tapon. 2007. "Sorting and Incentive Effects of Pay-for-Performance: An Experimental Investigation." *The Academy of Management Journal* 50(2): 387-405.

Camerer, C.F., and D. Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 89(1): 306-318.

Caves, Richard E. 2000. *Creative Industries: Contracts between Art and Commerce.* Cambridge: Harvard University Press.

Couger, J.D., and R.A. Zawacki. 1980. *Motivating and Managing Computer Personnel*. New York: Wiley.

Cusumano, M., A. Macormack, C. Kemerer, and B. Crandall. 2003. "Software Development Worldwide: The State of the Practice." *IEEE Software* 20(6): 28-34.

Cusumano, M.A. 2004. *The Business of Software: What Every Manager, Programmer, and Entrepreneur Must Know to Thrive and Survive in Good Times and Bad*. New York: Free Press.

Dal Bó, P., A. Foster, and L. Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100(5):2205-2229.

Dasgupta, P., and P.A. David. 1994. "Towards a new economics of science." *Research Policy* 23: 487-524.

Dohmen, T., and A. Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *The American Economic Review* 101(2): 556-590.

Eriksson, T., and M.C. Villeval. 2008. "Performance-Pay, Sorting and Social Motivation." *Journal of Economic Behavior & Organization*, 68: 412-421.

Eriksson, T., S. Teyssier, and M. Villeval. 2009. "Does Self-Selection Improve the Efficiency of Tournaments?" *Economic Inquiry* 47(3): 530-548.

Franceschelli, I., S. Galiani, and E. Gulmez. 2010. "Performance pay and productivity of low- and high-ability workers." *Labour Economics* 17: 317-322

Frey, B.S. 1994. "How Intrinsic Motivation is Crowded Out and In." *Rationality and Society* 6: 334-352

Frey, B.S. 2000. *Arts and Economics: Analysis and Cultural Policy*. Berlin: Springer Verlag.

Gneezy, U., M. Niederle, and A. Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118(3): 1049-1074.

Harris, M., and Artur Raviv. 1978. "Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement." *The American Economic Review* 68(1): 20-30.

Hölmstrom, B. 1979. "Moral hazard and observability." *The Bell Journal of Economics* 10(1): 74-91.

Hölmstrom, B. 1989. "Agency costs and innovation." *Journal of Economic Behavior and Organization* 12: 305-327.

Jovanovic, B. 1979. "Job Matching and the Theory of Turnover." *The Journal of Political Economy* 87(5): 972-990.

King M., S. Ruggles, J.T. Alexander, S. Flood, K. Genadek, M.B. Schroeder, B. Trampe, and R. Vick. 2010. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].

Kreps, D.M. 1997. "The Interaction Between Norms and Economic Incentives." *The American Economic Review* 87(2): 359-364.

Larkin, I., and S. Leider. 2010. Why Do Firms Use Non-Linear Incentive Schemes? Experimental Evidence on Sorting and Overconfidence." Harvard Business School Working Paper 10-078.

Lazear, E.P. 2000. "Performance Pay and Productivity." *American Economic Review* 90(5): 1346-1362.

Lazear, E.P., U. Malmendier, and R.A. Weber. 2011. "Sorting, Prices and Social Preferences." Unpublished manuscript.

Lepper, M., and D. Greene. 1978. *The Hidden Cost of Reward: New Perspectives on the Psychology of Human Motivation.* New York: John Wiley.

Maas, Han L. J. van der, and Eric-Jan Wagenmakers. 2005. "A Psychometric Analysis of Chess Expertise." *The American Journal of Psychology* 118(1): 29-60.

Manso, Gustavo. 2011. "Motivating Innovation." *Journal of Finance* forthcoming.

Mowery, D.C. 1996. *The International Computer Software Industry: A Comparative Study of Industry Evolution and Structure*. New York: Oxford University Press.

Nasar, J. L. 1999. *Design by Competition: Making Design Competitions Work*. Cambridge: Cambridge University Press.

Oberholzer-Gee, F., and R. Eichenberger. 2008. "Fairness in Extended Dictator Game Experiments." *The B.E. Journal of Economic Analysis & Policy* 8(1): art. 16.

Rosen, S. 1986. "The theory of equalizing differences." *Handbook of Labor Economics*.

Salop, J., and S. Salop. 1976. "Self-Selection and Turnover in the Labor Market." *Quarterly Journal of Economics* 90(4): 619-627.

Schneiderman, B. 1980. *Software Psychology: Human Factors in Computer and Information Systems*. Scott Foresman & Co.

Shearer, Bruce. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *Review of Economic Studies* 71: 513-534.

Shi, L. 2010. "Incentive Effect of Piece-Rate Contracts: Evidence from Two Small Field Experiments." *The BE Journal of Economic Analysis & Policy* 10(1).

Stern, Scott. 2004. "Do scientists pay to be scientists?" *Management Science* 50(6): 835-854.

Throsby, D. 1994. "The production and consumption of the arts: A view of cultural economics." *Journal of Economic Literature*.

Weisbrod, B.A. 1983. "Nonprofit and proprietary sector behavior: Wage differentials among lawyers." *Journal of Labor Economics* 1(3): 246-263.

Figures

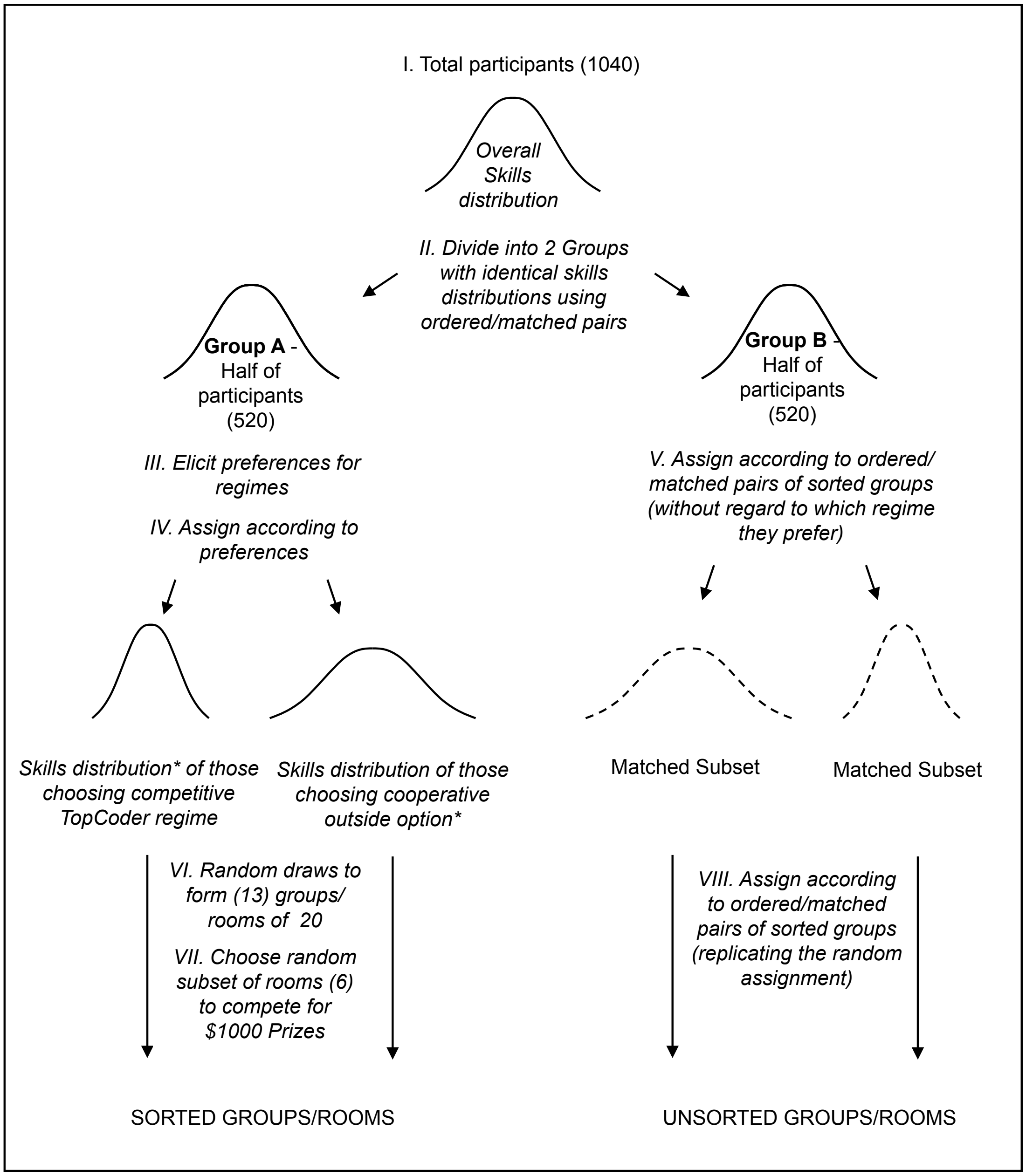

I. Total participants (1040)
Overall Skills distribution
II. Divide into 2 Groups with identical skills distributions using ordered/matched pairs
Group A - Half of participants (520)
Group B - Half of participants (520)
III. Elicit preferences for regimes
IV. Assign according to preferences
V. Assign according to ordered/ matched pairs of sorted groups (without regard to which regime they prefer)
Skills distribution* of those choosing competitive TopCoder regime
Skills distribution of those choosing cooperative outside option*
Matched Subset
Matched Subset
VI. Random draws to form (13) groups/ rooms of 20
VII. Choose random subset of rooms (6) to compete for $1000 Prizes
VIII. Assign according to ordered/matched pairs of sorted groups (replicating the random assignment)
SORTED GROUPS/ROOMS
UNSORTED GROUPS/ROOMS

Figure 1 Illustration of the Assignment Procedure

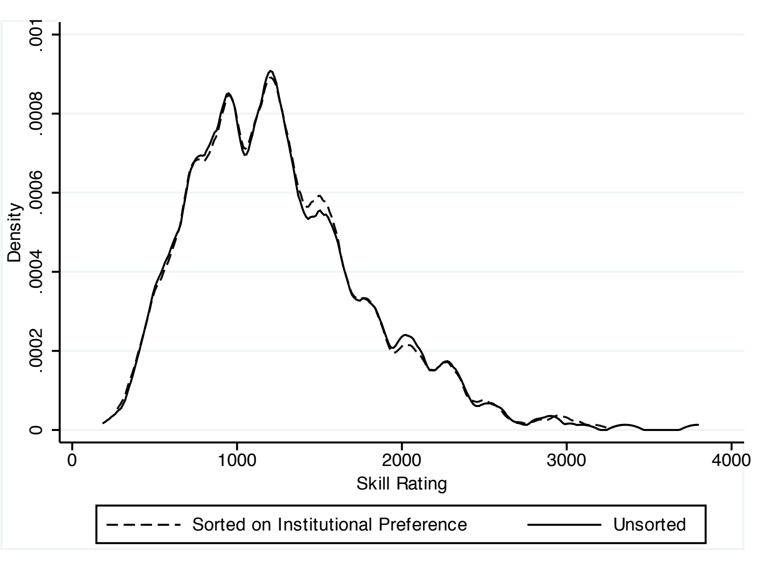


Figure 2 Kernel Density Skills Distribution for Sorted and Unsorted Groups

Note: The lines in the figure are kernel density estimates of the frequency of observations across different levels of the problem-solving skill rating. The density is estimated with an Epanechnikov kernel. A very narrow bandwidth of 50 was chosen to highlight the closeness of the skills distributions.

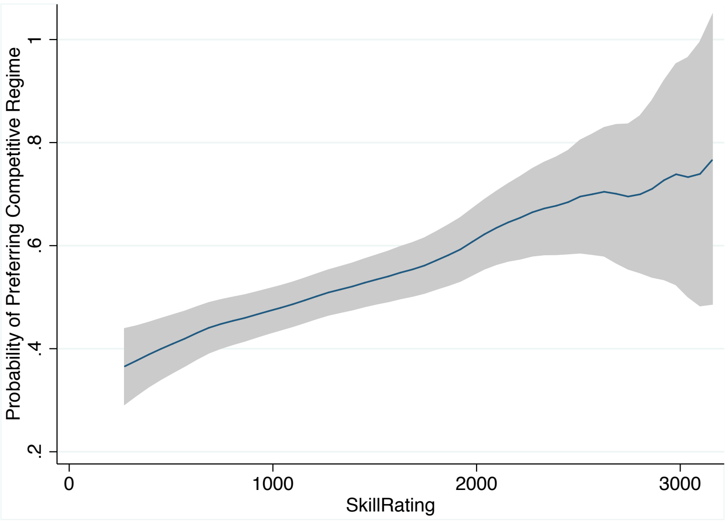


Figure 3 Proportion Preferring Competitive TopCoder Regime Over the Outside Option (Cooperative Regime), by Skill Level

Note: The line fits to a series of 1's and 0's depending on whether the individual preferred TopCoder's competitive regime (1) or the outside option (0). The line is a locally-weighted fitted second-order polynomial. The local weighting is based on an Epanechnikov kernel with a bandwidth of 300. The shaded grey region represents the 90% confidence interval for the estimate.

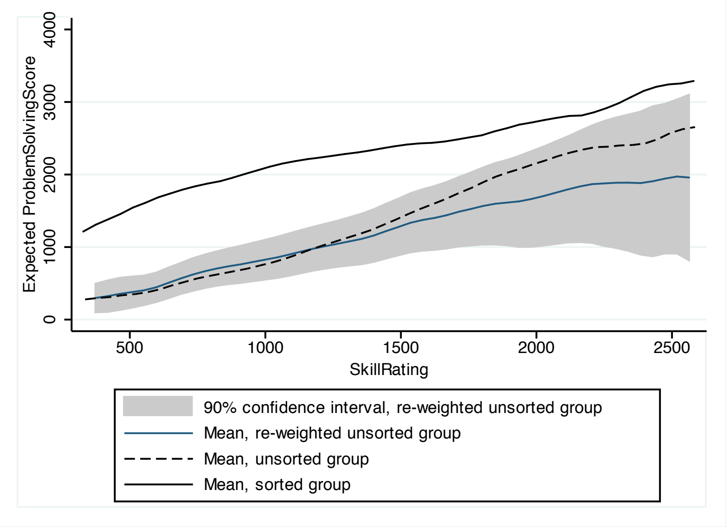


Figure 4 Non-Parametric Regression of Problem-Solving Performance, Stratified by Treatment

Note: Each of the lines fits a locally-weighted fitted second-order polynomial, with local weighting based on an Epanechnikov kernel with a bandwidth of 300. The solid black line is the relationship for the group that has been sorted on the basis of its preference for the TopCoder competitive regime. The dashed black line is the relationship for a group that has not been sorted on its preferences, but with an identical skills distribution. The blue line is the same unsorted group, the data points of which have been re-weighted according to steps described in Section 5.3. The shaded grey region represents the 90% confidence interval for the estimate.

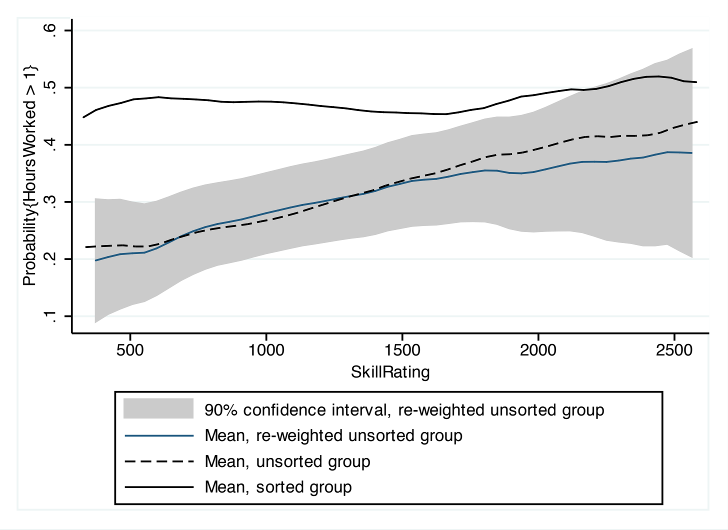


Figure 5 Non-Parametric Regression of Probability of Working More than Minimum Level on Skills, Stratified by Treatment

Note: Each of the lines fits a locally-weighted fitted second-order polynomial, with local weighting based on an Epanechnikov kernel with a bandwidth of 300. The solid black line is the relationship for the group that has been sorted on the basis of its preference for the TopCoder competitive regime. The dashed black line is the relationship for a group that has not been sorted on its preferences, but with an identical skills distribution. The blue line is the same unsorted group, the data points of which have been re-weighted according to steps described in Section 5.3. The shaded grey region represents the 90% confidence interval for the estimate.

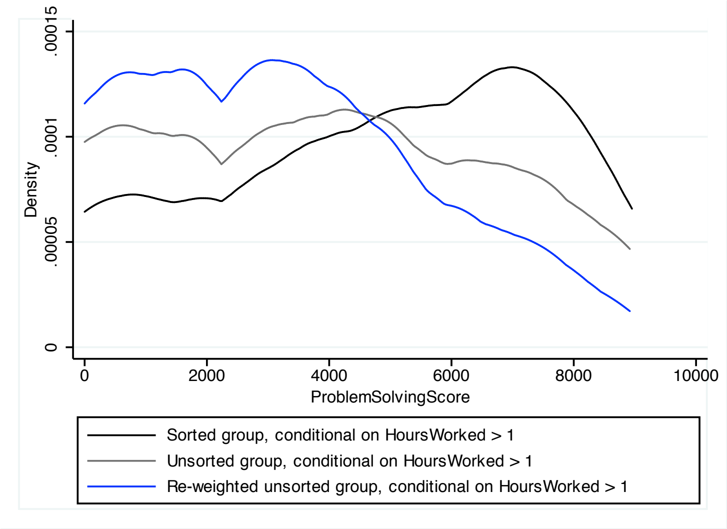


Figure 5 Kernel Density of Problem-Solving Performance, Stratified by Treatments

Note: The lines in the figure are kernel density estimates of the frequency of observations across different levels of the problem-solving skill rating. The density is estimated with an Epanechnikov kernel with a bandwidth of 1,000. The black line is the density for the group that has been sorted on the basis of its preference for the TopCoder competitive regime. The grey line is the density for a group that has not been sorted on its preferences, but with an identical skills distribution. The blue line is the same unsorted group, the data points of which have been re-weighted according to steps described in Section 5.

Tables

Table 1 Key Features of the Competitive TopCoder Regime and the (Cooperative) Outside Option Regime

	COMPETITIVE TOPCODER REGIME	OUTSIDE OPTION REGIME
Size of a Group	20 competitors	4 x 5-person teams (assigned)
Payoffs	Total: $1000, divided 5-ways among top 5 submiters ($500, $200, $125, $100, $75)	Total: $1000, divided 5-ways among winning team members (according to average of team members' suggestions)
Communications & Code Sharing	None	A private team-message board and ability to send directed messages
Information	Competitors "see" who else is in the group, their ratings and top code submissions to date	Competitors "see" best scores to date of other teams; detailed information on the statistics and background of their own team members

Table 2 Variable Definitions

Variable	Definition
ProblemSolvingScore	Numerical score awarded to a solution as an assessment of overall quality, based on automated test suite
NumSubmissions	Number of solutions submitted to be compiled, tested and scored by an individual participant during the course of the experiment
HoursWorked	Number of hours worked by an individual participant during the course of the experiment
SortedonPreference	Indicator switched to one for participants who were asked their preferences regarding the regimes and subsequently assigned to their preferred regime
Prize	Indicator switched to one for participants within a group of 20 that competed for a $1000 cash prize
SkillRating	Measure of general problem solving ability in Algorithmic problems based on historical performance on TopCoder platform

Table 3 Means, Standard Deviations and Correlations

	Variable	Mean	Std. Dev.	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)
(1)	*ProblemSolvingScore*	1736	2802	0	8957	1.00					
(2)	*NumSubmissions*	2.89	6.32	0	42	.74	1.00				
(3)	*HoursWorked*	10.80	20.68	0	190	.61	.55	1.00			
(4)	*SortedonPreference*	.50	.50	0	1	.18	.14	.21	1.00		
(5)	*Prize*	.43	.50	0	1	.25	.22	.22	.00	1.00	
(6)	*SkillRating*	1344	546	328	3354	.22	.17	-.02	.00	.02	1.00

Table 4 Comparison of Mean Outcomes, Stratified by Treatment

UNSORTED ON INSTITUTIONAL PREFERENCE					
NO CASH PRIZE			CASH PRIZE		
Variable	Average	Standard Deviation	Variable	Average	Standard Deviation
ProblemSolvingScore	578	582	*ProblemSolvingScore*	2070	3052
NumSubmissions	.78	2.50	*NumSubmissions*	3.55	6.76
HoursWorked	3.48	3.29	HoursWorked	10.70	17.17

SORTED ON INSTITUTIONAL PREFERENCE					
NO CASH PRIZE			CASH PRIZE		
Variable	Average	Standard Deviation	Variable	Average	Standard Deviation
ProblemSolvingScore	1682	2754	*ProblemSolvingScore*	2976	3214
NumSubmissions	2.58	5.92	*NumSubmissions*	5.38	8.53
HoursWorked	10.16	19.39	*HoursWorked*	21.42	30.32

Table 5 Baseline OLS Regression Results

	Dependent Variable = *ProblemSolvingScore*				
Model:	(5-1)	(5-2)	(5-3)	(5-4)	(5-5)
Explanatory Variables	Two-Way Correlation	Linear Sklls Control	Skills-Level Dummies	Matched Pair Differences	Prize Control
SortedonPreference	1,016*** (243)	1,016*** (237)	1,009*** (236)	1,042*** (235)	1,010*** (229)
Prize					1,324*** (239)
Skills-Level Dummies			Yes		Yes
Constant	1,223*** (153)	-248 (281)			
Adj R-Squared	.03	.08	.08	.12	.14

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity robust standard errors reported; number of observations = 516 participants.

Table 6 Results of Effort and Activity Regressions

	Number of Submissions	*Hours Worked*	*Problem Solving Score*
Model:	(6-1)	(6-2)	(6-3)
Specification	Count Model, Multiplicative Matched Pair Fixed Effects		Linear model, matche pair differences
SortedonPreference	.65*** (.19)	.84*** (.16)	216 (163)
HoursWorked			157*** (16)
HoursWorked^2			-.85*** (.20)
Log-Likelihood	-707	-2298	
Adj R-Squared			.55

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity robust standard errors reported; number of observations = 516 participants.

Table 7 Results of Tests for Social Interactions

	Dependent Variable = ProblemSolvingScore		
Model:	(5-3)	(7-1)	(7-2)
Explanatory Variables	Matched Pair Differences		
SortedonPreference	1,009*** (236)	1,055*** (354)	1,082*** (361)
Others in Same Room			
Mean		-.04 (.24)	-.13 (.49)
Variance			.49 (1.09)
Skew			-49 (534)
Max			-.22 (.31)
Skills-Level Dummies	Yes	Yes	Yes
Adj R-Squared	.08	.12	.11

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity robust standard errors reported; number of observations = 516 participants.

Table 8 Logit Model Results of Probability / Propensity to Prefer Competitive Regime

Dependent Variable:	*Dependent Variable = I{Prefer Competitive TopCoder Regime*				
Model:	(8-1)	(8-2)	(8-3)	(8-4)	(8-5)
Explanatory Variables	Skills Level Dummies	Motivations	Age	Employed	Countries of Origin
<u>Raw Problem-Solving Skill</u>					
900 ≤ SkillRating < 1200	.54**	.52**	.52**	.53**	(.23)
	(.25)	(.26)	(.26)	(.26)	(.28)
1200 ≤ SkillRating < 1250	.47*	.45*	.45	.47*	.07
	(.26)	(.27)	(.28)	(.28)	(.29)
1500 ≤ SkillRating < 2200	.98***	.93***	.97***	.99***	.59**
	(.26)	(.27)	(.27)	(.27)	(.30)
2200 ≤ SkillRating	1.18***	1.15***	1.21***	1.24***	.81*
	(.43)	(.43)	(.43)	(.44)	(.47)
<u>Motivation for Joining TopCoder</u>					
"Cash Prizes"		.33	.26	.24	.20
		(.30)	(.31)	(.31)	(.32)
"Employment Opportunity"		-.28	-.33	-.41	-.44
		(.36)	(.36)	(.37)	(.39)
"Technology Competition"		.64***	0.53**	.50**	.45*
		(.22)	(.23)	(.23)	(.24)
<u>Age</u>					
18-28			.41	.37	-.35
			(.57)	(.57)	(.29)
25-34			.59	.37	-.55
			(.59)	(.63)	(.40)
35-44			1.32*	1.05	-.01
			(.70)	(.75)	(.61)
≥45			2.36*	2.18*	.98
			(1.24)	(1.25)	(1.16)
Declined to Answer			.57	.39	-.35
			(.84)	(.86)	(.76)
<u>Other</u>					
"Professional" (versus "Student")				.28	.41
				(.27)	(.29)
Country of Origin Dummies					Yes
Log Likelihood	-345	-338	-333	-332	-319

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; number of observations = 520 participants.

Table 9 Regression Results with Synthesized Control Group

	Dependent Variable = *ProblemSolvingScore*		
Model:	(5-5)	(9-1)	(9-2)
Explanatory Variables	Unweighted	Re-Weighted by Propensity for Competitive Regime	
SortedonPreference	1,010*** (229)	1,140*** (235)	1,198* (626)
Sorted x Skill			.03 (.45)
CashPrize	1,324*** (239)	1,360*** (246)	-1028 (643)
Prize x Skill			1.86*** (.50)
Skills-Level Dummies	Yes	Yes	Yes(*)
Adj R-Squared	.14	.07	.03

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity robust standard errors reported; number of observations = 516 participants in the cooperative regime; 524 in the cooperative regime. (*) A linear control for SkillRating is included in addition to the dummies.

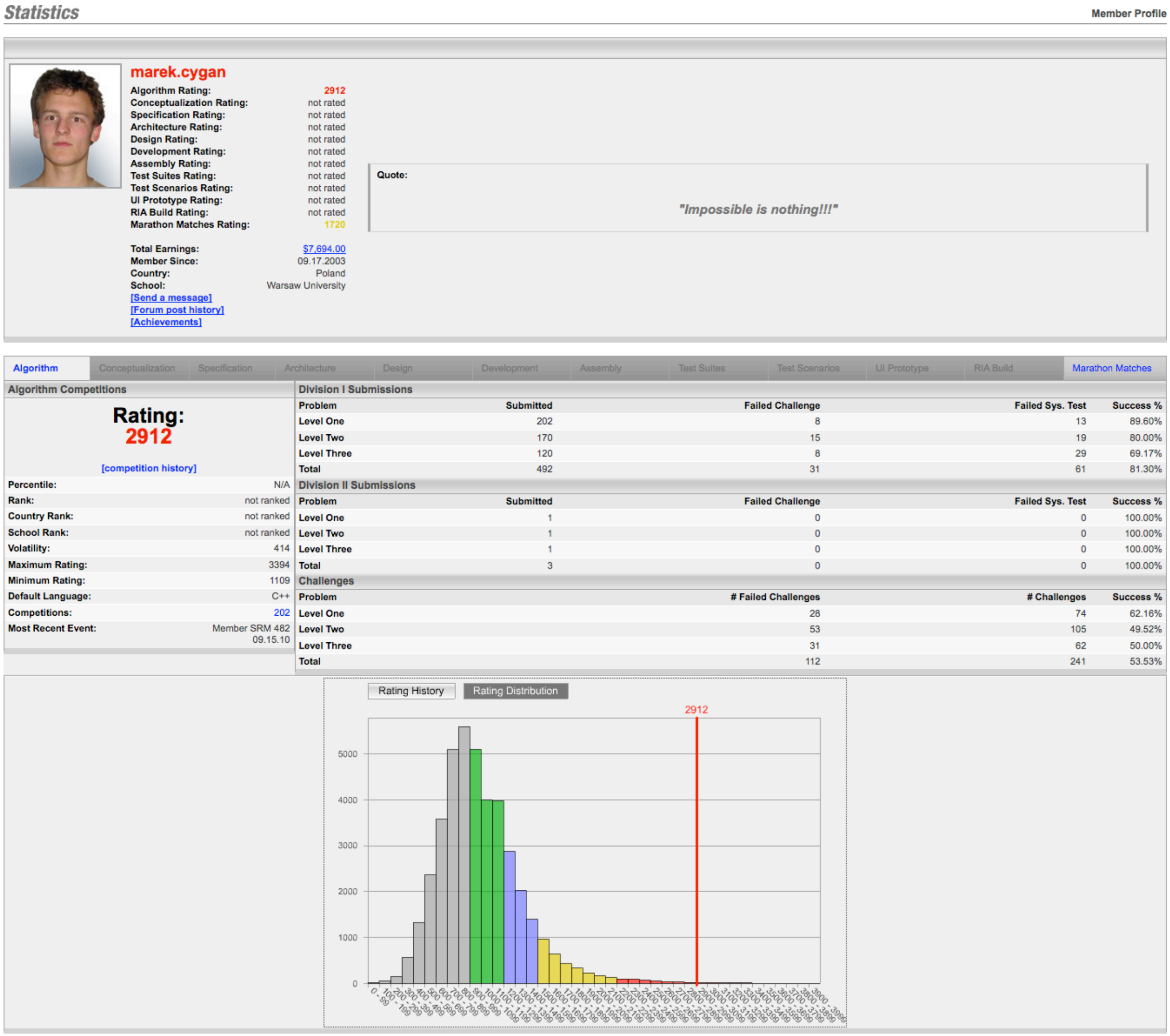
Table 10 Probability of Working Greater than the Minimum Amount and Problem-Solving Performance, Conditional on Working Greater than the Minimum Level of Effort

Dependent Variable:	*I{HoursWorked > 1}*		*ProblemSolvingScore | HoursWorked > 1*	
Model:	1	2	3	4
Explanatory Variables	Synthesized Control Group	Add Interactions with Skill	Synthesized Control Group	Add Interactions with Skill
SortedonPreference	.16*** (.04)	.31*** (.12)	1,301*** (416)	2,467* (1437)
Sorted x Skill		0 (0)		-.83 (1.05)
CashPrize	0.2424*** (.05)	-.12 (.12)	413.70 (391.07)	492.22 (1134.60)
Prize x Skill		0.0003*** (.00)		-.07 (.79)
Skills-Level Dummies	Yes	Yes(*)	Yes	Yes(*)
Adj R-Squared	.09	.11	.14	.14

Notes. *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively; heteroskedasticity robust standard errors reported; number of observations = 516 participants in the cooperative regime; 524 in the cooperative regime. (*) A linear control for SkillRating is included in addition to the dummies.

Appendices : NOT FOR PUBLICATION

APPENDIX 1: Participant Ratings



Statistics

Member Profile

marek.cygan

Algorithm Rating:	2912
Conceptualization Rating:	not rated
Specification Rating:	not rated
Architecture Rating:	not rated
Design Rating:	not rated
Development Rating:	not rated
Assembly Rating:	not rated
Test Suites Rating:	not rated
Test Scenarios Rating:	not rated
UI Prototype Rating:	not rated
RIA Build Rating:	not rated
Marathon Matches Rating:	1720
Total Earnings:	$7,694.00
Member Since:	09.17.2003
Country:	Poland
School:	Warsaw University

[Send a message]
[Forum post history]
[Achievements]

Quote:

"Impossible is nothing!!!"

Algorithm Competitions

Rating:
2912

[competition history]

Percentile:	N/A
Rank:	not ranked
Country Rank:	not ranked
School Rank:	not ranked
Volatility:	414
Maximum Rating:	3394
Minimum Rating:	1109
Default Language:	C++
Competitions:	202
Most Recent Event:	Member SRM 482 09.15.10

Division I Submissions

Problem	Submitted	Failed Challenge	Failed Sys. Test	Success %
Level One	202	8	13	89.60%
Level Two	170	15	19	80.00%
Level Three	120	8	29	69.17%
Total	492	31	61	81.30%

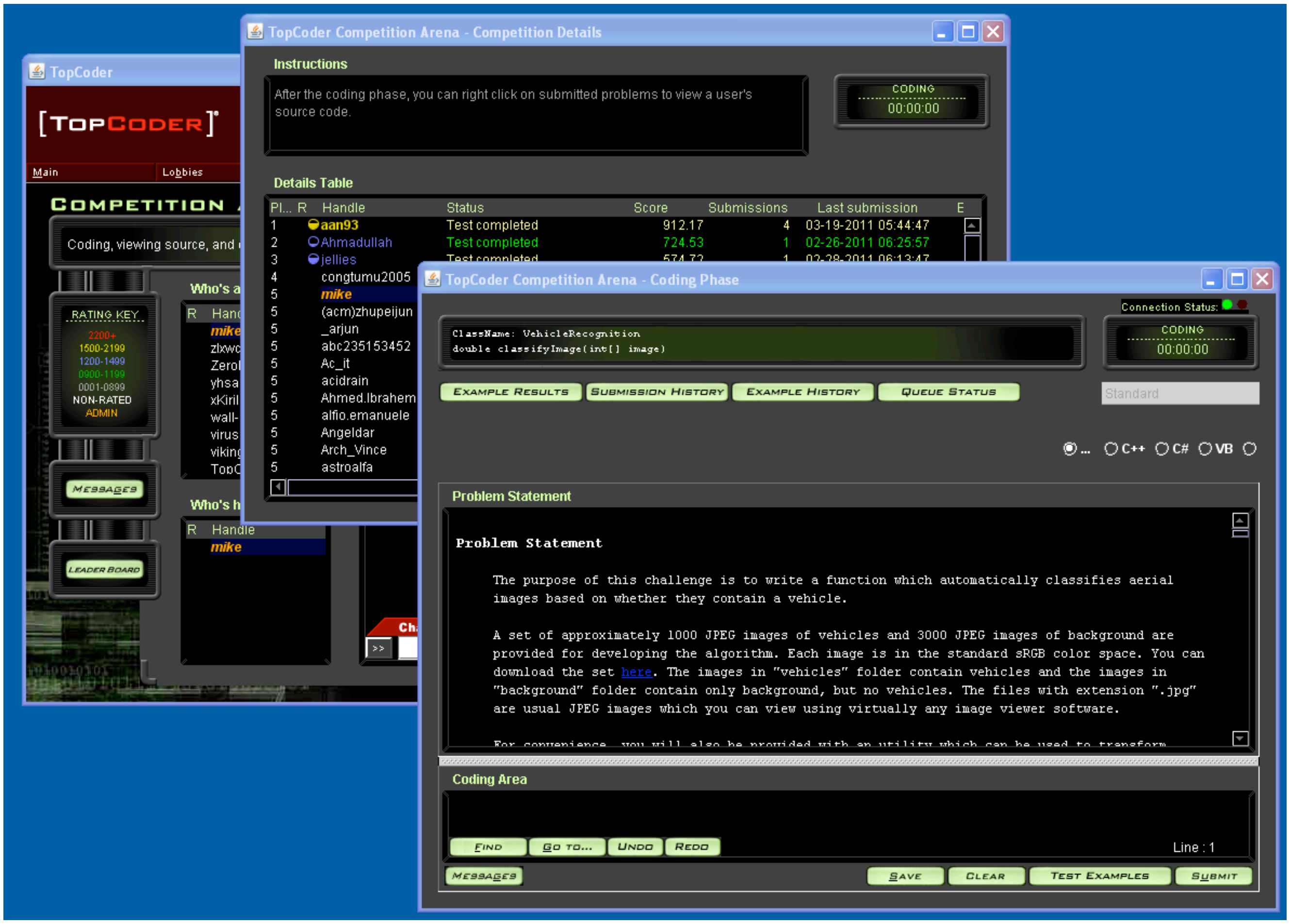
Division II Submissions

Problem	Submitted	Failed Challenge	Failed Sys. Test	Success %
Level One	1	0	0	100.00%
Level Two	1	0	0	100.00%
Level Three	1	0	0	100.00%
Total	3	0	0	100.00%

Challenges

Problem	# Failed Challenges	# Challenges	Success %
Level One	28	74	62.16%
Level Two	53	105	49.52%
Level Three	31	62	50.00%
Total	112	241	53.53%

Appendix 2: TopCoder Participant Arena Screen Shots



TopCoder Competition Arena - Competition Details

Instructions

After the coding phase, you can right click on submitted problems to view a user's source code.

CODING 00:00:00

Details Table

Pl...	R	Handle	Status	Score	Submissions	Last submission	E
1		aan93	Test completed	912.17	4	03-19-2011 05:44:47	
2		Ahmadullah	Test completed	724.53	1	02-26-2011 06:25:57	
3		jellies	Test completed	574.72	1	02-28-2011 06:13:47	
4		congtumu2005					
5		mike					
5		(acm)zhupeijun					
5		_arjun					
5		abc235153452					
5		Ac_it					
5		acidrain					
5		Ahmed.Ibrahem					
5		alfio.emanuele					
5		Angeldar					
5		Arch_Vince					
5		astroalfa					

TopCoder Competition Arena - Coding Phase

Connection Status:

```
ClassName: VehicleRecognition
double classifyImage(int[] image)
```

CODING 00:00:00

Standard

● ... ○ C++ ○ C# ○ VB ○

Problem Statement

Problem Statement

The purpose of this challenge is to write a function which automatically classifies aerial images based on whether they contain a vehicle.

A set of approximately 1000 JPEG images of vehicles and 3000 JPEG images of background are provided for developing the algorithm. Each image is in the standard sRGB color space. You can download the set here. The images in "vehicles" folder contain vehicles and the images in "background" folder contain only background, but no vehicles. The files with extension ".jpg" are usual JPEG images which you can view using virtually any image viewer software.

For convenience, you will also be provided with an utility which can be used to transform

Coding Area

Line : 1

APPENDIX 3: Problem Statement

Problem: SpaceMedkit

You have been asked to assist the space medicine community in stocking a space vehicle with appropriate medical resources to mitigate the likelihood of medical evacuation of crew members during space flights. The space vehicle has mass and volume constraints that limit the amount of medical resources that can be flown. To complete this task, you have agreed to create an optimization algorithm that identifies the best possible medical kit (medkit) for meeting constraints on the number of crew member evacuations (**P**) while minimizing the medical resource mass and volume.

For your optimization, the space medicine community will provide you with a list of approved medical resources, with unit mass and volume. Medical resources in the list will be classified as consumable or non-consumable. Consumable resources can be used only once; non-consumables can be used multiple times.

In order to build the optimization, you will be provided with data from a previously developed mission simulation. Each trial in the simulation provides data for a fully treated (all required medical resources are available), and an untreated (not all required medical resources are available) scenario, including the occurrence of a crewmember evacuation. In the simulation, full treatment of a condition does not always prevent evacuation, but it does generally lower the probability of evacuation.

Inputs

The parameters described below will be constant for all tests, and are also available for download. The only parameters that will vary between tests are **P** and **C**.

1. **availableResources** -- this parameter will give you the different medical resources that you may include in your medkit. Each element will be formatted as "RID CONSUMABLE MASS VOLUME".
   - RID is an alphanumeric identifier specific to the resource.
   - CONSUMABLE is either 0 or 1, where 1 indicates that the resource will be used up in treatment (like a drug, for instance) and 0 indicates that the resource can be reused (like a thermometer).
   - MASS and VOLUME are self-explanatory
2. **requiredResources** -- this parameter will describe the different medical events that might occur on the missions. Each event can take one of two courses: a best case course, and a worst case course. These two courses require different resources for treatment. For simplicity, there is no middle ground; the event will follow one of these two courses. Each element of this parameter will be formatted as "MID RID BEST WORST".
   - MID is an alphanumeric identifier specific to the medical event. Note that multiple elements will have the same MID.
   - RID is the resource ID (matching the previously described input).
   - BEST is the amount of this type of resource required if the event takes the best course.
   - WORST is the amount of this type of resource required if the event takes the worst course.

   (MID,RID) is a unique key for this input, and thus no two elements will have the same value for both of these fields.

3. **missions** -- this parameter will describe a number of missions. Your goal is to design your medkit tailored to these missions. This input should be considered the training data, as your medkit will be evaluated on a different set of missions generated by the same simulation. Each element will be formatted as "MISSION ORDER MID WORST TREATED UNTREATED".
   - MISSION is an id number for the mission.
   - ORDER specifies the order within a mission that events occur (each mission will be sorted by this in the input).
   - MID is an alphanumeric identifier specific to the medical event.
   - WORST is 1 if the worst case course of this event occurred, and 0 otherwise (best case).
   - TREATED specifies the number of evacuations if this event is treated.
   - UNTREATED specifies the number of evacuations if this event is untreated.

Output

You should design a medkit and return a String[] wherein each element is formatted as "RID QUANTITY", indicating that the resource QUANTITY of RID should be included (this may be a floating point value).

Your return will be evaluated on each mission independently (resources are restocked between missions). For a mission, the events will be evaluated one by one (according to ORDER). If all of the resources are available to treat the event (under the condition -- best or worst -- that occurs), those resources will be used to treat it. The number of evacuations from the event for the treatment status that occurs will be added to the total number of evacuations. Note that, for simplicity, each medical event is considered independent of the outcome of previous events. This total will be evaluated over all missions. In pseudocode:

```
foreach mission
    restock resources according to your output
    foreach event in mission (in order)
        if all resources available to treat event
            evacuations += event.treated
            decrement consumed resources
        else
            evacuations += event.untreated
```

Scoring

For each test case, your input will be evaluated on a set of 10,000 missions, randomly selected from a corpus of 200,000. The average number of evacuations per mission must be no more than the input **P**. Thus, the total number of evacuations summed over all missions must be no more than **P*10000**. Given that, your score will be 1000 / (mass + **C** * volume), where **C** is an input parameter. If the evacuations rate exceeds **P**, your score will be 0 for that test case. Your overall score will be the sum of your individual scores.

Appendix 4: Eliciting Preferences

Choice Survey Email Communication:

Subject: Mandatory Survey - NASA-TopCoder Challenge – Please respond in 24 hours

Dear <Handle Name>,

We are considering you as one of the participants for next week's TopCoder-NASA Marathon Match Challenge. We would like you to complete a short, three question survey regarding the contest. Please complete the survey within 24 hours.

We appreciate your attention to this. Given the experimental nature of this event, we require that you not disclose the existence of these questions through personal communications, email, blog postings, forum postings or any other means---or risk disqualification.

Thank for your help and cooperation!

Best,
Mike Lydon
Chief Technology Officer

Please proceed to the following link: <insert link>

Survey Questions:

a) Version 1

Q1 As you know, we are investigating new ways of participating in TopCoder experiments. Some people will be able to work in teams.

Might you be interested in joining a team to compete against other teams?

I DEFINITELY would prefer to join a team
I MIGHT prefer to join a team
I am indifferent or I am not sure

I MIGHT prefer to compete on my own
I DEFINITELY would prefer to compete on my own

Q2 As further clarification, both teams and individual competitors will be in groups of 20 (4x5-person teams or 20 individuals). There will be 5 cash prizes awarded in each group, either to the winning team or each of the top 5 individuals. So the chances of winning---in terms of the prizes per each group of 20 people--are the same for both individual and group formats.

Team members will be free to share ideas and code with one another over a private discussion board. The team will be evaluated as a group, with the best submission of the group representing the group's final submission.

Please confirm or adjust your previous answer:

I DEFINITELY would prefer to join a team
I MIGHT prefer to join a team
I am indifferent or I am not sure
I MIGHT prefer to compete on my own
I DEFINITELY would prefer to compete on my own

Q3 Finally, here is a hypothetical question. Imagine if TopCoder were always to offer the options of joining a team or competing on your own. What is the best guess of the percentage of events for which you would join a team:

I would always join teams (100% of the time in teams)
I would mostly join teams (>80% of the time in teams)
I would frequently join teams (60%-80% of the time in teams)
I would join both roughly equally (40%-60% of the time in teams)
I would somewhat regularly join teams (20%-40% of the time in teams)
I would occasionally join teams (<20% of the time in teams)
I would never join teams (0% of the time in teams)

As a reminder please do not disclose this survey to anyone else.

b) Version 2

Q1 As you know, we are investigating new ways of participating in TopCoder experiments. Some people will be able to work in teams.

Might you be interested in joining a team to compete against other teams?

I DEFINITELY would prefer to compete on my own

I MIGHT prefer to compete on my own
I am indifferent or I am not sure
I MIGHT prefer to join a team
I DEFINITELY would prefer to join a team

Q2 As further clarification, both teams and individual competitors will be in groups of 20 (4x5-person teams or 20 individuals). There will be 5 cash prizes awarded in each group, either to the winning team or each of the top 5 individuals. So the chances of winning---in terms of the prizes per each group of 20 people--are the same for both individual and group formats.

Team members will be free to share ideas and code with one another over a private discussion board. The team will be evaluated as a group, with the best submission of the group representing the group's final submission.

Please confirm or adjust your previous answer:
I DEFINITELY would prefer to compete on my own
I MIGHT prefer to compete on my own
I am indifferent or I am not sure
I MIGHT prefer to join a team
I DEFINITELY would prefer to join a team

Q3 Finally, here is a hypothetical question. Imagine if TopCoder were always to offer the options of joining a team or competing on your own. What is the best guess of the percentage of events for which you would join a team:

I would never join teams (0% of the time in teams)
I would occasionally join teams (<20% of the time in teams)
I would somewhat regularly join teams (20%-40% of the time in teams)
I would join both roughly equally (40%-60% of the time in teams)
I would frequently join teams (60%-80% of the time in teams)
I would mostly join teams (>80% of the time in teams)
I would always join teams (100% of the time in teams)

As a reminder please do not disclose this survey to anyone else.