Proceedings of the ASME 2018 International Design Engineering Technical Conferences & Computers in Engineering Conferences
IDETC/CIE 2018
August 26-29, 2018, Quebec City, Quebec, Canada

DETC2018-85457

DRAFT: DESIGNING IMPROVED TEAMS FOR CROWDSOURCED COMPETITIONS

Christopher McComb
Engineering Design
The Pennsylvania State University
University Park, PA, USA
mccomb@psu.edu

Torsten Maier
Industrial Engineering
The Pennsylvania State University
University Park, PA, USA
torstennamaier@psu.edu

ABSTRACT

Teams are ubiquitous, woven into the fabric of engineering and design. Often, it is assumed that teams are better at solving problems than individuals working independently. Recent work in engineering, design, and psychology has indicated that teams may not be the problem-solving panacea that they were once thought to be. Crowdsourcing has seen increased interest in engineering design recently, and platforms often encourage teamwork between participants. This work undertakes an analysis of the performance of different team styles and sizes in crowdsourced competitions. This work demonstrates that groups of individuals working independently may outperform interacting teams on average, but that small interacting teams are more likely to win competitions. These results are discussed in the context of motivation for crowdsourcing participants.

1. INTRODUCTION

Large and small businesses, sports clubs, and social groups rely on teams to complete shared goals. In this way, team efficiency and performance are often crucial to task completion. Teams are particularly common in engineering and business [1,2]. Many emerging platforms for crowdsourcing can be viewed as large teams, or as multi-team systems, that compose and organize many different individuals with the objective of solving a single problem. The crowdsourcing of engineering and design problems has seen increased interest in recent years (e.g., [3,4]). This work explores teams at the intersection of crowdsourcing and engineering design by investigating the potential similarities between teams engaged in crowdsourced data science competitions and more traditional engineering design teams. This exploration utilizes data from a number of crowdsourced data science competitions in the Kaggle platform. Data science shares a number of similarities with design [5], including large solution spaces, procedural similarities, and the potential for formalization through C-K theory [6]. These characteristics make the data collected from the Kaggle platform useful for drawing potentially generalizable insights.

Even though teams are common in industry [1,2], there is little agreement between definitions of the word *team* that have been proposed in the literature (for examples see [7–10]). However, most definitions share two concepts: multi-agency (the composition of a team as two or more individuals) and communication (the ability of those individuals to exchange information) [11]. These two concepts are fundamental to understanding team-based solutions. This work aims to provide a deeper knowledge base of both components within the context of data science.

Interaction frequency affects team performance and the sharing of information [12]. Homogeneity and heterogeneity of new venture teams has been shown to have an effect on persistence and performance over time [13]. For this study, teams were categorized into two sub-categories: true and nominal teams. True teams are comprised of team members who share information and resources to create a single combined solution. Nominal teams are defined as coalitions of individuals who work independently but use their best solution.

Team size is an important factor, but findings related to team size are mixed. Several studies have identified a positive relationship between group size and factors such as dissatisfaction and cognitive conflict [14,15]. A meta-analysis of 31 published field studies determined negative results for increased team sizes citing low efficiency as a leading cause [16]. In support of these findings, higher rates of social loafing (the "reduction in motivation and effort when individuals work collectively compared with when they work individually or coactively") have also been linked to increasing group size [17–19]. In contrast, other studies have shown that larger teams may experience increased effectiveness as the heterogeneity of tasks increases [20]. This would indicate that larger teams are more

effective as scope increases, most likely, due to their enlarged experience base. A quantitative review of 93 studies found a slight increase in performance of project and management teams when they include more members [21]. However, further analysis determined that task autonomy led to increased performance for specific tasks [21]. This may indicate that optimal team size is task-dependent. In addition, computational work has demonstrated that smaller teams are capable of making decisions that have better axiomatic characteristics [22].

Team communication frequency is a common metric used when modeling or investigating team dynamics [23,24]. Studies have found a positive relationship between communication frequency and team cognition and cooperation [23,24]. Pentland (2012) found team communication patterns to be the strongest indicator of team success and found that revising the break schedules at call centers to increase teammate socialization lead to increased performance [25]. E-mail and face-to-face interaction frequency was found to have a curvilinear relationship with team performance [26]. Optimal performance was located between high and low interaction frequency. In contrast to these findings, "best-member strategy" has shown to yield similar results to conventional teams [27]. These weak interactive ties facilitate a diverse perspective amongst team members, beneficial in the problem-solving process [28,29].

Agile project management strategies are rapidly growing in popularity in the software development domain, which is relevant for the data science focus of the Meta Kaggle dataset. These are projects that value individuals and interactions, working software, customer collaboration, and responding to change over process and tools, documentation, contract negotiation, and following a plan [30]. In 2002, agile projects accounted for less than 5% of new application development projects [31]. As of 2011, agile projects accounted for 29% of new application development projects [31]. Scrum is the most popular agile framework for developing software products; it can be broken down into 3 key components: the team, events, and artifacts. For the purposes of this paper, only team elements will be discussed. For full information see The Scrum Guide [32].

The scrum team is composed of the Product Owner, the Development Team, and the Scrum Master [32]. The Product Owner is responsible for assessing and maximizing the value of the product. The Development Team is composed of the members developing the product. The Scrum Master is responsible for facilitating scrum processes. Teams work in Sprints that last 1-4 weeks [32]. At the end of each Sprint, product goals are reassessed. New goals are developed based on the Product Backlog and a new Sprint begins. This cyclic system creates iterative solutions to short term projects that comprise the final product.

Optimal scrum team size is based on Miller's number (7±2) [33], the average number of information chunks that a person can hold in short-term memory [34]. It is now used in information theory and user interface design [35,36]. Miller's (1956) work indicates a natural human tendency to chunk, or group, information so that it is manageable for our short-term memory. Studies have shown that team communication amongst large teams can be optimized by dividing them into smaller scrum teams [33]. However, the effectiveness of this is dependent upon how easily tasks can be partitioned and on if significant overhead communication is required [37].

The current work specifically examines software development within the context of data science competitions. The focus of this work on data science teams is not driven by mere convenience - some researchers have claimed that design theory and methodology can provide a necessary framework for data science [5]. We specifically examine teams participating in Kaggle Competitions. Kaggle is a company that hosts online data science and machine learning competitions [38]. The Meta Kaggle data set is public data provided by Kaggle that describes their competitions, submissions, and scores. These competitions range from underwater image analysis that assesses ocean health to analytics in cardiology that assesses heart functionality [39]. This data is analyzed here to compare the effect of different team sizes and team interaction frequencies on team performance. Specifically, this study attempts to answer the following research questions:

1. *Do teams perform better than individuals in Kaggle competitions?*
2. *How do team size and interaction type affect overall solution quality?*
3. *How does the probability of winning (and expected winnings) vary as a function of team size and interaction type?*

The rest of this paper is organized as follows. Section 2 summarizes the Meta Kaggle dataset and details the variables used in this study. Section 3 provides an explanation of the methodology used in this work and outlines the numerical experiments that were conducted. Section 4 details the results of the numerical experiments. Finally, Section 5 concludes the paper with a summary and a discussion of future directions.

2. DATA

This work specifically utilizes the 2016 Meta Kaggle dataset [40] which contains information on 316 different competitions run through Kaggle's platform. This data contains information for 594,892 different users and 7,949 distinct teams.

The Meta Kaggle dataset includes 24 different tables of information for each of those competitions. Of those tables, only five were used for this study: users, teams, submissions, competitions, and team memberships. The Users dataset contains data related to individual Kaggle user's profiles. The Teams dataset contains data related to unique teams (of one or more users). The Submissions dataset corresponds to the data recorded when a competition entry is submitted. The Competitions dataset contains information pertinent to completed Kaggle competitions. The Team Memberships dataset is a key used to link users with their teams. A full list of

the variables used, and their descriptions are provided in Table 1.

TABLE 1. DATASET VARIABLES AND DESCRIPTIONS

Dataset	Variable Name	Description
Users	ID	Unique identifier for the user
	Display Name	User's display name
Teams	ID	Unique identifier for the team
	Team Name	The team's name
	Competition ID	Unique identifier for the competition that the team participated in
	Score	The team's final score
	Ranking	The team's final ranking
Submissions	ID	Unique identifier for the submission
	Submitted User ID	Unique identifier for user who made submission
	Team ID	Unique identifier for the team corresponding to the user who made the submission
Competitions	ID	Unique identifier for the competition
	Competition Name	URL slug for the competition
	Title	title for the competition
Team Memberships	ID	unique identifier for the team membership
	Team ID	Unique identifier for the team that the user belongs
	User ID	Unique identifier for the user

3. METHODOLOGY

This section introduces the analytical methodology employed to investigate the three research questions. First, the primary competitive styles observed in Kaggle competitions are introduced and defined (individual, team, and *nominal* teams simulated here). Second, the metrics used to evaluate and compare these competitive styles are introduced.

3.1. Competitive Styles

Competitors in Kaggle's competitions may compete either as individuals or band together with other individuals as part of a team. We analyze both of those conditions in this work. It should be noted that individual competitors are not completely isolated from the ideas of others. Kaggle implements an extensive forum on which players may share approaches, code, or advice. In addition, we analyze the performance of *nominal teams*. Nominal teams are groups of individuals who do not interact. Rather, they work independently towards a full solution, with the best individually-generated solution being provided as the solution of the team. A growing body of literature demonstrates that nominal teams may outperform interacting teams in tasks such as concept generation [41,42]. In the current work, nominal teams are simulated via a bootstrapping approach in which several individuals from the Meta Kaggle dataset are randomly selected and the best solution found by any individual in that group is taken as the team solution. This methodology has been used in several other studies [11,43,44]. Although other work on concept generation tends to sum or average the outputs of individuals in nominal teams, the approach used here of taking the best solution depicts a reasonable approach to using nominal teams for solving tasks with well-defined and quantitative objectives. The primary comparison featured in this work is between nominal teams are compared against the teams directly available in the Meta Kaggle dataset, which will be referred to as *true teams*. Comparisons are also made against single individual competitors as a baseline.

For the analyses in this paper, only competitions with at least 10 individual competitors and at least 10 team competitors were used. Figure 1 shows the distribution of team sizes for the 127 competitions that met these criteria, and Figure 2 shows the distribution of team sizes for the winning team from each competition. These figures provide additional motivation for the research questions driving this work. Is the high rate of individual winners driven by the prevalence of individual competitors or by some differentiation in performance? This question will be further elucidated in Section 4.

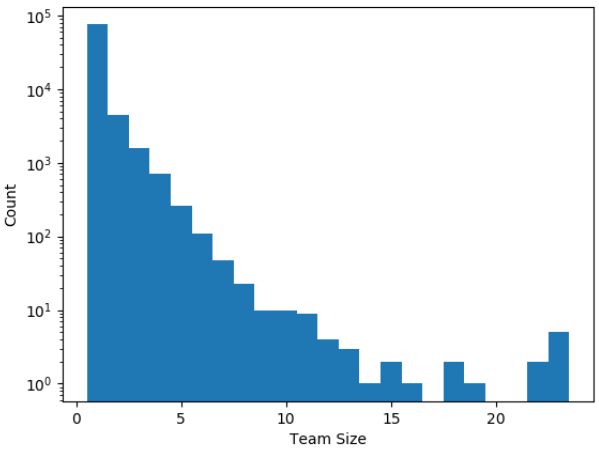

FIGURE 1. TEAM SIZE HISTOGRAM FOR ALL TEAMS.

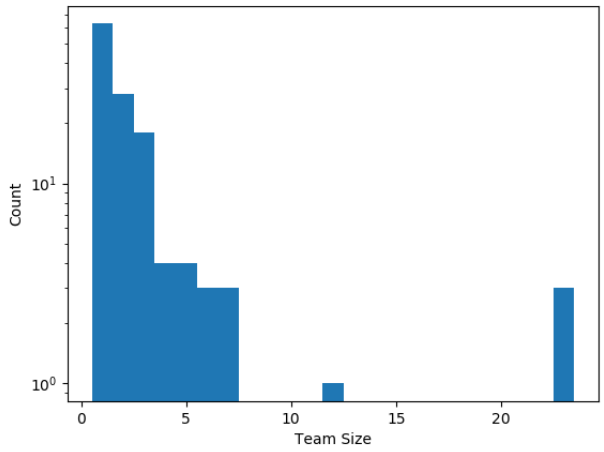

FIGURE 2. TEAM SIZE HISTOGRAM FOR WINNING TEAMS.

3.2. Metrics

Four metrics were utilized in this work to describe team and individual efficacy: solution quality, team effort, probability of winning, and expected payout per team member.

Solution quality is the final score of the algorithm submitted by a team or individual. These scores were normalized for those taking part in a given competition using the equation

$$q' = \frac{q - \mu_q^{indiv}}{\sigma_q^{indiv}}, \quad (1)$$

where q' is the normalized solution quality of a team or individual solution, q is the un-normalized solution quality, μ_q^{indiv} is the average solution quality for individuals competing in that competition, and σ_q^{indiv} is the standard deviation for individuals in that competition. Because of this normalization scheme, the distribution of quality for individuals has a mean of 0 and a standard deviation of 1. It should be noted that this normalization allows solution quality to take on negative values. This approach is relatively robust to the presence of outliers in the quality data, whereas an approach that scales the values to fall in a fixed range (e.g., [0, 1]) would be highly sensitive to outliers.

Effort was quantified as the number of submissions made by a team or individual. A submission enables a team to evaluate the approximate score of their solution – it is analogous to an objective function evaluation in optimization. Effort was normalized for those taking part in a given competition using the equation

$$e' = \frac{e}{\mu_e^{indiv}}, \quad (2)$$

where e' is the normalized effort of a team or individual, e is the un-normalized effort, and μ_e^{indiv} is the average effort for individuals taking part in that competition. This metric will always be positive, as it represents as cardinal quantity. It should be noted that Equations 1 and 2 were both designed to establish the individual condition as a baseline for comparison.

Probability of winning is the likelihood that a team with a given structure and approach will win an average competition. Within the simulation paradigm utilized in this work, this probability is estimated simply as

$$p_{win} = \frac{w}{m}, \quad (3)$$

where p_{win} is the probability of winning, m is the number of simulations conducted, and w is the count of instances in which a team of interest produces a winning solution.

Expected payout per team member quantifies the average fraction of the competition reward that an individual can expect to win when averaged over many competitions (e.g., average winnings over time). This expected fraction is computed by dividing the probability of winning by the number of members on the team as

$$r = \frac{w}{mn}, \quad (4)$$

where r is the expected payout per team member, n is the number of individuals on the team and w and m are as defined previously.

4. RESULTS

This section details the results of three numerical experiments conducted to better understand the role of team performance in the Kaggle competition environment. A full implementation of these analyses is available in the Python language under an MIT License.[1]

4.1. Do teams perform better than individuals?

The results of the comparison between the three competitive styles (individual, true team, and nominal team) are shown in Figure 3. The data for each competition was normalized separately and only aggregated after normalization within the competition. The data points plotted for the true teams and individuals are based directly on data from the Meta Kaggle dataset. Nominal teams were simulated by selecting several individuals and selecting the best individual solution as the team solution. By construction, the distribution of team sizes for the simulated nominal teams matches the distribution of team sizes for the true teams (see Figure 1). Note that the error bars for standard error of the individual data point are too small to appear in the figure because of the large sample size.

True teams put in more average effort (quantified by the number of submission, shown normalized on the x-axis) than individuals, and consequently are capable of producing solutions with higher quality (shown normalized on the y-axis). Nominal teams put in only slightly more effort than true teams and achieve substantially higher solution quality. Since nominal teams and true teams have the same distribution of team size, the difference in the number of submissions (our approximation of effort) is unexpected. For nominal teams, the mean

[1] https://github.com/HSDL/KaggleTeams

normalized effort is equal to the average team size of 2.61 individuals. However, the true teams only put in 2.42 individual worth of effort on average. In other words, members of true teams put in 92.7% of the effort that would be expected if they competed individually. This difference in expended effort between true and nominal teams is potentially evidence of social loafing.

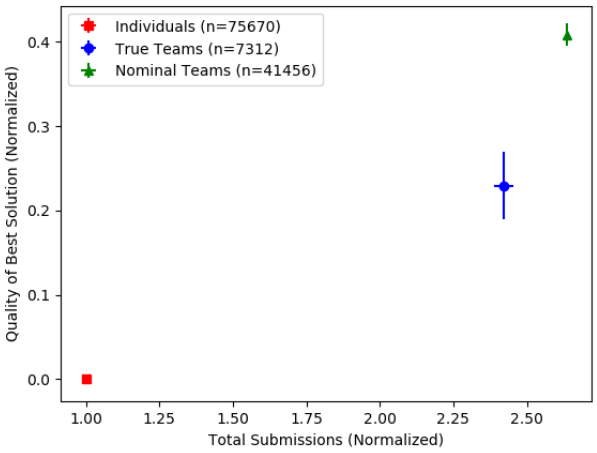

FIGURE 3. COMPARISON OF INDIVIDUALS, INTERACTING TEAMS, AND NOMINAL TEAMS. ERROR BARS SHOW ±1 STANDARD ERROR.

Moreover, the pattern of individual variability within each one of the competitive styles was assessed by examining the relationship between effort and solution quality. Results are summarized in Table 2. For each competitive style, the table provides the slope (rate of increased performance for increased effort) the standard error of the estimate of the slope, and the p-value of the estimate.

TABLE 2.CORRELATION BETWEEN EFFORT AND PERFORMANCE FOR DIFFERENT COMPETITIVE STYLES

Competitive Style	Slope	Standard Error	p-value
Individual	0.160	0.002	<0.001
True Team	0.073	0.009	<0.001
Nominal Team	0.015	0.001	<0.001

The strength of the relationship between performance and effort decreases as the average performance of the competitive style increases. Individuals can expect a performance increase of 0.16 if they increase their performance by 1 individual (essentially doubling their effort). True teams, however, can only expect half of that increase in performance for the same increase in effort. This provides some insight into the underlying cause of the social loafing effect observe in Figure 3. If individuals in a true team are aware that increasing their effort only marginally benefits the team, this could lead them to infer that *decreasing* their effort will only marginally *harm* the team, giving rise to social loafing behavior.

The effort-performance relationship for nominal teams is even weaker, more than an order of magnitude less than the relationship for individual competitors. Herein lies the potential peril of using a nominal teaming approach in practice. In individuals become aware that their increased efforts have only a small effect on team performance, then they may be more apt to engage in social loafing behavior. This possibility is not capture in the current work because nominal teams are simulated without a mechanism for social effects.

4.2. How do team size and interaction type affect overall solution quality?

The analyses in the remainder of this paper were limited to team sizes for which more than 10 instances were observed in the Kaggle dataset (see histogram in Figure 1). Thus, the maximum team size explored here is 8, as larger team sizes show substantial variance due to the low number of instances.

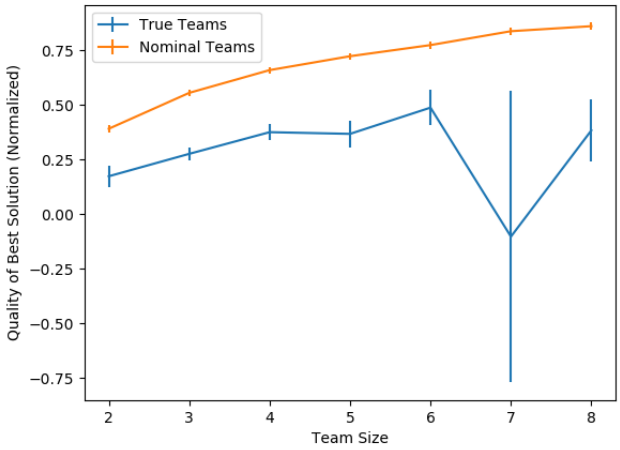

FIGURE 4. SOLUTION QUALITY VERSUS TEAM SIZE FOR TRUE AND NOMINAL TEAMS. ERROR BARS SHOW ±1 STANDARD ERROR.

The results of the analysis of team performance as a function of team size are provided in Figure 4. Exactly 1000 nominal teams were generated for each team size shown in the figure. The number of true teams for each team size is provided in Figure 1. This analysis confirms that the effect observed in Figure 3 is robust across team sizes. A Kruskal-Wallis test was used to compare the performance of true and nominal teams for each team size (Kruskal-Wallis was chosen over ANOVA because homoscedasticity was violated). Every comparison was statistically significant ($p < 0.001$). On average, nominal teams consistently outperform true teams by approximately 0.25 standard deviations. Of particular note is the fact that 2-person nominal teams offer performance that is on par with 4-person true teams. In addition, it should be recognized that the average solution quality for 7-person true teams is substantially lower than adjacent team sizes, and the standard error is higher. This

difference is driven by a small number of teams with very low solution quality.

Further, the relationship between cumulative team effort and team size is provided in Figure 5. A Kruskal-Wallis test was again used to compare the effort of true and nominal teams for each team size. Every comparison was statistically significant ($p < 0.001$). The nominal team effort follows a diagonal line with a one-to-one increase in effort for every individual added. This is expected, as nominal teams are simulated coalitions of individuals. The difference in effort displayed between the nominal and true team conditions grows at a constant rate for team sizes of up to 6, reaching a maximum of almost 50%. This aligns with other work on social loafing that indicates an higher predisposition towards the behavior in larger teams [17–19]. However, for team sizes of 7-8, the difference in effort shrinks and standard error increases. This indicates that some teams were prone to social loafing while others were not. One possible explanation is that some of these larger teams employed an approach that utilized sub-teams or a clear decomposition of work into sub-tasks, effectively negating negative social effects.

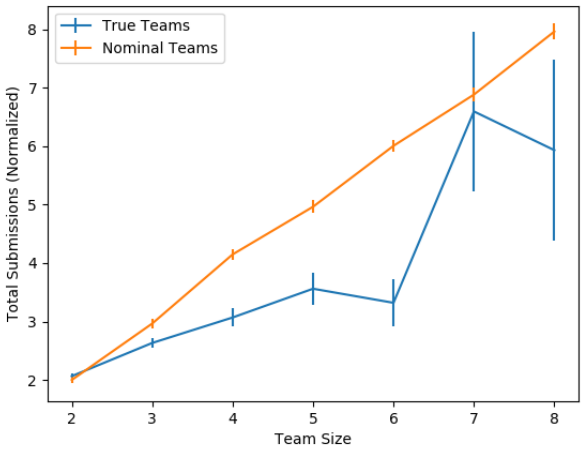

FIGURE 5. TOTAL SUBMISSIONS VERSUS TEAM SIZE FOR TRUE AND NOMINAL TEAMS. ERROR BARS SHOW ±1 STANDARD ERROR.

4.3. How does the probability of vary as a function of team size and interaction type?

Previous sections analyzed the average performance of different competitive styles and assessed them with respect to different team sizes. However, average performance of a group is not necessarily a strong indication of how that group will perform in an authentic environment. In this section, we simulate the actual competitiveness of both nominal and true teams of varying size in an average competition. Specifically, an average competition in the Meta Kaggle dataset has 653 competitive entities (either individuals or true teams). In order to simulate competitions, we employed a statistical bootstrapping approach that leveraged the available Kaggle data. To generate one competition, we first randomly sampled 653 competitive entities from the Meta Kaggle dataset in such a way that they followed the size distribution shown in Figure 1. Ten thousand such competitions were randomly generated. For a given team size and style, ten thousand instances of the team were sampled from the Meta Kaggle dataset. Then, every combination of the set of average competitions and the set of teams was assessed by injecting the simulated team into the simulated competition. This resulted in 10 million different simulated competitions. This made it possible to estimate the relatively low probabilities of winning with high accuracy. Figure 6 shows the probability of winning an average competition (i.e., having a solution with the highest score), and Figure 7 shows the expected individual payout per team member.

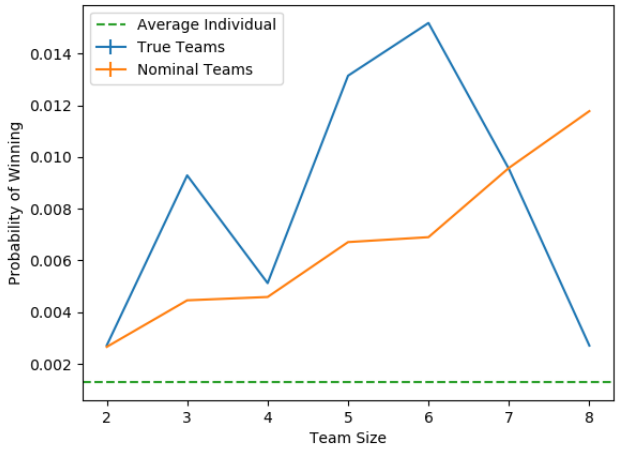

FIGURE 6. PROBABILITY OF WINNING AN AVERAGE COMPETITION AS A FUNCTION OF TEAM SIZE FOR BOTH TRUE TEAMS AND NOMINAL TEAMS. BARS FOR STANDARD ERROR ARE TOO SMALL TO APPEAR ON THIS FIGURE BECAUSE OF THE LARGE SAMPLE SIZE.

In Figure 6, the probability of winning for true teams following a roughly parabolic shape, with a peak at a team is of 6. For that size, the probability of winning is approximately 1.5%. For a competition with nearly 700 competitive entities, this is significantly above what would be expected by chance. The probability of winning for the nominal teams, in contrast, increases almost linearly for the team sizes shown. The probability of winning with any team competitive style is substantially higher than the probability of winning as an individual.

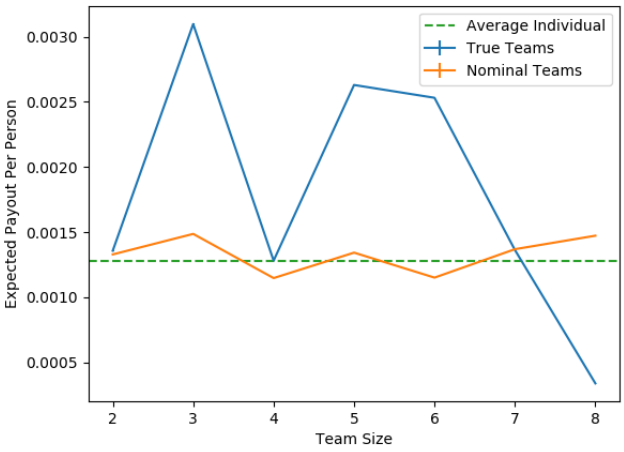

FIGURE 7. EXPECTED INDIVIDUAL PAYOUT IN AN AVERAGE COMPETITION AS A FRACTION OF TOTAL PRIZE AMOUNT. BARS FOR STANDARD ERROR ARE TOO SMALL TO APPEAR ON THIS FIGURE BECAUSE OF THE LARGE SAMPLE SIZE.

Figure 7 shows the average expected individual payout to each team member as a fraction of total competition reward (probability of winning divided by team size). This represents the average payout that an individual could expect if they consistently employ a single competitive style over time. The highest observed value for true teams occurs at a size of 3, although team sizes of 5 or 6 offer comparable payouts. Interestingly, the expected individual payout for members of nominal teams is essentially equivalent to the expected payout for an individual competing alone, regardless of team size. In other words, an individual has no personal monetary incentive to join a nominal team; they can expect to reap the same return regardless of nominal team size. However, joining true teams with sizes of 3, 5, or 6 could yield significantly higher returns.

The results for both probability of winning and individual payout seem counter-intuitive, especially in comparison to Figure 4 that clearly demonstrates that nominal teams display better mean performance than true teams. However, in competitions such as these, mean performance of a group directly indicate a winning team. By their very nature, winning teams typically fall in the extreme tail of the distribution they belong to. Therefore, the parameter of interest in predicting win probability is not strictly mean performance, but mean performance in combination with standard deviation. To illustrate this fact in greater detail, the 95th percentile of solution quality as a function of team size is provided in Figure 8. This demonstrates that the highest performing true teams deliver solutions with higher quality than the highest performing nominal teams, helping to explain the results in Figures 6 and 7. This result provides a new lens through which to view the growing literature on the superiority of nominal teams relative to interacting teams. Although nominal teams may provide better mean performance, that performance has smaller standard deviation. The higher standard deviation of interacting teams makes higher performance accessible with a higher probability, increasing the likelihood that they will succeed in competitive environments. This may help to explain the prevalence of teams in practice for both engineering and business [1,2].

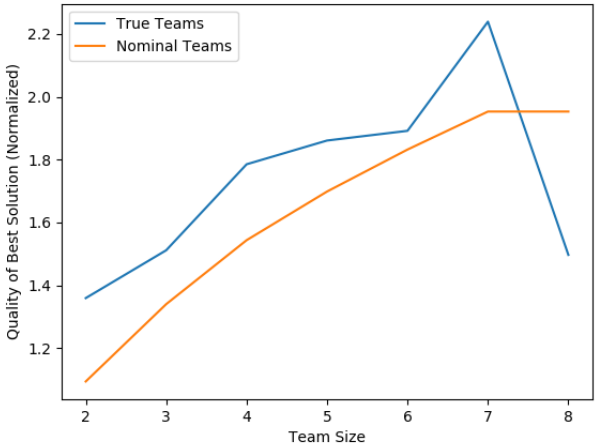

FIGURE 8. 95TH PERCENTILE OF SOUTION QUALITY AS A FUNCTION OF TEAM SIZE FOR TRUE AND NOMINAL TEAMS.

These analyses also have deeper significance for interpreting the motivations of the participants in Kaggle's competitions. If these competitors were motivated purely by the prestige associated with winning, it is expected that they would be competing in larger teams of five or six. If they were motivated purely by monetary rewards they would likely compete in teams of three to maximize individual payout. However, the dominant competitive style in Kaggle is individual. This hints that these extrinsic motivations (prestige and compensation, respectively) may pale in comparison to intrinsic motivation offered by the Kaggle platform and community. This competition between intrinsic and extrinsic motivation in crowdsourcing has been noted by other authors [45–47].

5. CONCLUSIONS

This work examined the behavior of teams engaged in crowdsourced data science competitions in order to gain insights relevant to engineering design as both researchers and practitioners embrace crowdsourcing. Specifically, the analysis investigated the Meta Kaggle dataset, which contains data from a large number of data science competitions conducted in the Kaggle crowdsourcing platform. This data was utilized to conduct three numerical experiments.

The first experiment assessed the relative performance of three different competitive styles: individual, true team (those teams observed in the Meta Kaggle dataset) and nominal teams (simulated coalitions of individuals for which the best individual solution becomes the team solution). This analysis

showed that true teams perform better than individuals (though with much greater effort). However, nominal teams outperform true teams with only minimal additional effort.

The second experiment explored the robustness of the first experimental result for teams of different sizes. It was observed that nominal teams provide solutions of higher quality than true team for every team sizes assessed. It was also observed that the effect of social loafing increases with team size until a size of 7-8, at which point teams may naturally develop some degree of internal structure of task decomposition.

The third experiment moved beyond the assessment of average performance and injected nominal and true teams into a series of simulated competition environments. The results from these simulated competitions ran counter to those from the first two experiments: true teams were more likely to win competitions nominal teams. This is caused by the fact that true teams experience higher performance variability than nominal teams. Therefore, although their mean performance is lower, the upper tail of their distribution reaches higher. This helps to explain the prevalence of interacting teams in industry contexts.

One important feature of the data used here is the inherently quantitative evaluation of potential solutions that is afforded by the formulation of typical data science tasks. Often, tasks in engineering and design lack objectives that are easy to quantify. However, the rising prevalence of ratings-based surveys and crowdsourced evaluation approaches (e.g. [48,49]) promises to make quantifiable objectives more common in engineering design.

Social loafing was a considerable factor in the true teams analyzed here. However, the nominal teams that were simulated had no mechanism for social loafing. It is possible that simply belonging to a team could lead to social loafing even in the absence of communication with teammates. Therefore, future work should evaluate the role of social loafing in nominal teams. To this end, validated models of team behavior such as Cognitively Inspired Simulated Annealing Teams framework [29] may be useful in addition to behavioral human studies.

Other research has shown that expertise plays an important role in determining crowdsourcing outcomes [50], but the current work does not approximate the expertise of Kaggle competitors. This should be remedied in future work to offer greater insight for the motivation that competitors have to either join teams or compete individually. It may be possible to assess expertise by examining performance on past competitions as well as forum posts (free-form text comments available in the Meta Kaggle dataset).

The primary results from this work are the following: (1) nominal teams may provide superior average performance in crowd-sourced environments; (2) social loafing plays a large role in true teams and should be investigated in nominal teams as well; and (3) the inherent variability of true teams may enable them to outperform nominal teams in competitive environments. This work exists at the interaction of engineering design, crowdsourcing, and data science, and thus these results may be of use to each of these communities.

ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency through cooperative agreement N66001-17-1-4064. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] Paulus, P. B., Dzindolet, M. T., and Kohn, N., 2011, "Collaborative Creativity-Group Creativity and Team Innovation," *Handbook of Organizational Creativity*, M.D. Mumford, ed., Elsevier, New York, pp. 327–357.

[2] Reiter-Palmon, R., Herman, A. E., and Yammarino, F. J., 2008, "Creativity and Cognitive Processes: Multi-Level Linkages between Individual and Team Cognition," *Multi-Level Issues in Creativity and Innovation (Research in Multi Level Issues, Volume 7)*, M.D. Mumford, S.T. Hunter, and K.E. Bedell-Avers, eds., pp. 203–267.

[3] Ren, Y., Bayrak, A. E., and Papalambros, P. Y., 2016, "EcoRacer: Game-Based Optimal Electric Vehicle Design and Driver Control Using Human Players," J. Mech. Des., **138**(6), p. 61407.

[4] Ball, Z., and Lewis, K., 2018, "Observing Network Characteristics in Mass Collaboration Design Projects," Des. Sci., **4**, p. e4.

[5] Kazakci, A. O., 2015, "Data Science as a New Frontier for Design," *International Conference on Engineerring Design*, pp. 1–10.

[6] Hatchuel, A., and Weil, B., 2009, "C-K Design Theory: An Advanced Formulation," Res. Eng. Des., **19**(4), pp. 181–192.

[7] Cannon-Bowers, J. A., Salas, E., and Converse, S., 1993, "Shared Mental Models in Expert Team Decision Making," *Individual and Group Decision Making: Current Issues*, N.J. Castellan, ed., pp. 221–246.

[8] Dyer, D. J., 1984, "Team Research and Team Training: A State-of-the-Art Review," *Human Factors Review: 1984*, F. Muckler, ed., Human Factors Society, Santa Monica, pp. 285–323.

[9] Orasanu, J. M., and Salas, E., 1993, "Team Decision Making in Complex Environments," *Decision Making in Action: Models and Methods*, G. Klein, J. Orasanu, R. Calderwood, and C. Zsambok, eds., Ablex Publishers, Norwood, NJ, pp. 327–345.

[10] Morgan, B., Glickman, A., Woodard, E., Blaiwes, A., and Salas, E., 1986, *Measurement of Team Behaviors in a Navy Environment*.

[11] McComb, C., Cagan, J., and Kotovsky, K., 2017, "Optimizing Design Teams Based on Problem Properties: Computational Team Simulations and an Applied Empirical Test," J. Mech. Des., **139**(4), p. 41101.

[12] Mello, A. S., and Ruckes, M. E., 2006, "Team Composition," J. Bus., **79**(3), pp. 1019–1039.

[13] Steffens, P., Terjesen, S., and Davidsson, P., 2012, "Birds of a Feather Get Lost Together: New Venture Team Composition and Performance," Small Bus. Econ., **39**(3), pp. 727–743.
[14] Mullen, B., Symons, C., Hu, L.-T., and Salas, E., 1989, "Group Size, Leadership Behavior, and Subordinate Satisfaction," J. Gen. Psychol., **116**(May), pp. 155–170.
[15] Amason, A. C., and Sapienza, H. J., 1997, "The Effects of Top Management Team Size and Interation Norms on Cognitive and Affective Conflict," J. Manage., **23**(4), pp. 495–516.
[16] Gooding, R. Z., 2018, "A Meta-Analytic Review of the Relationship between Size and Performance : The Productivity and Efficiency of Organizations and Their Subunits Author (S): Richard Z . Gooding and John A . Wagner III Source : Administrative Science Quarterly , Vol . 30 , ," **30**(4), pp. 462–481.
[17] North, A. C., Linley, P. A., and Hargreaves, D. J., 2000, "Social Loafing in a Co-Operative Classroom Task," Educ. Psychol. An Int. J. Exp. Educ. Psychol., **20**(4), pp. 454–475.
[18] Lam, C., 2015, "The Role of Communication and Cohesion in Reducing Social Loafing in Group Projects," Bus. Prof. Commun. Q., **78**(4), pp. 454–475.
[19] Karau, S., and Williams, K., 1993, "Social Loafing: A Meta-Analytic Review and Theoretical Integration," J. Pers. Soc. Psychol., **65**(4), pp. 681–706.
[20] Majuika, R. J., and Baldwin, T. T., 1991, "Team-Based Employee Involvement Programs for Continuous Organizational Improvement: Effects of Design and Administration," Pers. Psychol., **44**, pp. 793–812.
[21] Stewart, G. L., 2006, "A Meta-Analytic Review of Relationships between Team Design Features and Team Performance," J. Manage., **32**(1), pp. 29–55.
[22] McComb, C., Goucher-Lambert, K., and Cagan, J., 2017, "Impossible by Design? Fairness, Strategy, and Arrow's Impossibility Theorem," Des. Sci., **3**(e2).
[23] He, J., Butler, B., and King, W., 2007, "Team Cognition: Development and Evolution in Software Project Teams," J. Manag. Inf. Syst., **24**(2), pp. 261–292.
[24] Pinto, M., 1990, "Project Team Communication and Cross-Functional Cooperation in New Program Development," J. Prod. Innov. Manag., **7**(3), pp. 200–212.
[25] Pentland, A., 2012, "The New Science of Building Great Teams Analytics for Success The New Science of Building Great Teams : Analytics for Success," Harv. Bus. Rev., (April).
[26] Patrashkova-Volzdoska, R. R., McComb, S. A., Green, S. G., and Compton, W. D., 2003, "Examining a Curvilinear Relationship between Communication Frequency and Team Performance in Cross-Functional Project Teams," IEEE Trans. Eng. Manag., **50**(3), pp. 262–269.
[27] Yetton, P. W., and Bottger, P. C., 1982, "Individual versus Group Problem Solving: An Empirical Test of a Best-Member Strategy," Organ. Behav. Hum. Perform., **29**(3), pp. 307–321.
[28] Granovetter, M., and Granovetter, M., 1983, "The Strength of Weak Ties: A Network Theory Revisited," Sociol. Theory, **1**, pp. 201–233.
[29] McComb, C., Cagan, J., and Kotovsky, K., 2015, "Lifting the Veil: Drawing Insights about Design Teams from a Cognitively-Inspired Computational Model," Des. Stud., **40**, pp. 119–142.
[30] Chow, T., and Cao, D.-B., 2008, "A Survey Study of Critical Success Factors in Agile Software Projects," J. Syst. Softw., **81**(6), pp. 961–971.
[31] The Standish Group, 2011, "CHAOS Manifesto: The Laws of CHAOS and the CHAOS 100 Best PM Practices."
[32] Schwaber, K., and Sutherland, J., 2017, "Scrum Guide."
[33] Al Qurashi, S., and Qureshi, M. R. J., 2014, "Scrum of Scrums Solution for Large Size Teams Using Scrum Methodology," Life Sci. J., **11**(8), pp. 443–449.
[34] Miller, G. A., 1956, "The Magical Number Seven," Psychol. Rev., **63**, pp. 81–97.
[35] Garg, T., 2016, "Graphical User Interface for Antimicrobial Peptide Database."
[36] Jamieson, R. K., Nevzorova, U., Lee, G., and Mewhort, D. J. K., 2016, "Information Theory and Artificial Grammar Learning: Inferring Grammaticality from Redundancy," Psychol. Res., **80**(2), pp. 195–211.
[37] Di Penta, M., Harman, M., Antoniol, G., and Qureshi, F., 2007, "The Effect of Communication Overhead on Software Maintenance Project Staffing: A Search-Based Approach," IEEE Int. Conf. Softw. Maintenance, ICSM, pp. 315–324.
[38] Carpenter, J., 2011, "May the Best Analyst Win," Science, **331**(6018), pp. 698–699.
[39] Kaggle, 2017, "Data Science Bowl 2017 | Kaggle," Kaggle.
[40] Kaggle Inc., 2016, "Meta Kaggle" [Online]. Available: https://www.kaggle.com/kaggle/meta-kaggle.
[41] Hinsz, V. B., Tindale, R. S., and Vollrath, D. a, 1997, "The Emerging Conceptualization of Groups as Information Processors.," Psychol. Bull., **121**(1), pp. 43–64.
[42] Salomon, G., and Globerson, T., 1989, "When Teams Do Not Function the Way They Ought to," Int. J. Educ. Res., **13**(1), pp. 89–99.
[43] McComb, C., Jonathan, C., Kotovsky, K., Cagan, J., and Kotovsky, K., 2017, "Validating a Tool for Predicting Problem-Specific Optimized Team Characteristics," *Volume 7: 29th International Conference on Design Theory and Methodology*, pp. 1–10.
[44] Wright, D. B., 2007, "Calculating Nominal Group Statistics in Collaboration Studies," Behav. Res. Methods, **39**(3), pp. 460–70.

[45] Zheng, H., Li, D., and Hou, W., 2014, "Task Design, Motivation, and Participation in Crowdsourcing Contests," Int. J. Electron. Commer., **15**(4), pp. 57–88.
[46] Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., and Vukovic, M., 2011, "An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.
[47] Paolacci, G., and Chandler, J., 2014, "Inside the Turk: Understanding Mechanical Turk as a Participant Pool," Curr. Dir. Psychol. Sci., **23**(3), pp. 184–188.
[48] Wu, H., Corney, J., and Grant, M., 2015, "An Evaluation Methodology for Crowdsourced Design," Adv. Eng. Informatics, **29**(4), pp. 775–786.
[49] Yumer, M. E., Chaudhuri, S., Hodgins, J. K., and Kara, L. B., 2015, "Semantic Shape Editing Using Deformation Handles," ACM Trans. Graph., **34**(4), p. 86:1-86:12.
[50] Burnap, A., Ren, Y., Gerth, R., Papazoglou, G., Gonzalez, R., and Papalambros, P. Y., 2015, "When Crowdsourcing Fails: A Study of Expertise on Crowdsourced Design Evaluation," J. Mech. Des., **137**(3), p. 31101.