Analyzing Participant Behaviors in Design Crowdsourcing Contests Using Causal Inference on Field Data

Ashish M. Chaudhari
School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907

Zhenghui Sha
Department of Mechanical Engineering,
University of Arkansas,
Fayetteville, AR 72701

Jitesh H. Panchal
School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907

Crowdsourcing is the practice of getting ideas and solving problems using a large number of people on the Internet. It is gaining popularity for activities in the engineering design process ranging from concept generation to design evaluation. The outcomes of crowdsourcing contests depend on the decisions and actions of participants, which in turn depend on the nature of the problem and the contest. For effective use of crowdsourcing within engineering design, it is necessary to understand how the outcomes of crowdsourcing contests are affected by sponsor-related, contest-related, problem-related, and individual-related factors. To address this need, we employ existing game-theoretic models, empirical studies, and field data in a synergistic way using the theory of causal inference. The results suggest that participants' decisions to participate are negatively influenced by higher task complexity and lower reputation of sponsors. However, they are positively influenced by the number of prizes and higher allocation to prizes at higher levels. That is, an amount of money on any following prize generates higher participation than the same amount of money on the first prize. The contributions of the paper are: (a) a causal graph that encodes relationships among factors affecting crowdsourcing contests, derived from game-theoretic models and empirical studies, and (b) a quantification of the causal effects of these factors on the outcomes of GrabCAD, Cambridge, MA contests. The implications of these results on the design of future design crowdsourcing contests are discussed. [DOI: 10.1115/1.4040166]

1 Introduction

During the past decade, crowdsourcing has been widely used as a way to facilitate open innovation where ideas from outside an organization are sought to supplement internal activities [1]. In crowdsourcing, a problem is outsourced to a large group of people, predominantly online communities. In a specific form of crowdsourcing, prizes are announced along with a problem, the crowd submits solutions to the problem, and the best solutions are awarded the prizes. This form of crowdsourcing is called a *crowdsourcing contest*.

There are many decisions to be made in setting up crowdsourcing contests [2], for example, selecting among (i) single-stage versus multistage tournament, (ii) open entry versus restricted entry versus entry fee, (iii) a grand competition for the entire system versus multiple smaller competitions for subsystems, (iv) fixed prize versus auctions, (v) winner-takes-all versus multiple-prizes, (vi) alternate prize amounts and the duration of the contest. These decisions affect the outcomes of contests, which can be measured in terms of quality of solutions, number of contributors, amount of effort invested by participants, overall cost of running the contest. For example, higher prize amount attracts greater participation, and participation by a large number of individuals indicates a greater probability that someone in the participant pool will be able to solve the problem.

Overall outcomes of contests are driven by individuals' decisions such as whether to participate or not, and the amount of effort and cost to invest in. Individual decisions are, in turn, dependent on decisions made by other participants, and various factors related to the problem, the design of the contest, the sponsor, and the individual. For instance, individuals are less likely to participate if many other experts are also participating, because the chances of winning are low [3,4]. Individuals are more likely to participate if the problem is simple, or within their area of expertise. Similarly, individuals are more likely to be motivated to participate if the contest is sponsored by reputed organizations. The interdependence between these factors and their effects on the outcomes are shown in Fig. 1.

Understanding the interdependence between these factors is essential for effective design of crowdsourcing contests. Thus, our *objective in this paper* is to understand how the outcomes of crowdsourcing contests are affected by sponsor, contest, problem, and individual-related factors.

1.1 Review of Relevant Research. Research on crowdsourcing in the *engineering design* community has focused on demonstrating the benefits of crowdsourcing in design-related tasks (e.g., idea generation and evaluation), and developing ways to improve the effectiveness of crowdsourcing in design. For example,
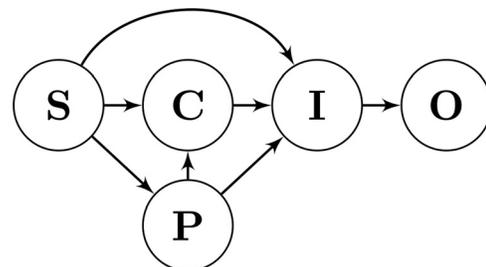

Fig. 1 Interdependence of set of sponsor-related (S), problem-related (P), contest-related (C), individual-related (I), and outcome-related (O) factors in crowdsourcing contests

---

Burnap et al. [5] present a method to identify experts in a crowd for design evaluation tasks. Green et al. [6] use crowdsourcing to evaluate creativity of design using university students as subjects. Their results indicate that it is possible to evaluate originality reliably using a crowd of nonexperts. Gerth et al. [7] present a hypothesis that task factors and individual expertise affect the quality of solution. The hypothesis is tested using a multi-agent simulation for data generation and machine learning techniques for data analysis. Kudrowitz and Wallace [8] crowdsource a task of evaluating a large set of ideas on Amazon Mechanical Turk [9], where the ideas are evaluated based on metrics such as creativity, novelty, clarity, and usefulness. While these studies demonstrate the usefulness of crowdsourcing for engineering design tasks, they do not analyze the effects of different factors (including competition) affecting the outcomes of crowdsourcing.

Contests have been extensively studied in economic sciences using *game theory* [3,4,10] to analyze the effects of contest factors (e.g., prize, number of prizes, and entry fee) on participants' decisions. Considering individual participants as the actors in contests, game-theoretic models theoretically evaluate participants' decisions and their dependence on contest-related factors. For example, models by Taylor [4], Che and Gale [3], and Szymanski [11] are used to analyze the effects of different incentive structures on individuals' decisions, and the resulting outcome quality in innovation contests. Although game theoretic models of innovation contests are helpful in establishing the influence of contest-related factors on the outcomes, and can be adapted for engineering design, they are typically based on a number of assumptions about the problem type, information available to the individuals, and simplifications about the nature of the design process. As an example, in Ref. [12], the innovation process is modeled as a random draw from a known distribution. Depending on the assumptions, the predictions from the models can be substantially different (even contradictory, in some cases). Further, due to the complexity of interactions between different factors (as shown in Fig. 1), game-theoretic models typically fix all the factors, except for a small subset. Therefore, it is difficult for contest designers to determine which model is an appropriate representation of their specific situation.

Laboratory experiments and *field studies* are two commonly used methods for empirical research on crowdsourcing to observe participants behaviors. In laboratory experiments, all aspects of experiment setting such as participants' identity and background, task and problem context, and duration are controlled. The experimenter can also control information provided to participants. In contrast, field studies are conducted on online platforms, and offer partial or no control over experiment settings. Participation in field studies is open to all or group of subscribers of respective platforms.

Laboratory experiments have been used to assess the validity of predictions made by game theoretic models, where participants (typically, university students) are asked to solve well defined tasks [12–14]. All other factors are controlled in such a way that the correspondence between assumptions in the game-theoretic models and experiments is maintained. The advantage of using laboratory experiments is that they allow causal inferences by varying one factor at a time. However, the limitation (as with game-theoretic models) is that it is difficult to establish whether idealized settings used in experiments are representative of real engineering design tasks.

At the other end of the spectrum, field data from real crowdsourcing platforms are used to analyze the relationships among factors affecting crowdsourcing, and the resulting outcomes. For example, researchers [15–17] have observed the effects of various factors such as the nature of the problem and the associated uncertainty, participants' expertise, resource availability, and competition on the outcome quality and participation. Li et al. [15] show that the length of problem specification and software complexity (measured in terms of inheritance, coupling, and cohesion) affect the software quality on TopCoder.com platform. These studies,

however, are restricted to either software development problems on the website link[1] [18] or microtasks for problem solving on Amazon Mechanical Turk [16]. Using field data in research has additional methodological challenges including lack of control and lack of observability of information (e.g., effort invested by participants and cost incurred by them), which results in the inability to attribute the outcomes to specific factors. Because of the lack of control, field data, by itself, can only be used for correlation analysis but not for making causal inferences, which is necessary for understanding how outcomes are affected by various factors.

In summary, engineering design literature is focused on identifying ways to utilize crowdsourcing for design-related tasks but lacks the analysis of participants' decisions under competitive settings. On the other hand, literature on game-theoretic models and empirical studies has focused on analyzing decisions in contests but the relevance of assumptions to engineering design situations is not clearly established. Laboratory experiments provide limited insights because of limited representativeness of the tasks and subject pools, whereas analysis of field data is challenged by lack of control. In this paper, we address the lack of understanding about the relevance of game-theoretic models to design crowdsourcing, and investigate how different factors affect participants' decisions in real engineering design crowdsourcing contests.

1.2 Approach. In order to maintain representativeness of real engineering design crowdsourcing contests, we use field data from crowdsourcing contests. To address the challenge of lack of control, our approach is to use the theory of causal inference [19], which helps in addressing the trade-off between control and realism. The theory of causal inference has formal ways of handling confounding effects if the factors are explicitly modeled in the causal graph. Using this theory, we can quantify the impact of controlling different factors even from field data under assumed causal dependencies between such factors. Using this theory as a foundation, our approach in this paper consists of three steps: (i) evaluating the relevance of assumptions in game-theoretic models to engineering design to identify predicted relationships among factors, (ii) synthesizing the predictions from game-theoretic models and existing empirical studies (laboratory and field experiments) into a single causal graph, and (iii) applying field data from GrabCAD challenges [20] to the hypothesized causal graph to quantify the impact of different factors.

The outline of the paper is as follows: First, we contextualize the predictions about the effects of *contest, problem*, and *individual* factors for the engineering design context. The predictions from game-theoretic models are presented in Sec. 2, and the predictions from existing empirical studies are presented in Sec. 3. Using these predictions, we create a causal graph of the relationships between different factors. The causal graph is presented in Sec. 4. Finally, we employ the causal graph for testing a set of predictions on a dataset acquired from GrabCAD's crowdsourcing platform. The results of the analysis are presented in Sec. 5.

2 Predictions of Participants' Behaviors From Game-Theoretic Models

Game theoretic analysis has been extensively performed in the economics and management literature to predict the outcomes of contests and to optimally design them [3,4,13,21]. The models generate theoretical predictions about participants' decisions in crowdsourcing contests. The models are based on the assumption that each participant is a rational decision maker who strives to maximize his/her expected payoff. Since contests involve multiple participants, the models incorporate beliefs about others' decisions, and uncertainty in receiving desired outcomes in their formulation. Each participant best responds to others' decisions,

---

[1]TopCoder.com

leading to Nash equilibrium. Decisions in Nash equilibrium are dependent on contest, problem, and individuals factors.

A game-theoretic model of contests starts with identifying two types of players: the sponsor and the participants. The sponsor defines the problem, the number of prizes (M), payment structure, and prize amounts (Π), entry fee (F), time deadline (T), etc. Based on the details of a contest, each participant decides whether to participate by paying the entry fee. If participant i decides to participate, he/she incurs cost of resources (C_i) for solving the problem, which results in a solution of quality (q_i). Participants submit their solutions to compete for prizes. All players are assumed to be expected payoff maximizers. The sponsor's net expected payoff (π_S) is related to the best solution from those submitted by all participants, and is modeled as

$$\pi_S = \max_i \mathbf{E}[q_i - \Pi] \tag{1}$$

where $\mathbf{E}$ denotes the expectation operator.

On the other hand, participant i's expected payoff (π_i) is influenced by the prize amount (Π), the probability of winning the contest (f_i), and the resource cost (C_i). Entry fee F is absent in most contests, thus, we ignore it in this analysis. A higher prize amount implies higher return on the participant's effort, and higher probability of winning implies higher chances of receiving this return. The cost, however, simulates a sunk investment on the part of the participant, and has negative impact on payoff. Mathematically, for a single prize contest, participant i's expected payoff is then

$$\pi_i = \Pi f_i - C_i \tag{2}$$

where participant i's probability of winning is a function of his/her own solution quality q_i, as well as others' quality q_{-i} (which are not known to i).

Game-theoretic models are formed by using different assumptions about factors such as Π, f_i, C_i, and q_i. These factors can be related to characteristics of the contest, the problem, and the individuals, as highlighted in Table 1. In Sec. 2.1, we present typical assumptions used in such game-theoretic models of contests, and their relationships with the above factors.

2.1 Typical Assumptions in Game-Theoretic Models

2.1.1 Prize Structure. The prize structure can be of different types such as fixed single, fixed multiple, or auction. When a sponsor specifies single-prize amount ($\Pi = $ constant) for the best solution, the contest is referred to as a *fixed single-prize* contest. On the other hand, when participants submit bids (p_i) for prizes along with their quality, the contest is called an *auction-style* contest. In that case, the prize received by participant i is equal to his/her bid, i.e., $\Pi = p_i$. When prize amounts are fixed by the sponsor, and are offered to a fixed number (M) of highest quality submissions, the contest is called a *fixed multiple-prize* contest. In the case of multiple prizes, $\Pi = \{\Pi_1, \Pi_2,\ldots, \Pi_M\}$, participant i's expected payoff is

$$\pi_i = \sum_{m=1}^{M} \Pi_m f_{i,m} - C_i \tag{3}$$

where $f_{i,m}$ is the probability of winning mth prize [22], assuming that a participant can win at most one prize.

Table 1 List of factors modeled in game-theoretic models

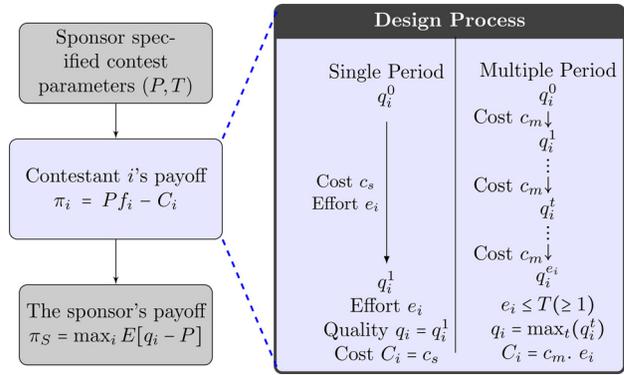| Type of factor	Factors modeled
Contest	Number of prizes, prize amount
Problem	Uncertainty in quality
Individual	Risk preference, cost, effort

Fig. 2 Structure of game-theoretical models

2.1.2 Cost Incurred. The cost C_i in Eq. (2) is a generic variable which includes all resources invested, e.g., monetary resources, human effort, computational cost, etc. It is dependent on how these resources are allocated and expended during the design process. The resources can either be used in a single period (referred to as a *single-period* process), or in a sequential manner (referred to as a *multiperiod* process). The distinction between single-period and multiperiod processes, illustrated in Fig. 2, is dependent on the number of resource usage steps until the final solution is picked for submission by the participant. In both processes, cost (or effort) is a strategic variable, whereas quality is the outcome.

In a single-period process, a decision maker first decides how many resources (money, computational or laboratory experiments) to invest c_s, and then uses all the investment for design. Participant i receives quality outcome of $q_i = q_i^1$ as a result of investing $C_i = c_s$ or effort e_i in one step. A single period process does not imply that a designer is only carrying out a single task. It only refers to how resources are allocated and used.

In a multiperiod process, a decision maker sequentially decides on resource usage after observing the outcome of each design step. Suppose at each step t, participant i incurs cost c_m, and observes the outcome with quality q_i^t. Participant i performs e_i steps sequentially.[2] The solution picked for submission then has the best quality of all outcome qualities obtained during the multiperiod process ($q_i = \max_t(q_i^t)$). The total cost incurred in this process is a function of the number of steps e_i. In its simplest form, if the same cost (c_m) is incurred at each step, then the total cost is $C_i = c_m e_i$.

In engineering design, a period may refer to an iteration where a designer creates a physical/virtual prototype and evaluates it. The assumption of single-period process is typically used when creation of a prototype incurs small cost, and multiple prototypes can be created and evaluated in parallel. However, when each design prototype is expensive, it is necessary to learn the outcome before creating the next prototype. In such scenarios, the assumption of multiple-period process is more applicable [23].

2.1.3 Quality of Outcomes. Game-theoretic models make assumptions about the relationship between resources invested (C_i) and quality received (q_i). This relationship can be either *deterministic* or *stochastic*. In deterministic quality models, a complete control over the quality from a design process is exercised by defining the quality as an explicit function of C_i. In other words, a desired quality q_i can be achieved by investing resources given in terms of cost ($q_i = q(C_i)$). On the other hand, the quality from a design process is stochastically dependent on effort or cost in stochastic quality models. The uncertainty in defining quality may be due to technical uncertainty associated with a design process or uncertain evaluation criteria used by the sponsor. In all these cases, for a given cost C_i, the outcome of the design process

---

[2]In multiperiod processes, the total number of steps is often used as a proxy for the effort invested.
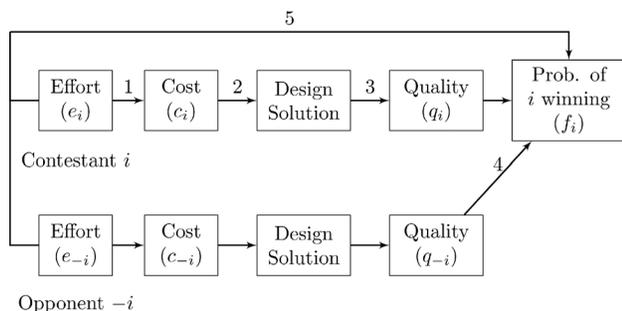
Fig. 3 Various uncertainties in deciding participant i's probability of winning: 1—Uncertain cost, 2—Uncertainty in design process, 3—Uncertain evaluation criteria, 4—Uncertainty in quality of opponent given by CDFs, and 5—Uncertainty in winning given by CSFs

remains uncertain. The quality of an outcome is thus modeled as a stochastic variable $q_i = q(C_i) + \epsilon$ where ϵ is a random variable. The assumption of deterministic quality is applicable to design tasks where the concepts are well understood and the evaluation criteria are well defined on a numerical scale (e.g., topology optimization of a bracket with the objective of maximizing strength-to-weight ratio), whereas stochastic quality is more appropriate in concept design tasks.

2.1.4 Winning Probability. The probability of winning (f_i) in Eq. (2) represents the uncertainty about participant i winning the prize. The uncertainty in winning originates from unknown quality (q_{-i}) of opponents' solution, and/or from the uncertain quality (q_i) of participant i's own solution (see Fig. 3). If the design process or quality evaluation criteria of a contest is unknown, every participant's solution quality is uncertain. Such uncertainties are modeled in game-theoretic models through the probability of winning. Two commonly used models of probability of winning (f_i) are *contest success functions* (CSFs) [24], and *cumulative distribution functions* (CDFs) [4].

Contest success functions are defined as functions of effort invested by a participant, and efforts invested by their opponents. A general form of CSFs for N participants is $f_i(e_i, e_{-i}) = (h(e_i) / \sum_{j=1}^{N} h(e_j))$, where $h()$ represents the effectiveness of any participant i's effort e_i toward winning the contest. When opponents' qualities are unknown, the probability of winning can be modeled based on their effort using CSFs. CSFs predict higher probability of winning for a participant with higher effort. CDFs, on the other hand, are probability distributions defined over quality. If the uncertainty about an opponent's quality is defined by a probability distribution $g(q_{-i})$, then the probability that participant i with quality q_i wins against his/her opponent with quality q_{-i} is given as $G(q_i) = \int g(q_i > q_{-i}) dq_{-i}$. CDFs predict higher probability of winning for higher quality q_i.

Contest success functions model how effective a participant's effort is toward winning, while CDFs model how effective his/her quality is toward winning. Uncertainty in quality from a design process may exist in contests modeled by both CSFs and CDFs.

This uncertainty is defined using a quality distribution in CDFs but is not explicitly modeled in CSFs.

2.2 Instantiation of Models and Corresponding Predictions About Crowdsourcing Contests. In game-theoretic models, participants' decisions to participate and the amount of effort are analyzed based on his/her expected payoff as given in Eq. (2). Once the payoff function is defined, the best response strategy for any participant is to maximize his/her expected payoff in response to the strategies of other participants. All models frame contest as a simultaneous game with perfect information (except for Ref. [25] discussed in Sec. 2.2.1), i.e., participants play the contest concurrently without knowing other participants' strategies (effort, cost, or quality) but they are aware of others' payoff function (cost structure, different uncertainties). Since other participants' strategies are unknown, they are sometimes estimated based on beliefs about others' quality, effort, and cost. Finally, each participant chooses to participate if his/her expected payoff is non-negative, i.e., $\Pi f_i \geq C_i$ [26], and invests effort (or cost) that corresponds to the Nash-equilibrium predicted from a model [3,4].

Different structural assumptions about the prize Π, the probability of winning f_i, and the cost C_i lead to different game-theoretic models. For example, a model by Sheremata [13] considers a fixed single-prize for the winner, models probability of winning as a function of effort using a CSF, and assumes a single-period cost function dependent on effort. Within this model, the overall effort by all participants, instead of the best quality, is defined as the contest outcome. This model is different from Taylor's model [4] of fixed single-prize contest with uncertainty in design process. In this model, the cost is incurred over multiple periods but the winning probability is defined in terms of a CDF of quality $G(q)$.

Examples of existing models from the literature are listed in Table 2. For the purpose of relating structural assumptions to outcomes predicted in game-theoretic models, we classify the models based on their formulation of quality outcome (q_i) and cost (C_i). q_i is modeled either as deterministic or stochastic, while C_i either as single-period or multiperiod function. However, some models do not consider the type of quality as they use CSFs for probability of winning. Because these models can be interpreted as stochastic or deterministic type, we add *undefined quality* as the third type. Also, no model exists for the combination of deterministic quality and multiperiod cost, because the relation between the two is known in deterministic problems, and the resources are therefore always expended in single period. Given such categorization, we present predictions from sample models in different categories.

2.2.1 Deterministic Quality, Single-Period Cost Function. A model by Che and Gale [3] falls under this category. It assumes a single-period design process where q_i is a deterministic function of C_i. The model analyzes equilibrium strategies for two participants, symmetric or asymmetric with respect to the cost function. The outcomes of contests are predicted as mixed-strategy equilibrium meaning that a distribution over equilibrium quality is present, given by a CDF. This model predicts results comparing auction-style contests to all forms of single-prize contests. The model predicts that a first-price auction with two symmetric

Table 2 Categorization of game-theoretic models along model assumptions, and corresponding examples

Quality outcome from design process	Cost function	Example models
Deterministic quality	Single-period	Che and Gale [3][a,s,d], Archak [18][s,m,d]
Deterministic quality	Multiperiod	No models
Stochastic quality	Single-period	Schottner [25][a,s,d], Terweisch and Xu [21][s,m,d]
Stochastic quality	Multiperiod	Taylor [4][s,d], Fullerton et al. [12][a,d]
Undefined quality	Single-period	Szymanski and Valetti [11][s,m,c], Sheremata [13][s,m,c]

Notes: Payment structures used within models are denoted by superscripts: [a]auction; [s]fixed single-prize; [m]fixed multiple-prize. Probability of winning assumptions used within models is denoted by superscripts: [c]CSFs; [d]CDFs.

participants offers higher expected payoff for the sponsor than a fixed single-prize contest. This prediction is labeled $P1$.

In a model of multiple-prize contest of this category, the influence of relative magnitude of each prize is studied by Archak [18]. In this model, the relation between quality and effort is deterministic, participants are asymmetric with different cost functions $C_i(q_i)$, and the probability of winning is modeled using a CDF. The quality outcome from a contest is presented as a symmetric equilibrium where all participants submit the same quality. Based on this equilibrium, the model predicts the following: jth prize's influence on the quality of kth winner is of the order $\mathcal{O}(j^{k-1}/2^{j-1})$. Consequently, every prize positively influences quality, quality of the kth prize is most influenced by the kth prize, and the influence of every added prize on quality outcomes is decreasing. This prediction is labeled prediction $P2$.

2.2.2 Stochastic Quality, Single-Period Cost Function.
Schottner [25] compares auction-style contests to fixed-single-prize contests in the presence of uncertain quality outcomes from a design process. Uncertainty in quality outcome is modeled as a sum of a deterministic part and a random part, ($q_i = f(e_i, C_i) + \epsilon_i$). A distribution over the random part of quality is used to define the probability of the winning function. The model assumes that the quality (q_i) picked for submission by every participant is observable to others before placing an auction bid for price (p_i). Quality outcome from contests is presented as a pure-strategy Nash equilibrium. The model predicts that large differences in qualities of two participants due to a large variation in the random part ϵ_i of quality elicit high bid prices from participants. This results in higher cost of prize for the sponsor. In comparison, the model predicts that a controlled contest with a fixed single-prize may be optimal (prediction $P3$).

Terwiesch and Xu [21] model a single-period design process with the uncertainty in quality due to unclear evaluation criteria, called *uncertainty in evaluation*. They argue that contests (*expertise-based*) with problems of low uncertainty in evaluation have clear quality evaluation metric, and well-behaved solution space. At the other extreme, contests (*ideation*) with problems of high uncertainty in evaluation have undefined specifications resulting in subjective evaluations by the sponsor. The authors suggest that the effect of participant expertise (ability) is dominant in expertise-based contests as compared to ideation contests. Participants with lower expertise are less likely to participate in expertise-based contests. In ideation contests, however, higher uncertainty around the quality outcome and lower requirement of expertise attracts participation, independent of expertise. Ideation contests, therefore, draw more participation compared to expertise-based contests (prediction $P4$).

2.2.3 Stochastic Quality, Multiperiod Cost Function.
Taylor [4] models the uncertainty in quality in a multiple-period process for fixed single prize contests using a quality distribution $G(q_i)$. The probability of winning of any participant is defined using CDF of $G(q_i)$. A participant's quality at each period is randomly drawn from $G(q_i)$. The overall quality outcome is arrived based on a strategy to stop spending resources, called z-stop strategy, according to which the participants invest resources in multiple steps sequentially until the pure-strategy Nash equilibrium quality is achieved. This equilibrium predicts that the number of participants N spending effort is proportional to (Π/c_m). This leads to the hypothesis that the number of participants in a single-prize contest is proportional to the prize amount (Π) (prediction $P5$).

Fullerton et al. [12] extend this model to auction-style contests. They argue that the bid price p_i is a monotonically increasing function of quality q_i. Based on this assumption, the model predicts that the sponsor's expected payoff in auction-style contests is strictly less than his/her expected payoff in fixed-prize contests (prediction $P1$).

2.2.4 Undefined Quality, Single-Period Cost Function.
Szymanski and Valetti [11] present a model where every participant's effort is defined as a contest outcome. The sponsor's payoff is given as an aggregated effort of all participants. In the model, the probability of winning is defined as a function of effort, given by CSFs. Further, all participants are symmetric in their cost which is same as effort ($C_i = e_i$), and all participants are risk-neutral which lead them to follow a pure Nash equilibrium strategy where everyone invests the same amount of effort. This Nash equilibrium effort is expended in one step. Sheremata [13] makes similar structural assumptions, and derives the Nash equilibrium for multiple-prize contests. A comparison of single-prize contests with multiple-prize contests, in these models, suggests that symmetric and risk-neutral participants generate higher effort in fixed single-prize contests than fixed multiple-prize contests (prediction $P6$).

Szymanski and Valetti [11] extend the model to participants with asymmetric costs, and analyze their effort levels with respect to the addition of a second prize. The model predicts that in the case of asymmetric costs assuming risk-neutral participants, multiple-prizes generate more effort than the single-prize contest with the same total prize amount. Additionally, deviation from risk-neutral behavior is more likely to support higher effort from multiple-prizes (prediction $P7$) [27]. The model also predicts that in fixed multiple-prize contests, the equilibrium effort by participants increases with the increase in prize amounts (prediction $P8$).

2.3 Summary of Predictions From the Models.
In summary, the models discussed above predict contest outcomes assuming particular forms of uncertainty in quality and the cost function. Some models neglect uncertainty while others model it explicitly. Most models consider either single-period cost function or multiperiod cost function to establish the relationship between cost and effort. We therefore compare predictions from different game-theoretic models to observe the effects of these factors.

From the comparison of predictions from the models, we observe that the uncertainty in quality affects contest outcomes, wherever the cost structure does not. For example, the desirability of auctions reduces with an increase in uncertainty about quality outcomes, as uncertain quality results in large price bids submitted by participants (predictions $P1$, $P3$). Uncertainty in problem specifications and quality evaluation motivates larger participation in contests, i.e., ideation contests are expected to attract higher participation compared to expertise based contests (prediction $P4$). In contrast, an auction-style contest, in both single and multiperiod cost functions, is expected to provide higher payoff to the sponsor than a fixed single-prize contest (prediction $P1$). Effects of prize amounts and number of prizes do not change with cost structure (predictions $P5$, $P8$). Higher prize amounts, for both cost structures, induce greater participation.

In design crowdsourcing, the outcomes of interest also include the number of participants which improves the diversity of solutions, and the quality of winning solution. The predictions from game-theoretic models about effort and quality of outcomes can be extended to these two factors, as discussed below.

The *quality of the winning solution* is the maximum of quality outcomes from all participants. By assuming that quality of winning is positively correlated with the outcome quality from each participant, and quality is deterministic as formulated in prediction $P2$ by Archark [18], we propose that quality of winning solution increases with prize amounts for single-prize and multiple-prize contests (prediction $P9$).

The *number of participants* is the count of participants who expend nonzero effort. To extend the insights about effort as an outcome to the number of participants as an outcome, the decision to participate is regarded as spending nonzero effort. For crowdsourcing contests, we assume that the number of participants is positively correlated with the effort invested by each participant. Higher effort by all participants would also mean higher number of individuals participating. Additionally, assuming that participants of crowdsourcing contests are asymmetric and risk-averse,

predictions $P7$ and $P8$ can be extended to the number of participants. So, we predict that the number of participants increases with prize amounts for single-prize and multiple-prize contests (prediction $P10$). Further, the number of participants increases with the number of prizes (prediction $P11$).

3 Predictions About Participants' Behaviors From Existing Empirical Studies

3.1 Laboratory Experiments. Laboratory experiments assess the predictive power of game-theoretical models [12–14]. These experiments are observed to support predictions from respective game-theoretic models. For example, the laboratory experiment by Fullerton et al. [12] compares the outcomes of fixed-prize contests and auctions. From this experiment, the prediction $P1$ that auction generates better expected payoff for the sponsor than fixed single-prize contest is found to be true. Another experiment by Sheremata [13] compares participants' effort in single-prize contests to their effort in multiple-prize contests. The results of this experiment show that subjects' behaviors are consistent with the corresponding model's prediction $P6$. Sha et al. [14] observe the effects of effort on quality and winning probability using a laboratory experiment. The results show that quality from a multiple-period process with uncertainty is an increasing log-linear function of effort, which attains the highest quality at high effort. In this process, participants who win have put more effort. The probability of winning therefore is observed to increase with effort.

3.2 Field Studies. Game-theoretic models and laboratory experiments evaluate the effects of contest-related factors on participants' decisions. However, participants of a crowdsourcing contest are also influenced by problem and sponsor-related factors, which are typically analyzed using field studies [28,29]. Examples of problem-related factors are task size, technical difficulty, variability of required skills, uncertainty in solution, feasibility, etc. Examples of sponsor-related factors are sponsor interaction, feedback to participants' questions, trustworthiness, reputation, popularity, etc. We broadly categorize these factors into *task complexity*, *sponsor-type*, and *sponsor interaction* to study their impact on participants' decisions.

The *task complexity* in contests can negatively affect individuals' decisions to participate. Simple problems have minimal information content and decoupled functional parameters, whereas complex problems have more content and coupled functional requirements [30,31]. Complex problems demand more resources and time spent on research as compared to simple problems. Thus, a simple problem is preferred over a complex problem where size and coupling are necessary to be evaluated [32]. In field studies on software development [15] and language processing [33], problems with smaller size and lower complexity are observed to generate better outcomes. Consequently, we propose that higher task complexity decreases participation (prediction $P12$).

Sponsor type is broadly defined as reputation and trustworthiness of the sponsor of a contest. It has a significant impact on participation decisions. Participants are motivated by recognition they can earn from the sponsor [28]. A reputed sponsor by association provides a better visibility to their submissions and improved recognition of their effort. Additionally, participants who are sponsor's "fans" are more likely to participate [34]. Bigger reputation generally implies a larger fan base, and thus possibly larger participation. Furthermore, the field study by Ref. [28] suggests that sponsor's trustworthiness inspires trust in participants while its fraudulent behavior negatively impacts participation. Therefore, we predict that reputation and trustworthiness of the sponsor improves participation (prediction $P13$).

Sponsor interaction is defined as the interaction between sponsor and participants beyond initial problem statement. During a contest, the sponsor may communicate with participants on contest platform through instant messages, comments, or feedback to answer participants' questions. Previous studies on crowdsourcing in electronic commerce [29,34] suggest that sponsor interaction is important to transfer sponsor's ideas to participants so that participants can generate more and creative solutions based on constructive feedback. Their field studies show that participants' solution quality is positively correlated with sponsor interaction. Similar trend is expected in design crowdsourcing where uncertainty in evaluation and quality is present, as pointed in Sec. 2, and more information or details can improve solutions. Therefore, we propose that sponsor interaction increases effort and solution quality by participants (prediction $P14$).

4 An Integrated Causal Model of Influencing Factors

The predictions from the game-theoretic models and empirical studies are summarized in Table 3. Based on the predictions, we develop a causal model [19] considering the factors identified in existing studies game-theoretic models and empirical studies that influence participants' decisions and contest outcomes. The factors included in the causal model are listed in Table 4. The causal model is depicted as an acyclic directed graph in Fig. 4. This causal graph formalizes and encodes the belief about plausible causal relations between different factors. In the rest of the paper, we perform causal inference analysis on this graph using field data to quantify the strengths of different dependencies. The analysis is based on the *assumption* that these encoded relationships are correct.

The approach adopted in this paper is exploratory rather than confirmatory [35]. In an exploratory approach, a model is specified based on prior scientific knowledge, tested on sample data, and components of the model are validated or respecified. In contrast, in a confirmatory approach a well-accepted model is tested on sample data. Our explanations for proposed causal relationships include a review of existing game-theoretic and empirical studies, and also arguments for plausible confounding. A complete validation (or falsification) of our causal model however requires new data from further experimentation. Since the current knowledge about design crowdsourcing lacks such a causal model and the true model is unknown, future experiments can be built upon the proposed causal model.

In the causal model, we assume that the contest-, problem-, and sponsor-related factors are interdependent, and their effects on the contest outcomes may be confounded. Because sponsors design tasks for contests and decide prizes, their characteristics (called sponsor type) may influence contest design. For instance, popularity, reputation, and trustworthiness will likely influence individuals' preferences toward more effort. Sponsors with these types are likely to have received those types due to their ability to provide higher rewards. Conversely, sponsors with lower popularity may not be able to reward as much, or trusted with completing payment. In such cases, the sponsor type will influence the context design. Further, the sponsor types may influence problem factors. The reputation of sponsor could have emerged due to the sponsor's advancements in particular technology, in that case higher reputation may be correlated with higher task complexity. Alternatively, a sponsor may acquire lower reputation due to vague and unclear problem descriptions in contests, in which case the sponsor type will be correlated with the uncertainty in quality and evaluation.

When assessing relations between problem- and contest-related factors, it is expected that task complexity will factor in the design of contest, e.g., higher complexity may demand higher prize amount. Further, the uncertainty in quality is expected to be correlated with the contest design. As we know from Ref. [21], uncertainty in quality (especially due to unclear evaluation) differentiates ideation contests from expertise-based contests. The contest design can be influenced by whether the task involves ideation or well-defined problem solving, as both require different levels of expertise and knowledge in different domains [36].

Table 3 Summary of all predictions from game-theoretic models and empirical studies, and corresponding assumptions

	Predictions	Assumptions
Game-theoretic models	$P1$: Auction instead of fixed single-prize $\xrightarrow{+}$ Expected payoff of sponsor	Deterministic quality, single-period cost, or Stochastic quality, multiperiod cost
	$P2$: ith prize $\xrightarrow{+}$ jth solution quality	Deterministic quality, single-period cost
	$P3$: Fixed single-prize instead of auction $\xrightarrow{+}$ Expected payoff of sponsor	Stochastic & non-private quality, single-period cost
	$P4$: Uncertainty in quality due to evaluation $\xrightarrow{+}$ No. of participants	Stochastic quality, single-period cost
	$P5$: Prize in single-prize contest $\xrightarrow{+}$ No. of participants	Stochastic quality, multiperiod cost
	$P6$: Single-prize instead of multiple-prize contest $\xrightarrow{+}$ Effort	Undefined quality, single-period cost
	$P7$: Single-prize instead of multiple-prize contest $\xrightarrow{-}$ Effort	Undefined quality, single-period cost, risk-averse or asymmetric participants
	$P8$: Prizes in multiple-prize contest $\xrightarrow{+}$ Effort	Undefined quality, single-period cost
Extensions from the models	$P9$: Prizes in single- or multiple-prize contest $\xrightarrow{+}$ Quality of winning soln.	Winner's quality positively correlated to each participant's quality
	$P10$: Prizes in single- or multiple-prize contest $\xrightarrow{+}$ No. of participants	No. of participants is positively correlated to each participant's effort. Asymmetric or risk-averse participants
	$P11$: Number of prizes $\xrightarrow{+}$ No. of participants	
Empirical studies	$P12$: Task complexity $\xrightarrow{+}$ No. of participants	—
	$P13$: Reputation & trustworthiness of sponsor $\xrightarrow{+}$ No. of participants	—
	$P14$: Sponsor interaction $\xrightarrow{+}$ Effort, Quality	—

Table 4 List of factors considered in the proposed causal model

Type	Factors
Sponsor	Type (reputation/trustworthiness), Sponsor interaction
Contest	Number of prizes, Prize amount, ith prize ($i = 1, 2, \dots M$)
Problem	Uncertainty in quality, task complexity
Individual	Belief, Preferences, Cost, Effort, Payoff, Quality
Outcomes	Number of participants, Quality of winning solution

Therefore, we posit a model of causal dependence between all influencing factors which helps to specify confounding relations of these factors to the contest outcomes (see Fig. 4). The causal graph of relations among all factors is assumed to satisfy the rules of graphical models for causal inference [19]. Any node in the graph is directly affected by only its parent nodes, and indirectly affected by its ancestor nodes. All exogenous factors, i.e., factors that do not have any parents, are assumed to be independent, and may be used as control variables.

This causal graph acts as a formal model of our assumptions when testing predictions. A causal relationship between two factors X and Y postulated in the predictions can be mapped to all
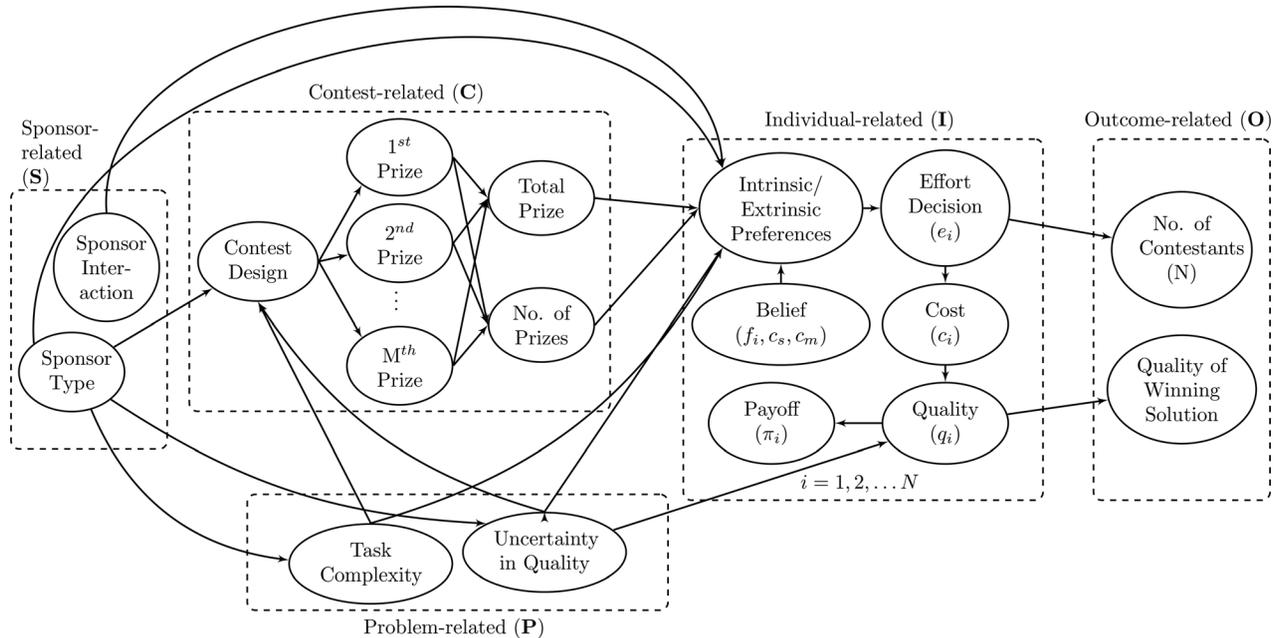

Fig. 4 Graph of hypothesized causal relations between influencing and outcome factors, developed based on existing game-theoretic models and empirical studies on crowdsourcing

connected paths between X and Y in the causal graph. For example, the relationship between total prize ($X = \Pi$) and quality of winning solution ($Y = Q$) is related to paths: (1) ($X = \Pi \rightarrow$ Preferences $\rightarrow e_i \rightarrow C_i \rightarrow q_i \rightarrow Y = Q$), and (2) ($X = \Pi \rightarrow$ Preferences $\rightarrow e_i \leftarrow$ Uncertainty in quality $\rightarrow q_i \rightarrow Y = Q$). In Sec. 5, we discuss whether and to what extent such relationships can be tested on the field data from GrabCAD challenges.

5 Testing Predictions Using Field Data From GrabCAD Challenges

GrabCAD is an open computer-aided design (CAD) sharing community of individuals with diverse specialization [20]. Organizations have sponsored challenges on GrabCAD for interior design, graphic design, industrial design, and product design. A typical GrabCAD challenge starts with posting (i) a detailed description of motivation and the goal behind the problem, (ii) resources such as designs, drawings, or tutorials on required software for participants, (iii) specific solution requirements such as STEP, IGES files or rendering images, and (iv) judging criteria for choosing the winners. In problem description, contest details such as number of prizes, monetary prize amounts, gifts such as iPad, 3D printers, camera etc. with their monetary values, and deadline of final submissions are made public. Participants upload their submissions to the challenge web-page anytime between the start and the end of that challenge. After the final deadline has passed, all submissions are evaluated by a jury which consists of representatives from the sponsor company. The jury rates solutions based on some judging criteria and picks the highest rated solutions. Out of these top submissions, a pre-set number of submissions are awarded prizes.

5.1 Data Collection.
We analyzed all challenges hosted on GrabCAD platform between 2011 and 2016. From this set, we filtered out challenges that involved promotional tasks with no specific problem statement or prizes, e.g., community challenges with no prizes hosted by GrabCAD to promote crowd's involvement. We removed challenges that have crowd submissions set to private. Such challenges are dropped from our analysis because participants' behaviors in them are either unknown, or most likely caused by intrinsic factors which we cannot observe. Finally, we selected data of 96 challenges which provide measurable information on influencing and outcome-related factors. Scrapy [37], a Python-based web crawler, was used to capture the accessible data in accordance with regulations of GrabCAD.

The dataset for each challenge includes a description of the sponsor company, the problem description, the number of monetary prizes and gifts, corresponding prize amounts (worth price for gifts), the number of submissions, a unique identity of each submission and corresponding participant, and the identity of the winners. Based on participant identities and their submissions, the number of participants for each challenge is derived after

removing repetitive and blank submissions so that each participant is counted only once. The number of prizes is counted as the sum of number of monetary prizes and number of gifts.

5.2 Measurable Factors in GrabCAD Challenges.
Some of the factors considered in the causal graph (Fig. 4) are measurable in the GrabCAD dataset while others are not. In this section, we analyze which factors are measurable and how they are measured. For consistency, factors are called *measurable* if they are directly or indirectly inferred from available information, and can be quantified or categorized given some rules. A list of measurable factors in GrabCAD challenges is given in Table 5.

Contest-related factors: Most of the contest-related factors are measurable. In the contest design, fixed single- and multiprize structures are observed on GrabCAD challenges. Auctions are not used for payment in GrabCAD challenges. Number of prizes and prizes are measured.

Individual-related factors: None of the individual-related factors is measurable for participants of GrabCAD challenges. Among the two cost structures assumed in game-theoretic models, single and multiperiod costs, none are observable and therefore not measured for GrabCAD challenges. The resources, both effort and cost, spent by participants are unknown. Although participants' submissions are observed, no consistent measure of quality is available across all challenges. Due to this, participants' beliefs about others' quality and own winning probability are also not measured.

Problem-related factors: Some problem-related factors are inferred from problem descriptions in challenges, and measured based on specific rules described in the following.

(1) *Uncertainty in quality of outcomes*: As discussed in Sec. 2.1.3, this uncertainty arises from two sources: (i) uncertain relationship between resources invested and quality of solution, and (ii) uncertainty in evaluation and problem description, which relates to how a sponsor defines quality measure for submissions to choose the winners. The first source of uncertainty is not measured because cost and quality are not measured, and their relationship varies among participants. The second source of uncertainty is measured from problem descriptions and requirements.

Based on the uncertainty in evaluation, we categorize challenges into *ideation* and *expertise-based* challenges. This categorization follows the guidelines suggested by Terwiesch and Xu [21]. The uncertainty in evaluation, which relates to unclear problem requirements and judging criteria, dictate these guidelines. Problems with high uncertainty in evaluation are categorized as ideation contests, while problems with low uncertainty in evaluation are categorized as expertise-based contests. For example, the Alcoa airplane bracket bearing design challenge [38] had an existing bearing bracket which participants were asked to optimize. Submissions in this challenge were evaluated based

Table 5 List of measurable factors in GrabCAD challenges, relevant statistics of their values, and related predictions. Average (a), median (md), and range ([min, max]) of values of each factor are given.

		Measurable factors	Values	Related predictions
Influencing factors	Contest-related	Number of prizes	[0, 20] a: 4.52, md: 4	P6-7, P11
		First prize (USD)	[0, 6000] a:1576, md:1000	P2, P5, P8-9
		Second prize (USD)	[0, 2000] a: 772, md: 500	P2, P8-10
		Third prize (USD)	[0, 4000] a:448, md:254.5	P2, P8-10
		Fourth prize (USD)	[0, 2000] a: 184, md: 0	P2, P8-10
		Fifth prize (USD)	[0, 2000] a: 138, md: 0	P2, P8-10
		Total prize (USD)	[0, 30000] a:3254, md:2100	P5, P8-10
	Problem-related	Uncertainty in quality (Task type)	Ideation or expertise-based	P4
		Task complexity	Part or system-design	P12
	Sponsor-related	Sponsor type (Web traffic rank)	[321, 2e07] a: 2.94e06, md: 5.5e05	P13
Outcome-related factors		Number of submissions	[2, 638] a: 113.9, md: 77	–
		Number of participants	[0, 300] a: 67.5, md: 53	–

Table 6 Pearson correlation coefficients and corresponding *p*-values (in parentheses) between P, C, and S factors

	Total prize	No. of prizes	Task complexity	Uncertainty in quality
No. of prizes	**0.22 (0.031)**	—		
Task complexity	0.03 (0.805)	−0.12 (0.243)	—	
Uncertainty in quality	−0.02 (0.814)	**0.23 (0.022)**	**−0.33 (0.001)**	—
Web traffic rank	−0.22 (0.549)	−0.18 (0.124)	0.18 (0.120)	−0.20 (0.0844)

on the ratio of ultimate strength and weight. Well-defined design requirements put this challenge in the category of low uncertainty in evaluation. Also, well-behaved solution landscape implies low technical uncertainty. Thus, it is an expertise-based challenge. On the other hand, in the Robot Gripper Test Object challenge [39], participants were asked to come up with ideas for gripping mechanisms but the evaluation criteria were not well defined. Here, any implementation of gripping mechanism can be verified using simulation or prototypes. However, unclear problem requirements result in uncertainty in comparing any two mechanisms. This challenge therefore fits into the category of ideation-based contests.

(2) *Task complexity*: We measure task complexity for Grab-CAD challenges as the elemental count of subcomponents in task, i.e., the number of components required to be designed in challenge. The challenges of single component design are categorized under the *part-design* category, while the challenges that require the design of multiple components and interaction mechanism between these components are categorized under the *system-design* category. The number of components and the existence of complex interactions among them are aimed to represent size and coupling metrics of complexity commonly used for engineering design tasks [30]. For example, consider the autodesk robot gripper arm design challenge [39] where the task was to design a light-weight robotic arm. Since the task here was to design a single component (the arm), it is categorized as a part-design challenge. In contrast, the task in the dirt-bike tire changing tool challenge [40] was to design a tire changing tool. This task involved design of a tool frame, a lever, and gears. It also required the design of mechanism of interaction between these parts and a wheel tire. This challenge is therefore classified as a system-design challenge.

Sponsor-related factors: Sponsor interaction and sponsor type are not directly measurable. But from the information about sponsor's identity available in the dataset, sponsor type, which relates to reputation and trustworthiness, can be indirectly measured. We measure a sponsor's popularity and reputation using the global traffic rank of that sponsor's website as given on Alexa [41]. Low rank refers to high traffic while higher rank to lower traffic. For sponsors with very low web traffic, traffic ranks are not known from Alexa, and a high number (rank of 20 million) is assigned to their traffic ranks. In this measurement, we assume that the web traffic of sponsor's website is an indirect measure of sponsor's reputation. Indeed, according to Refs. [42] and [43], the number of visitors on the website is higher if the online trust in the website's sponsor is higher. Additionally, the online trust is influenced by the sponsor's reputation for reliable behavior, perceived size, trustworthiness, and credibility.

5.3 Testing Hypotheses on GrabCAD Dataset. Given that a subset of the factors presented in the proposed causal model are measurable, only a subset of the predictions can be tested on GrabCAD data. Specifically, predictions *P4*, *P5*, *P10*, *P11*, *P12*, and *P13* may be tested, and are hereafter referred to as hypotheses *H1*, *H2*, *H3*, *H4*, *H5*, and *H6* respectively.

When testing each of the above hypotheses, we make few assumptions. We assume that causal relationships between the factors are defined according to the proposed causal model, where each arrow represents a causal effect of the corresponding starting node on the target node. We do not test alternate structures of the causal model, rather we use this hypothesized causal model to quantify the strength of these dependencies using field data. Next, we assume that each causal relationship is governed by a linear model, i.e., the target node is a linear function of the starting node. This assumption is a natural starting point for analysis when the functional form of the dependencies is not known. Also, we assume that any uncertainty around this linear function affecting the target node is exogenous, and independent of all factors in the proposed causal model. The assumption does not imply that uncertainty in the target node arising from the uncertainty in the starting node is ignored; it is in fact adjusted using the method of linear regression when testing for causal effects, as described in the following.

5.3.1 Assessment of Confounding in Measurable Factors. In this section, we check for confounding between measurable factors in the GrabCAD data before evaluating their effects on participation. We identify confounds and include ones not captured in the causal model to the causal analysis. From a correlation analysis (see Table 6), we observe that significant correlations are consistent with the assumed causal model. An additional association within problem-related factors (between uncertainty in quality *U* and task complexity *T*) not captured in the causal model is found significant. This association is consistent with a view of design complexity as an uncertainty in relating current information content with the amount of information required to satisfy the problem [30].

Shadish et al. [44] offer experimental designs to identify and account for confounds in measured factors. These designs rely on measuring outcomes before treatment (*pretest*), after treatment (*posttest*), and separating groups with and without treatments (*control*). For the GrabCAD data, contest outcomes are only measured after challenges are over and not at different times, i.e., a pretest is lacking. As a result, our study can be viewed as a "posttest-only design with non-equivalent groups." This design is subject to *selection bias* whereby treatment effects may be confounded with population differences since population in control groups is self-selected. For our case, populations are different groups of challenges. The selection bias then can be addressed by identifying differences in challenge characteristics, and whether they are confounded or correlated. Therefore, we evaluate the effects of measurable factors while considering confounding between them. The analysis may still be subject to selection bias due to unobserved external factors, which is a limitation of this study. To evaluate the effects, we employ a statistical approach as discussed in Sec. 5.3.2.

5.3.2 Finding Causal Effects While Adjusting for Confounding Factors. In theory, the causal effect of factor X on factor Y is evaluated by fixing $X = x$ denoted by operation $do(X)$, and calculating the probability $\Pr(Y = y | do(X = x))$, i.e., the probability that $Y = y$ when we intervene to make $X = x$ [19]. When evaluating this effect, we say that factor Z has confounding effect if Z coincides with the parents of X, and directly affects Y or has another path to Y not passing through X. Multiple factors may be identified as confounding factors for a pair of X and Y. The confounding effect of a set of Z is neutralized using the adjustment formula

$$\Pr(Y = y | do(X = x)) = \sum_Z P(Y = y | X = x, Z = z) P(Z = z) \quad (4)$$

In our case, these calculations are simplified due to the assumption of linear causal effects. When X, Z, and Y are interrelated by

linear relationships, the adjustment formula converts to a linear regression formula, given as

$$Y \sim \alpha_1 X + \alpha_2 Z + U_Y \qquad (5)$$

where α_1 and α_2 are the causal effects of X and Z, respectively, on Y, and U_Y is the uncertainty in Y arising from exogenous factors not related to values of X or Z. By including Z in the linear regression model of Y on X, any effect that Z has on X is incorporated in the model, and that effect is thereby adjusted. Here, the causal effect α_1 is a direct effect if the connecting path(s) X to Y has no intermediate factors, i.e., X has no child other than Y. Similarly, the causal effect α_2 is a direct effect if Z also has a direct link to Y, in addition to the direct link to X. Otherwise, these effects are indirect effects.

As an example, when testing the effect of total prize (X) on the number of participants (Y), the confounding factors (sponsor type, task complexity, and uncertainty in quality), which affect both total prize and number of participants, need to be adjusted. Since all directed paths from these confounding factors to total prize reach through the set of individual prizes ($\Pi_1 \ldots \Pi_M$), which are parents of total prize, all individual prizes comprise the set Z. If the effect of the parent node set Z on total prize is adjusted, the effects of all ancestor nodes are adjusted as a result. One way to adjust the effects of the set Z is to keep values of all its factors constant. Then, a linear regression between number of participants and total prize would quantify the corresponding linear effect. However, when the values of the Z factors are varying as is the case for the GrabCAD data, the implementation of linear regression according to Eq. (5) is a better approximation of the linear effect of total prize on the number of participants.

5.3.3 Results of Hypothesis Testing.
According to the procedure explained above, factors X, Y, and a set of factors Z are identified for each hypothesis to be tested, and a linear regression is carried out using the ordinary least squares method. The results obtained from this method are listed in Table 7. These estimated parameters are indirect effects as paths from X to Y include other factors. This table reports indirect effects of X and Z on Y but does not report direct effects of Z on X. Since the actual models of causal relationships are unknown and can be non-linear, the results may possess high variability (low R^2). However, because we are interested in quantifying the significance of X's effect on Y, we emphasize on estimates of p-values.

We observe that the effects of X on Y in multiple regressions for hypotheses $H1$ and $H2$ are not statistically significant. This may be due to the disparity in sample sizes and small sample size. For hypothesis $H1$, 59 challenges in the GrabCAD dataset are of expertise-type while remaining 27 are of ideation-type (Cohen-d effect size without adjusting Web-traffic = 0.37, prob. of Type II error in multiple regression model = 0.37). We are also not able to test the effect of Web traffic rank (Z) on uncertainty in quality (X) as posited in the proposed casual model, because the linear regression of Z on X ($Z \sim \alpha X + U_z$) provides statistically insignificant results. To test hypothesis $H2$, 15 single-prize challenges are available from the data. The data from these 15 challenges do not provide statistically significant results, because the sample size is insufficient, and the probability of Type II error (= 0.45) in the multiple regression model is large. Also, the effects of the corresponding Z factors on prize (X) of these 15 single-prize challenges are insignificant.

We observe that the effects of X on Y in multiple regressions for hypotheses $H3$, $H4$, $H5$, and $H6$ are statistically significant, and the null hypotheses can be rejected. The probability of Type

Table 7 Regression coefficients and corresponding *p*-values (in parentheses) of the causal effects of factor *X* on factor *Y* and adjusted factors *Z*. Statistically significant results of interest are in bold.

Hypothesis (Prediction)	Y	X		Z		
H1(P4)	No. of participants	Uncertainty in quality −10.58(0.415)	Web traffic rank −0.24 (0.074)	Task complexity −57.18(0.000)	—	—	—
H2(P5)	No. of participants	Prize in single-prize contest 0.022(0.320)	Uncertainty in quality −43.29(0.531)	Task complexity −2.18(0.975)	Web traffic rank –0.20(0.682)	—	—
H3(P10)	No. of participants	First prize **0.011 (0.002)**	Uncertainty in quality −8.84(0.469)	Task complexity −62.71(0.000)	Web traffic rank −0.17(0.201)	—	—
	No. of Participants	Second prize **0.021 (0.000)**	Uncertainty in quality −9.87(0.404)	Task complexity −62.64(0.000)	Web traffic rank −0.15(0.237)	—	—
	No. of participants	Third prize **0.029 (0.000)**	Uncertainty in quality −7.67(0.523)	Task complexity −62.9(0.000)	Web traffic rank −0.16(0.213)	—	—
	No. of participants	Fourth prize **0.044 (0.002)**	Uncertainty in quality −7.21(0.555)	Task complexity −59.46(0.000)	Web traffic rank −0.16(0.207)	—	—
	No. of participants	Fifth prize **0.045 (0.001)**	Uncertainty in quality −6.90(0.571)	Task complexity −58.32(0.000)	Web traffic rank −0.16(0.206)	—	—
	No. of participants	Total prize **0.022 (0.008)**	First prize −0.021(0.036)	Second prize 0.0043(0.829)	Third prize −0.064(0.030)	Fourth prize 0.02(0.695)	Fifth prize −0.028(0.625)
H4(P11)	No. of participants	No. of prizes **4.89 (0.003)**	First prize 0.0008(0.904)	Second prize 0.032(0.056)	Third prize −0.041(0.127)	Fourth prize 0.0008(0.987)	Fifth prize −0.043(0.406)
H5(P12)	No. of participants	Task complexity −**57.18 (0.000)**	Web traffic rank −0.24 (0.074)	Uncertainty in quality −10.58(0.415)	—	—	—
H6(P13)	No. of participants	Web traffic rank –**0.33 (0.027)**	—	—	—	—	—

II error in multiple regression models for testing hypotheses $H3$, $H4$, and $H5$ is < 0.01 while it is 0.28 for $H6$. The effects of ith prize and total prize on number of participants are positive and significant at a statistical significance level of 0.05. The causal effect (the regression coefficient) of the ith prize on the number of participants increases as i increases, which suggests that every dollar allocated to the $(i + 1)$th prize has more impact on participation than a dollar allocated to the ith prize. The effects of corresponding adjusted factors (Z) on the ith prize (X) are found to be statistically insignificant. However, the effects of ith prize (Z) on total prize (X) are significant for all i, and accordingly adjusted when finding the effect of total prize (X) on the number of participants (Y).

The number of prizes is observed to have a positive impact (coeff $= 4.89$, p-value $= 0.003$) on the number of participants. In obtaining this result, the effects of all prizes (Z) on number of prizes (X) are adjusted, but they are not tested as these effects are found to be statistically insignificant.

In the results for hypothesis $H5$, task complexity is observed to have a significant negative impact on the number of participants (coeff $= -54.3$, p-value < 0.001). The estimated value of intercept parameter in the linear regression model of binary task complexity variable is 106.7. This suggests that the average number of participants in part-design challenges is 106.7 while it is 52.4 (intercept minus coeff) in system-design challenges.

The web traffic rank of a sponsor is observed to negatively affect the number of participant. This effect (coeff $= -0.33$, p-value $= 0.027$) is moderately significant at level $\alpha = 0.05$. The observed trend suggests that the number of participants decreases by 0.33 with every 100,000 increase in the web traffic rank (for an increase of 10 million in rank, the number of participants decreases by 33). This drop may be small; however, it is an aggregate result for all challenges assuming linear relationships. It is possible that the number of participants is an exponentially decreasing function of web traffic rank. For instance, the Handrail Clamp Assembly Challenge by NASA (Web traffic rank $= 640$, $N = 297$), despite having a similar prize structure and problem type as the throttle pedal design challenge by Microtechnologies (Web traffic rank ≈ 4 million, $N = 87$), observed almost four times higher participation than the latter. To test for linear–log relationship, we fit a regression model for hypothesis $H6$ given by $Y \sim \alpha \log X + U_Y$. This model provides a moderately better fit to the data (coeff $\alpha = -15.2$, p-value $= 0.003$).

5.3.4 Implications on Designing Crowdsourcing Contests for Engineering Design. Through the quantification of the causal effects, the results offer insights into what outcomes to expect of future design crowdsourcing contests on GrabCAD, or contests on other platforms that are similar in characteristics. Contest designers can use the results from the model to answer specific questions such as the following. If a sponsor firm has a fixed budget (say X dollars), how much should it increase the first prize, the second prize, and so on? How many prizes should it award? How many participants can a sponsor firm expect in comparison to another more popular and reputed sponsor firm? By how much the participation could decrease if coupling between multiple design components is introduced in a challenge problem? The estimated coefficients in Table 7, for example, show that increasing the first prize by \$5000 will likely increase the expected participation by 55, but distributing the same amount equally in top five prizes will likely increase the expected participation by 150. Dividing X dollars into ten prizes, rather than putting all money on a single prize, will likely increase the number of participants by about 50. Similarly, the number of participants ($N \approx 53$) for a coupled design problem in contest could be about half of that ($N \approx 107$) for a decoupled design problem in contest. Participation in contests by an unknown company will likely be much lower than a larger more reputed firm.

6 Conclusions

Emerging research on crowdsourcing in the engineering design community has studied feasibility and effectiveness of crowdsourcing for engineering design tasks [5–8]. Following this research interest, we investigate how the design of crowdsourcing contests, where participants are mainly incentivized by monetary rewards, affect their outcomes. Game-theoretic models and empirical studies identify factors that have influence on participants' decisions. We propose a causal model of interdependence among these factors, and their impact on contest outcomes. In the data analysis, we test a subset of the predictions from the proposed causal model on the field data from GrabCAD challenges. Apart from hypotheses $H1$ and $H2$, results of all tested predictions are observed to be statistically significant. These results suggest that participants' decision to participate is positively affected by higher number of prizes (more so at higher prizes than the first prize), more prize amounts, less complexity, and higher reputation and trustworthiness in sponsor. Overall, the favorability of multiple prizes and higher allocation to higher prizes provide a support for game-theoretic assumptions that participants are asymmetric/nonhomogeneous and risk-averse. The results are generalizable to other platforms if the hypothesized relationships among influencing and outcome-related factors, and features of these factors are similar to relationships and features present on GrabCAD.

The use of field data presents unique challenges. Some aspects of field data such as effort, cost, or quality are not observable or measurable. As a result, all predictions of the proposed causal model cannot be tested using the GrabCAD challenges data. For the predictions that are tested to be significant, underlying assumptions of the causal inference methodology that proposed interdependences are true, other external factors are absent, and dependencies are linear may limit generalization.

Broadly speaking, the limitations of the analysis are as follows: (i) causal inference using a single data source does not prove the validity of causal assumptions [45], and (ii) the validity of the results is dependent on the hypothesized causal model being true. It is possible that there are other factors influencing the outcomes that are not currently part of the causal graph. These include individual-specific factors such as their intrinsic motivation for participation, their expertise, the relation of the crowdsourcing task to their day-time job, the tools they have access to, and available time for working on problem. These factors cannot be observed from the field data, but can be incorporated in future work through other research methodologies such as controlled experiments, surveys, and interviews.

In future work, the effects of nonmeasurable factors in a field data can be evaluated using laboratory experiments. Selecting participants with specific backgrounds, testing for specific skills or preferences, and monitoring/managing their design processes are ways to achieve control over contest. This will enable attribution of contest outcomes to specific factors. For instance, auction-related predictions ($P1$, $P3$, $P6$), untestable using the GrabCAD dataset, can be tested by setting up an auction for payment in a controlled contest environment. Insights from laboratory experiments, however, may have limited external validity present in field data.

This paper contributes to the design research literature through its investigation of game-theoretic models and their underlying assumptions to understand participants' decisions. The mapping of assumptions in game-theoretic models to characteristics of engineering tasks provide a framework that can be used to develop new models of crowdsourcing contests for engineering design tasks. This study identifies that existing game-theoretic formulations largely ignore problem- and sponsor-related factors, which are key influencing factors in contests. Another contribution of this paper is the causal graph, which encodes the interdependence of factors derived from existing game-theoretic models and empirical studies. Future studies can leverage this causal graph to drive further experimentation into participants' decisions in design crowdsourcing contests.

References

[1] Chesbrough, H. W., 2006, *Open Innovation: The New Imperative for Creating and Profiting From Technology*, Harvard Business Press, Boston, MA.

[2] Panchal, J. H., 2015, "Using Crowds in Engineering Design Towards a Holistic Framework," International Conference on Engineering Design, Design Society, Milan, Italy, July 27–30, pp. 27–30.

[3] Che, Y.-K., and Gale, I., 2003, "Optimal Design of Research Contests," Am. Econ. Rev., **93**(3), pp. 646–671.

[4] Taylor, C., 1995, "Digging for Golden Carrots: An Analysis of Research Tournaments," Am. Econ. Rev., **85**(4), pp. 872–890.

[5] 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," ASME J. Mech. Des., **137**(3), p. 031101.

[6] Green, M., Seepersad, C. C., and Hölttä-Otto, K., 2014, "Crowd-Sourcing the Evaluation of Creativity in Conceptual Design: A Pilot Study," ASME Paper No. DETC2014-34434.

[7] Gerth, R. J., Burnap, A., and Papalambros, P., 2012, "Crowdsourcing: A Primer and Its Implications for Systems Engineering," Defense Technical Information Center, Warren, MI, Technical Report, No. 23063.

[8] Kudrowitz, B. M., and Wallace, D., 2013, "Assessing the Quality of Ideas From Prolific, Early-Stage Product Ideation," J. Eng. Des., **24**(2), pp. 120–139.

[9] Mturk, 2018, "Human Intelligence Through an API," Amazon Mechanical Turk Inc., Seattle, WA, accessed May 11, 2018, https://www.mturk.com

[10] Corchón, L. C., 2007, "The Theory of Contests: A Survey," Rev. Econ. Des., **11**(2), pp. 69–100.

[11] Szymanski, S., and Valletti, T. M., 2005, "Incentive Effects of Second Prizes," Eur. J. Political Econ., **21**(2), pp. 467–481.

[12] Fullerton, R., Lincster, B., McKee, M., and Slate, S., 2002, "Using Auctions to Reward Tournament Winners: Theory and Experimental Investigation," RAND J. Econ., **33**(1), pp. 62–84.

[13] Sheremata, R. M., 2011, "Contest Design: An Experimental Investigation," Econ. Inquiry, **49**(2), pp. 573–590.

[14] Sha, Z., Kannan, K. N., and Panchal, J. H., 2015, "Behavioral Experimentation and Game Theory in Engineering Systems Design," ASME J. Mech. Des., **137**(5), p. 051405.

[15] Li, K., Xiao, J., Wang, Y., and Wang, Q., 2013, "Analysis of the Key Factors for Software Quality in Crowdsourcing Development: An Empirical Study on Topcoder.com," IEEE 37th Annual Computer Software and Applications Conference (COMPSAC), Kyoto, Japan, July 22–26, pp. 812–817.

[16] Kittur, A., Chi, E. H., and Suh, B., 2008, "Crowdsourcing User Studies With Mechanical Turk," SIGCHI Conference on Human Factors in Computing Systems, (CHI) Florence, Italy, Apr. 5–10, pp. 453–456.

[17] Boudreau, K. J., Lacetera, N., and Lakhani, K. R., 2011, "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis," Manage. Sci., **57**(5), pp. 843–863.

[18] Archak, N., 2010, "Money, Glory and Cheap Talk: Analyzing Strategic Behavior of Contestants in Simultaneous Crowdsourcing Contests on Topcoder.com," 19th International Conference on World Wide Web, Raleigh, NC, Apr. 26–30, pp. 21–30.

[19] Pearl, J., 2009, "Causal Inference in Statistics: An Overview," Stat. Surveys, **3**, pp. 96–146.

[20] GrabCAD, 2018, "Engineering & Design Challenges," GrabCAD Inc., Cambridge, MA, accessed May 11, 2018, https://www.grabcad.com/challenges

[21] Terwiesch, C., and Xu, Y., 2008, "Innovation Contests, Open Innovation, and Multiagent Problem Solving," Manage. Sci., **54**(9), pp. 1529–1543.

[22] Clark, D. J., and Riis, C., 1998, "Competition Over More Than One Prize," Am. Econ. Rev., **88**(1), pp. 276–289.

[23] Loch, C. H., Terwiesch, C., and Thomke, S., 2001, "Parallel and Sequential Testing of Design Alternatives," Manage. Sci., **47**(5), pp. 663–678.

[24] Skaperdas, S., 1996, "Contest Success Functions," Econ. Theory, **7**(2), pp. 283–290.

[25] Schottner, A., 2008, "Fixed-Prize Tournaments Versus First-Price Auctions in Innovation Contests," Econ. Theory, **35**(1), pp. 55–71.

[26] Easley, D., and Ghosh, A., 2015, "Behavioral Mechanism Design: Optimal Crowdsourcing Contracts and Prospect Theory," 16th ACM Conference on Economics and Computation (EC), Portland, OR, June 15–19, pp. 679–696.

[27] Sisak, D., 2009, "Multiple-Prize Contests—The Optimal Allocation of Prizes," J. Econ. Surv., **23**(1), pp. 82–114.

[28] Zheng, H., Li, D., and Hou, W., 2011, "Task Design, Motivation, and Participation in Crowdsourcing Contests," Int. J. Electron. Commer., **15**(4), pp. 57–88.

[29] Zheng, H., Xie, Z., Hou, W., and Li, D., 2014, "Antecedents of Solution Quality in Crowdsourcing: The Sponsor's Perspective," J. Electron. Commer. Res., **15**(3), pp. 212–224.

[30] Summers, J. D., and Shah, J. J., 2010, "Mechanical Engineering Design Complexity Metrics: Size, Coupling, and Solvability," ASME J. Mech. Des., **132**(2), p. 021004.

[31] ElMaraghy, W., ElMaraghy, H., Tomiyama, T., and Monostori, L., 2012, "Complexity in Engineering Design and Manufacturing," CIRP Ann. Manuf. Technol., **61**(2), pp. 793–814.

[32] Suh, N. P., 1990, *The Principles of Design* (Oxford Series on Advanced Manufacturing), Oxford University Press, New York.

[33] Kittur, A., Smus, B., Khamkar, S., and Kraut, R. E., 2011, "Crowdforge: Crowdsourcing Complex Work," 24th Annual ACM Symposium on User Interface Software and Technology (UIST), Santa Barbara, CA, Oct. 16–19, pp. 43–52.

[34] Algesheimer, R., Borle, S., Dholakia, U. M., and Singh, S. S., 2010, "The Impact of Customer Community Participation on Customer Behaviors: An Empirical Investigation," Mark. Sci., **29**(4), pp. 756–769.

[35] MacCallum, R. C., 1995, *Structural Equation Modeling: Concepts, Issues, and Applications* (Model Specification: Procedures, Strategies, and Related Issues), Sage Publications, Thousand Oaks, CA, pp. 16–36.

[36] Dorst, K., 2004, "On the Problem of Design Problems-Problem Solving and Design Expertise," J. Des. Res., **4**(2), pp. 185–196.

[37] Scrapy, 2018, "A Fast and Powerful Scraping and Web Crawling Framework," Scrapinghub, Ballincollig, Ireland, accessed May 11, 2018, https://www.scrapy.org

[38] Alcoa, 2016, "Airplane Bearing Bracket Challenge," GrabCAD.com, Cambridge, MA, accessed May 11, 2018, https://grabcad.com/challenges/airplane-bearing-bracket-challenge

[39] Autodesk, 2013, "Autodesk Robot Gripper Arm Design Challenge," GrabCAD.com, Cambridge, MA, accessed May 11, 2018, https://grabcad.com/challenges/autodesk-robot-gripper-arm-design-challenge

[40] Rabaconda, 2012, "Dirtbike Tire Changing Tool Challenges," GrabCAD.com, Cambridge, MA, accessed May 11, 2018, https://grabcad.com/challenges/dirtbike-tire-changing-tool-challenge

[41] Alexa, 2018, "Alexa Website Traffic Statistics Tool," Alexa Internet Inc., San Francisco, CA, accessed May 11, 2018, https://www.alexa.com/siteinfo

[42] Wolk, A., and Theysohn, S., 2007, "Factors Influencing Website Traffic in the Paid Content Market," J. Mark. Manage., **23**(7–8), pp. 769–796.

[43] Shankar, V., Urban, G. L., and Sultan, F., 2002, "Online Trust: A Stakeholder Perspective, Concepts, Implications, and Future Directions," J. Strategic Inf. Syst., **11**(3–4), pp. 325–344.

[44] Shadish, W., Cook, T. D., and Campbell, D. T., 2002, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston, MA.

[45] Bollen, K. A., and Pearl, J., 2013, *Eight Myths about Causality and Structural Equation Models*, Springer, Dordrecht, The Netherlands, pp. 301–328.