Building Community Knowledge In Online Competitions: Motivation, Practices and Challenges

RUIJIA CHENG, University of Washington, USA
MARK ZACHRY, University of Washington, USA

Knowledge building is a prevalent feature in open online systems, but it is challenging to motivate participants to contribute and to maintain quality in the participants' contributions. Open online competitions, where participants compete for prizes with knowledge artifacts, offer a potential design model for online systems to incentivize community knowledge building activities. However, while there is evidence that participants contribute to public knowledge and share it during competitions, it remains unclear how and why they do so. In this study, through interviews with 14 participants in Kaggle Competitions, we investigate participants' motivation, practices, and challenges when contributing to community knowledge under a competitive structure. We find that competitive mechanisms impact expert and beginner participants very differently in their public knowledge building behaviors. Experts contribute to shared knowledge in order to compete for reputation, while they tend to form their own niches and only share knowledge artifacts that are abstract and not usable by less experienced participants. Beginners are often driven away from contributing to shared knowledge because of their vulnerable social image. We leverage Scardamalia's framework for Knowledge Building Communities to discuss the different challenges and opportunities that competitive design brings to expert and beginner participants. We offer design implications for effectively implementing competitive mechanisms that could benefit both expert and beginner participants in future knowledge building systems.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: open innovation contest; knowledge building community; interview

ACM Reference Format:
Ruijia Cheng and Mark Zachry. 2020. Building Community Knowledge In Online Competitions: Motivation, Practices and Challenges. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 179 (October 2020), 22 pages. https://doi.org/10.1145/3415250

1 INTRODUCTION

Nowadays, it is becoming increasingly common for people to build community knowledge with each other in open online systems [47] such as social Q&A [49], open source projects [50] and online creative communities [52]. In these systems, participants collectively advance knowledge that is publicly accessible to the community through distributed contributions, including sharing the artifacts they have created, exchanging feedback, initiating discussions, etc. Such activities are crucial to both the quantity and quality of user generated content, promoting self-regulated learning among members [20, 23, 25] and growth of the community [32]. Yet, as demonstrated in multiple CSCW and HCI studies, maintaining sustainable, high quality contribution among

Authors' addresses: Ruijia Cheng, rcheng6@uw.edu, University of Washington, Seattle, Washington, USA, 98195; Mark Zachry, zachry@uw.edu, University of Washington, Seattle, Washington, USA, 98195.

2573-0142/2020/10-ART179 $15.00
https://doi.org/10.1145/3415250

community members is challenging [4, 32, 45] due to a variety of reasons such as lack of motivation [45] and social barriers [21, 40].

To address such challenges, a specialized type of knowledge building system — online open innovation contests — has recently emerged. Such contests are designed and conducted by organizations to broadcast difficult challenges to a large crowd and to leverage the collective intelligence of many people to generate new ideas and innovations. These contests have proven to be an effective model for collecting innovative knowledge artifacts [15, 27] such as computer programs [3, 33, 44], designs [29, 53] and data science models [48]. Competitive mechanisms, including tangible rewards (e.g., monetary prizes) granted to the best solutions and gamified reputation systems (e.g., contributor rankings and achievement medals), are common features in these contests. Most interestingly, in addition to knowledge artifacts submitted by individual participants to the contests, researchers have observed lasting public knowledge building activities in such competitive communities, including public sharing of in-progress solutions [48] and open discussions around ideas [29]. These public sharing behaviors seem counterintuitive as they may lower individual competitors' chances of winning, yet they are surprisingly common in these competitive systems.

Therefore, we see open innovation contests as a potential new design model for increasing contribution and community knowledge building in online systems. In particular, we contend that system developers who seek to borrow the design of contests for use in their knowledge building systems should have a deeper understanding of the behaviors that such contests encourage and support. While previous research effort has recognized and described public knowledge building behaviors in online contests, few have investigated why competitors contribute to public knowledge when such contributions may harm their own success. Even less is known about how competitive mechanisms affect knowledge building behaviors among different levels of participants. Our study therefore is guided by the following exploratory research question:

How and why do participants with different levels of domain experience contribute to and consume shared knowledge in online competitions?

To answer this question, we conducted in-depth interviews with 14 users of Kaggle Competitions,[1] the world's largest data science open contest platform, wherein participants compete for best solutions and also build knowledge through public code sharing (Notebooks) and engaging in public discussions [48]. In our extended, semi-structured interviews, we asked about users' motivations and experiences related to activities such as sharing and using directly executable notebooks, contributing questions and ideas to discussions, and finding teammates to collaborate with. We found that participants at the two extreme ends of data science experience, namely, the *experts* and the *beginners*, have been very differently impacted by the competitive mechanisms in their engagement in building and using public knowledge. Interestingly, while Kaggle Competitions has been viewed as a successful and open platform for data science enthusiasts to showcase ideas and learn from each other, we found that such opportunities are not equally accessible to experts and beginners. While competitive mechanisms motivate experts to share knowledge, they also lead them to form niches and share solutions that are not readily understood by beginners, thereby impeding beginners' contribution and learning. We conclude with a discussion that leverages the theoretical framework of knowledge building communities [47], describing the potential opportunities and trade-offs of introducing competitive mechanisms to open knowledge building systems. We offer design implications for how developers might integrate competitive elements into systems designed for a community to facilitate public contribution.

[1] https://www.kaggle.com/competitions

2 RELATED WORK

Our study builds on previous research on knowledge building communities, as well as prior studies on gamification and online open competitions.

2.1 Knowledge Building Communities: Benefits and Challenges

Knowledge building, stemming from learning science theories, is "the creation, testing, and improvement of concept artifacts" shared publicly in a community [47]. The building of shared knowledge has been studied as a prevalent phenomenon in open online communities, such as collaborative problem-solving in social Q&A sites [49], co-creation of artifacts or wikis in open source projects [20, 50] and feedback exchanges in creative communities [40, 52].

Knowledge building is considered to be critically beneficial to the health and growth of online communities, because it results in the expansion of the body of shared content, thus attracting new individual members [32, 51]. Scardamalia et al. introduces a framework of principles that mark successful knowledge building in a community: authenticity of problem, improvable ideas, idea diversity, abstraction, epistemic agency, collective responsibility, democratization of knowledge, symmetric advancement of knowledge, pervasive knowledge building, constructive use of authoritative sources, knowledge-building discourse, and transformative assessment [47]. These principles are realized in a variety of knowledge building online communities: in online communities focused on creative content generation, members add their creations to a collection shared by the community, thus offering others potential inspiration [55] and the opportunity to build on their work [17]. Community discussions about and feedback on such creations benefit the creators themselves in terms of giving them encouragement [12, 20] and diverse helpful/critical perspectives [40, 54]. Additionally, such shared artifact-focused discussions benefit the community by interactively deepening their collective understanding of the creations [31]. Studies on social Q&A also show that online discussions can lead to collaborative problem-solving [37, 49], which generates new knowledge and promotes critical-thinking and innovation within the crowd [49].

However, despite these benefits, maintaining regular, widespread public contribution is challenging because of the large scale and open-ended nature of online communities. Public knowledge building is based on community members' willingness to publicly present their creations and offer intellectual input to others [50]. In most online communities, contributions are mainly made by a small group of individuals, leaving most other members as only occasional contributors or sometimes merely "free-riders" [45]. Such low rates of contribution are a loss for the dynamic and diverse knowledge in the community. One reason that contribution rates are low is that it is difficult to constantly motivate participants to invest effort to build in addition to consume public goods [6]. Furthermore, contributors often have psychological barriers when considering whether to expose themselves to the whole community [21, 40], especially when they are not that confident about the authenticity and maturity of their contribution [30], even as such work may indeed benefit the community. In addition, knowledge contributed to online discussions is often considered to be low quality [1] and sometimes does not meet the community's expectations in terms of timeliness, investment, and substantiveness [52]. Recognizing these challenges, we thus seek new ways to enhance knowledge building in open online systems.

2.2 Gamification and Open Innovation Contests

To encourage contributions to shared knowledge, many communities, including open innovation contest platforms, adopt gamification mechanisms. Such mechanisms include the use of extrinsic rewards (e.g., medals and rankings) as game elements for people to compete for in a non-gaming context. In some open online knowledge collaboration systems these mechanisms are prevalent (e.g., Stackoverflow) [11, 19]. Extrinsic rewards are designed to facilitate overall engagement and commitment [8, 13], because they address a type of social need for some community members to increase their self-determination and self-efficacy [46]. However, studies show that such extrinsic rewards can sometimes undermine engagement by suppressing community members' self-interest in the task itself [7]. In addition, rewards may result in less helping activities in the community, as members may be concerned about negative impacts on their rankings incurred by helping others [56].

Open innovation contests, as a special type of gamified design [42], have become an increasingly popular model for organizations to collect innovative solutions to open-ended problems from crowds [36]. By adopting traditional gamification features such as medals, user achievement rankings and tangible rewards (e.g., monetary prizes), such contests introduce a competitive dimension to the interactions, wherein only one or a small number of solutions are selected as winners. Such features are implemented with the hope of motivating more solutions to be submitted [43]. Prior studies show that tangible rewards are the main reason that people are drawn to participate in such competitions [2], and that the bigger the prize, the more likely that there will be an increased number of participants [56]. In addition, people participate because of reputation rewards in the form of social attention from other participants, as the result of public rankings [2, 28].

While previous work has shown that crowd workers on non-competitive platforms would share advice and work opportunities in their communities to provide support to each other [24], less is known about how and why public knowledge would be shared on competitive platforms. While gamification elements (e.g., leaderboards) could increase engagement, it remains ambiguous to what extent competitive mechanisms enhance or suppress contribution to community knowledge, especially because public contributions that benefit the community may diminish one's chance of winning rewards.

2.3 Community Knowledge Building in Open Innovation Contests

In open innovation contests, competitors interact with each other and jointly discuss their innovations, but at the same time, try to individually contribute the best solution [10]. A handful of studies have identified community knowledge building activities in open innovation contests. Some argue that competition can provide participants with common ground through which they teach each other domain knowledge related to the subject of competition, engaging those with less experience [44]. Participants ask questions, evaluate ideas, and share experiences and information in public discussions [29] where mutual commenting leads to more diverse solutions [5]. In competitions that allow exchanges of in-progress work, participants are able to revise and improve their own solutions by comparing their ideas with others' [34]. In competitions that allow people to form teams, participants contribute to broad discussions outside their immediate teams because they want to learn from different competitors [10, 22]; the better a team performs, the more they would share with the community [57]. On the other hand, competitive mechanisms may also result in less knowledge building activities, because only the winner will be recognized in the end [38]. Participants may lose interest in learning about unselected solutions in the community and thus place less value on participation in public discussions [14]. Research has also demonstrated that participation in public discussions dramatically decreases after teams become stably formed [41] and that many participants tend to freely take advantage of the ideas and artifacts shared by others without adding their own contributions [28]. A prior study of Kaggle Competitions shows that on average only a small portion of users share in-progress solutions in competitions, mostly when they are in an adverse situation such as not having enough time or teammates [48].

Despite all these ongoing conversations about community knowledge building activities in open innovation contests, little is known about the competitors' motivations for contributing to public knowledge. It is unclear whether and how they balance competing and contributing, and whether participants across experience levels work with the same motivations. We thus investigated these open questions through in-depth interviews of active community participants. By understanding their reasons for contributing, as well as the challenges they encounter in this process, we hope to gain insights that may help guide the implementation of competitive designs to support shared knowledge building.

3 EMPIRICAL SETTING: KAGGLE COMPETITIONS

In this paper we focus on Kaggle Competitions,[2] an important section of Kaggle.com. Kaggle is the world's largest data science online community, with 128,929 registered users from 194 countries at the time of this study. Kaggle has hosted 370 Competitions, sponsored by external organizations and companies seeking crowd-sourced solutions to real world data science challenges. Kaggle Competitions cover a variety of domains such as medical informatics, business intelligence, urban planning, etc., with a focus on prediction tasks, asking participants to compete for prediction accuracy.

We chose Kaggle Competitions as our empirical setting because, apart from its well-established competitive mechanisms (explained in 3.1), it affords community knowledge building activities such as public code sharing and social Q&A-based discussions. Public code sharing through Notebooks allows participants to directly share, replicate and build on each others' solutions [48]. It is a unique feature in Kaggle Competitions that is not on other open contest platforms such as TopCoder[3] and OpenIdeo.[4] In addition, Kaggle Competitions attract users with diverse background and experience levels (externalized by its user ranking system, explained in 3.1.2), allowing us to investigate the effect of competitive mechanisms on both expert and beginner participants. Finally, the diversity of topics covered by Kaggle Competitions also enables us to generalize our findings to different contest domains.

While in addition to Kaggle Competitions, there are many other features in the Kaggle eco-system, including user uploaded datasets, social news feed and online courses, in this paper we focused only on Competitions and related knowledge building features — Notebooks and Discussion.

3.1 Competitive Mechanisms

3.1.1 Prizes and Medals. At the time of this study, 307 out of 370 Kaggle Competitions offered tangible rewards (271 with money, 22 with swag, 14 with jobs) as prizes. In each competition, participants can submit their solutions multiple times as individuals or in self-formed teams. After submission, participants immediately receive a score for their prediction and a rank of their solutions among all the others in the same competition, as shown in Figure 1. In most cases, only the top three ranked submissions are awarded prizes. Besides monetary rewards, Kaggle Competitions also features awarded medals (gold, silver and bronze) based on performance in a given competition. The specific rules on how a medal will be granted can be found in Kaggle documentation on its user progression system. [5] Medals show up on a user's profile page as an indicator of user achievement status, and are also counted towards the global user ranking system.

3.1.2 User Ranking. All participants in Kaggle Competitions are publicly ranked according to their cumulative performance, presented on its global leaderboard. [6] The user ranking system consists of

[2]https://www.kaggle.com/competitions
[3]https://www.topcoder.com/challenges
[4]https://www.openideo.com/
[5]https://www.kaggle.com/progression
[6]https://www.kaggle.com/rankings

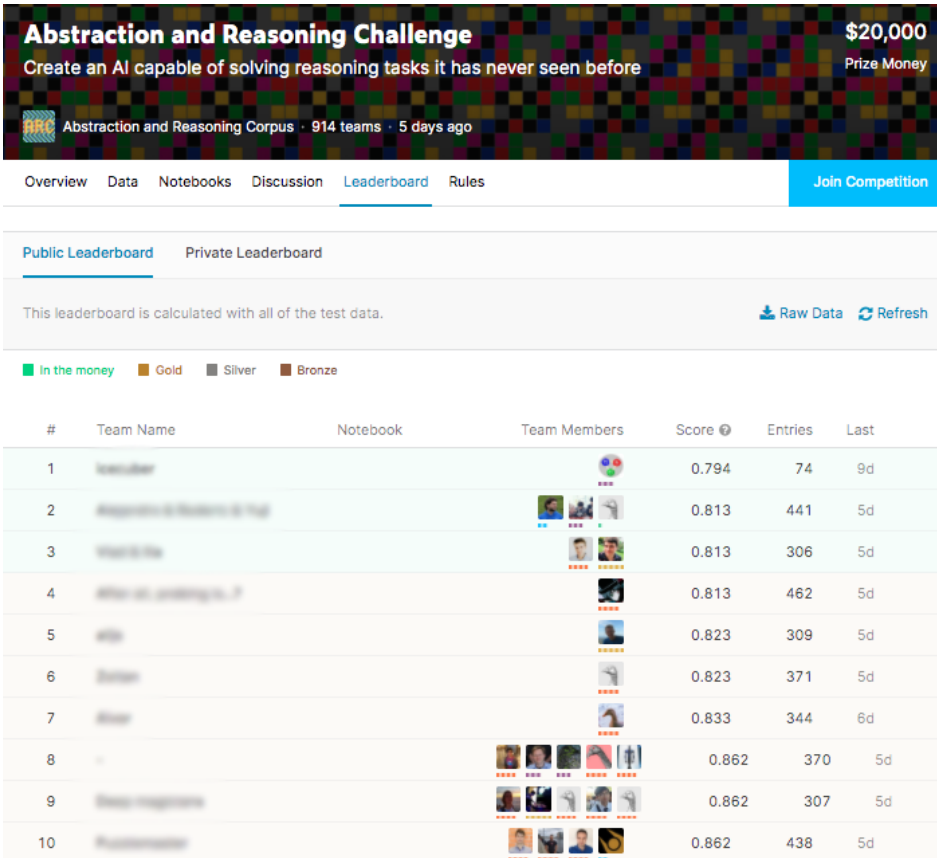

Fig. 1. A snapshot of a leaderboard in a completed competition. The winners and medal receivers, along with the names of all team/individual participants, the scores of their submissions and their ranks in the competition are displayed publicly. Only top 10 rankers are included in this snapshot.

five ranks (from the lowest to the highest): Novice, Contributor, Expert, Master and Grandmasters. The Novices rank is granted to users when they register. A user achieves the Contributor rank when they have their first submission to a competition. The ranks of Expert, Master and Grandmaster are based on the number of total medals a user gained.[5] Higher ranks in Competitions are more difficult to achieve. At the time of this study, there were 5,153 out of 128,929 (top 4.0%) users that achieved the Experts levels in Competitions, 1367 (top 1.1%) achieved Master level and only 171 achieved (top 0.13%) the Grandmaster level.

3.2 Community Knowledge Building Affordances

There are two major community knowledge building affordances in Kaggle Competitions: Notebooks for public code sharing and Discussion for social Q&A and text-based exchanges.

3.2.1 Notebooks. Notebooks[7] is a feature that allows users to share and execute others' code in a Jupyter Notebook environment embedded in and run by Kaggle. Users are allowed to modify

[7]Notebooks were called "Kernels". In the middle of our study, Kaggle changed its name to "Notebooks". In this paper we decided to call it "Notebooks" in order to stay consistent with the current configuration of the system. Some of our

others' shared notebooks and submit them as their own solution to a competition. At the time of this study, there were in total 11,658 notebooks shared on Kaggle, and as shown in a previous study [48], around 10% of users have experience sharing notebooks. We need to note that not all shared notebooks are attached to a specific competition — some notebooks are shared with an educational purpose about general data science methods. Due to the purpose of this study, we specifically asked about usage of notebooks that are connected to competitions during our interviews. Similar to Competitions, the Notebooks section also contains a gamified component in the form of medals and has its own ranking system. Medals are awarded to a notebook based on the level of community approval in the form of upvotes.[8]

3.2.2 Discussion. Competition discussions are the forums attached to specific competitions, where signed-in users can post text-based ideas, questions and solutions as discussion topics and reply to others' topics. In addition to Competition discussions, there is also a general discussion forum that is not connected to competitions. As we study community knowledge building behaviors under a competitive structure, in this study we refer to Discussions as those attached to specific competitions. Users can also earn Discussion medals for their discussion topics and their contributions to someone else's discussion topics. Discussion medals are offered according to the "net votes," which are the sum total of upvotes minus the sum total of downvotes to a topic or comment. Discussion has its own ranking system as well.[8]

4 METHOD

4.1 Participants

We conducted semi-structured interviews with 14 users to investigate their motivation and experiences when consuming and contributing to public knowledge in Kaggle Competitions. We posted our recruitment messages and a screening survey on the general discussion forum on Kaggle, related subreddits (e.g., r/kaggle, r/datascience) and slack channels for Kaggle users. We also sent individual messages to users on the user ranking leaderboard and the first author's social network. We recruited participants who have experience in participating in at least one competition (no matter if they successfully submitted a solution). We also purposefully sampled a pool of users from a variety of geographical regions to make our insights more generalizable. Each participant was compensated $10 for participating in the study.

We deliberately recruited both experienced and less experienced participants, as reflected by their professions and ranks in Competition, Notebooks and Discussion. In our analysis, we classified participants into two categories regarding their experience with data science: *experts* and *beginners*. Our definition of "experts" refers to participants who are self-identified as data science professionals AND have achieved Expert, Master or Grandmaster rank (top 4%, as explained in 3.1.2) in any of Competition, Notebooks or Discussion. We believe the combination of a career in data science and a top global rank is an indicator of advanced expertise in the domain. A total of 5 participants in our study fall into the category of experts. Compared to the world's top achievers, the other 9 participants are thus referred as "beginners." We chose this binary way to classify participants because we hope to see how competitive mechanisms affect experts and those with less experience in the same or different ways. We are aware that within the bracket of "beginners," there could still be nuances in levels of experience and expertise, which we did not consider in this analysis. We will address this point in the limitations section. After interviewing 14 participants, we were able to reach a saturation in the insights that we heard from them. In Table 1 we present the

interview participants still referred to it as "Kernels" in the quotes presented in the following sections. We added a following "(notebook)" to "kernels" in the quotes.

[8]https://www.kaggle.com/competitions

Index	ID	Classification	Gender	Profession	Competition	Notebook	Discussion	Region
1	B1	beginner	M	Student	Contributor	Contributor	Contributor	Europe
2	E1	expert	M	DS professional	Expert	Contributor	Contributor	Europe
3	B2	beginner	M	Student	NA	NA	NA	North America
4	B3	beginner	M	NDS professional	Novice	Novice	Novice	North America
5	E2	expert	M	DS professional	Grandmaster	Contributor	Expert	Asia
6	E3	expert	M	DS professional	Grandmaster	Contributor	Expert	Europe
7	B4	beginner	M	NDS professional	Contributor	Contributor	Contributor	Asia
8	B5	beginner	M	Student	Contributor	Contributor	Contributor	Asia
9	B6	beginner	M	Student	Contributor	Contributor	Contributor	North America
10	B7	beginner	M	DS professional	NA	NA	NA	Asia
11	B8	beginner	F	Student	Novice	Novice	Novice	North America
12	E4	expert	M	DS professional	Contributor	Master	Expert	Asia
13	E5	expert	M	DS professional	Master	Contributor	Grandmaster	Asia
14	B9	beginner	M	NDS professional	Contributor	Contributor	Contributor	Asia

Table 1: Characteristics of study participants. In the columns of Competition, Notebook and Discussion rank, B2 and B7 chose to not disclose their Kaggle profile, so "NA"s are presented. In the column of Profession, "DS professional" refers to data science professional and "NDS professional" refers to professionals who are not working on data science related jobs. The Classification column shows the experience level defined in our study (expert or beginner). Despite our effort to include diverse participants in our study, we were only able to recruit one participant who are self-identified as female, while the rest all self-identified as males.

characteristics of our participants, including their gender, profession, region of residence, ranks in Kaggle Competition, Notebook and Discussion, and classification in our analysis (expert or beginner).

4.2 Procedure

Each interview lasted 30-60 minutes and was conducted via an online conference call. Interview questions focused on participants' experiences with competitions, code sharing, and discussions. To ground the interview, we first asked participants to describe and provide the rationale behind their participation in one competition and then to share any knowledge building around that competition. We then asked them to reflect on their general practice and motivations more broadly when competing and contributing and using others' contributions on the platform.

The first author transcribed the interviews and then followed a thematic analysis procedure [26] to identify common themes across the interviews. As the themes were developed, they were discussed by the research team and checked for fidelity to coded samples selected from the dataset.

5 RESULTS

5.1 Competition Incentivize Knowledge Consumption

Competitiveness is an integral part of the Kaggle experience. Our study participants, however, acknowledged that it is very difficult to actually win prize money in Kaggle Competitions even for experts, let alone for beginners: "*even for Grandmasters, it is still very very hard to win the prize money*," said E2, a Grandmaster level participant. As E2 explained, "*the ratio of [time and effort] investment to [prize money] gain is so low.*" Although almost nobody regards the prize money as the primary incentive for participating in Kaggle Competitions, participants, especially experienced data scientists, still consider competing for prize money, medals and higher rankings to be an important part of the experience. Expert participants regard the prizes as a concrete goal with which they are trying to advance their expertise: "*for some of the competitions which I have a chance of winning, it's about trying to get the best result.*" (E5) Indirect financial benefits, such as new career opportunities, also drive participants to strive for a high ranking:

> "I worked as a consultant and most of my jobs come through Kaggle because people have seen my results in Kaggle. And so they offered me work. So I need[ed] to maintain my ranking." (E3)

The competitive atmosphere in the community, externalized by directly comparable scores and publicly visible rankings, motivates participants to improve their solutions so that they will compare favorably with others building knowledge about the challenge:

> "The way you do competitions is very different from the way you do your [data scientist] job. In competitions you will pay more attention to details, because you get a score after each submission; you get the ranks. In your job there is not a leaderboard, and you will not be directly compar[ed] with others, and you would think maybe you did okay, but actually you did not." (E2)

In general, participants regard the community as a healthy competitive platform and a resource for learning. Expert participants seek out opportunities to win, to build upon knowledge created by others, and to advance that shared knowledge. Beginner participants learn from the contributions of experts and enhance their own skills. In the following sections we elaborate on how experts and beginners leverage community knowledge differently.

5.1.1 Experts Seek for Diversity among Shared Knowledge. During a competition, expert participants tend to read through numerous different notebooks in order to explore different ideas shared by the community. Our study participants described how notebook exploring behavior is especially common in their initial data exploration stage when joining a competition. We found that they like to understand the data thoroughly before diving into an analysis. They believe that building such an understanding helps them choose more suitable methods and models. In particular, they usually search for a variety of notebooks that contain code about data pre-processing and visualization methods:

> "I usually end up with going through the list, and opening a new tab 20 different times. And then I just work with the kernels (notebooks) and write down all of the stuff that looks interesting and then try to incorporate into my own script." (E1)

Expert participants, even those who are Grandmasters in Kaggle Competitions, believe they can learn from reading beginners' notebooks. Because beginners have less experience with competitions, they are less likely to be constrained by conventional ways of looking at the data: "*newcomers are not here to win; they are here to try, so they really tr[y] out new methods*" (E5). Experts regard beginners' notebooks as creative angles for understanding data, which might be used to discover novel approaches in future analysis. Another expert participant, E2, also shared his strategy of taking advantage of ideas shared in the discussions:

> "In my team, I'm usually responsible for scrutinizing every relevant discussion post that I could find, even those by beginners. Sometimes they actually post a question, but I can see the useful insights hidden in it" (E2).

Notably, experts could benefit from community knowledge contributed by beginners. They discover value and opportunities for innovation from beginners' contributions — sometimes not even realized by the beginners themselves. Therefore, beginners could be encouraged to contribute discussion posts or notebooks, leading to the advancement of ideas in the community.

5.1.2 Beginners Feel Empowered Using Experts' Contributions to Get Started. Beginners, on the other hand, primarily use community knowledge generated by experts as an efficient way to enter into a competition. Code for data analysis and machine learning tasks can be very complex, typically including various stages and modules. Beginners take advantage of existing code so that they don't have to begin from scratch and then incorporate the code from those pre-existing notebooks into their own solutions. Beginners thus feel empowered when using notebooks that are highly upvoted and written by experts to get on-board:

> "You don't learn to build a bridge by inventing the bridge all over again. You go through the methods of bridge building from the experts that have come before you and then eventually innovate it in your own right." (B3)

Starting off from experienced competitor's notebooks can also give beginners a sense of how a good solution performs, or in many cases, understand what accuracy looks like compared with their own results: "*primarily I use the [existing] notebooks just to know what is the average good score*" (B7). Borrowing from experts' notebooks also helps beginners start with a relatively high rank in the competition and recognize potential directions that could lead to improved results: "*[high performance] notebooks can help you start with a very good leaderboard score early in the competition. So you just narrowed [the solutions] down*" (B7). Therefore, beginners appreciated notebooks posted by experts that are directly usable as complete solutions. Those solutions as good learning resources are guiding the community towards solving the problems:

> "Kaggle is a competitive platform, but more importantly in my mind, it's an educational platform that the few that competes on a competitive level supports the masses... Whether they win or lose, they're going to have a good kernel (notebook) on their hands that they can post… Someone has that first strong kernel, followed by the second strong kernel, by the third strong kernel and [then] we integrate them." (B3)

While beginners' learning could be scaffolded by experts' solutions, one concern is that using limited notebooks written be a few experts may sometimes suppress creativity in the community. Because notebooks are easy to run and use, beginners may simply take in code that apparently generates good results without knowing how and why it works. This form of uncritical borrowing sometimes means that no further modification or additions to notebooks with good performance are made, as observed by an expert participant:

> "There's sort of group[s] where everybody's copying each other. So all the notebooks are different variations on the same idea after a while… I feel like it's so easy for somebody who's entering a competition just to copy somebody else's code. Not really doing anything innovative or understand[ing] what the code is actually doing." (E3)

From an individual's perspective, a bias introduced by using common solutions in existing notebooks sometimes diminishes their motivation for coming up with novel ways to approach the challenge:

> "If you're starting [by] looking at a lot of other people, you might get biased. And if you see that everyone is using some particular techniques, basically you stop thinking about how you can approach a problem." (B6)

As a result, innovation within the broad community can be diminished, ultimately generating fewer learning resources. Unexpected, creative ways of exploring problems, however, are what competitors, including experts look for in notebooks because often they are what pushes the overall competitive performance to a higher level.

5.2 Experts' Contribution to Community Knowledge: Motivations and Limitations

In this section, we describe our findings on how and why expert participants in Kaggle Competitions engage in activities around community knowledge building:

5.2.1 Experts Build Reputation through Public Sharing. We found that in some cases, experienced data scientists contribute to notebooks and discussions for altruistic reasons. For example, E4 shared that his major motivation to spend extra effort to write clean and organized notebooks is that he would like beginners to benefit from his shared ideas. Nevertheless, we discovered that the major reason experts share knowledge to the community is to boost their reputations as data scientists, both on the platform as well as in the real world.

High reputation can potentially further lead to collaboration and network-building with experienced people in the community. While it is hard to win a competition as an individual, teams formed by experienced data scientists in general have more chances of actually winning the competition:

> "Once you achieve a certain status on Kaggle, it becomes easier to win other competitions because it's easier to form teams with talented people when you have a reputation." (E3)

Notably, in our interviews, we found that only expert participants had the experience of finding online teammates from Kaggle. Beginner participants, in contrast, did not find new people to collaborate with on the platform. The ranking system externalizes and quantifies skill level and expertise in a way that everyone in the community can see. Consequently, high achievers find it relatively easy to identify collaborators for teams, as their rankings serve as clear indicators of what they offer that would be mutually beneficial in a collaborative relationship:

> "When you get higher ranking in this platform, you get to interact with more highly ranked competitor[s], that gives you more opportunities for feedback and learning. So your ranking is your name card–like if I am ranked at top 10, I get to know the top 10 Grandmasters, versus if I were ranked at 1000th, then the top 10 people won't bother to talk to you." (E2)

The expert participants tend to find competitors who are at similarly high ranking, and who can also bring in different domain knowledge and techniques, in order to achieve higher scores. In a community of strangers from all over the world, public contributions are important ways to signal personal characteristics such as skills and domain expertise. E3 shared his experience learning about potential collaborators through their public contributions:

> "People write a lot on the discussion forums and they sort of drop clues about what kind of a solution they have. I just know, for example, there are people who are neural network experienced data scientists, [they say in discussion] 'I'm having trouble because I need more GPUs. So if they say that, then I know that they're using neural nets." (E3)

Notebooks and discussion posts could help experts get attention from other experienced data scientists who might become future collaborators. Therefore they share their contributions to communicate their status and expertise.

In addition to on-site reputation building, contributions to the community sometimes also lead to reputation building outside Kaggle. As Kaggle competitions are well-known as highly competitive and contain difficult data science problems, performance and reputation in Kaggle are linked to recognition in real life. Expert participants therefore contributed to notebooks in order to build up their real life data science portfolio:

> "I knew that I wasn't able to win this competition, but when I just posted a kernel that I already had, I could also improve my Google ranking. Like when people just look for my name, I might come up and they can say, hey, this guy knows about this or that topic." (E1)

Since Kaggle has a sophisticated scoring and evaluation system, a solid performance record on Kaggle is considered by some companies as an indicator of real world data science skills. Putting thought and effort into organizing an exceptional notebook is considered by experts to be an easier task than winning a competition, but one that offers publicly viewable skill sets can have a positive impact on the contributor's career development, leading to indirect financial benefits. Experts, who are more likely to receive good scores than beginners, are thus motivated to publicly share their notebooks.

5.2.2 Experts Primarily Share Abstract Solutions. While some experts share their solutions to competitions in well-explained notebooks and detailed discussion posts, we discovered a pattern from our expert participants that they tend to only share "abstract" solutions. Abstract solutions — usually in the forms of a notebook with snippets of code stripped from a full solution, or a text-based high-level description of methods in discussion posts — can be used to communicate ideas but not replicatable results. One reason that experts only share partial solutions is that in order to prevent ranking inflation that would result from many people simply running a high scored notebook, Kaggle moderation team urges contributors to remove notebooks or discussion posts containing highly scored full solutions posted before a competition ends: "*[Kaggle] let people use kernel (notebook) to communicate ideas, on an inspiration level, but not submittable solutions*" (E2).

Apart from being discouraged by the platform, posting full solutions publicly can also result in complaints from the community and potentially losing reputation. E5, a Master level participant in our interviews, recalled his experience posting detailed documentation of his highly scored solution in a competition discussion. He received down votes and protest comments from competition participants who had some experience themselves, as their rankings were negatively impacted by his sharing:

> "who really care[s] is the middle range [competitors], because what you do affect[s] their results... So these are the people that don't like people [to] reveal their secret[s]." (E5)

As a result, experts avoid sharing their solutions during a competition. Instead, in cases where they do contribute to notebooks and the discussion, they share only snippets of code or partial solutions described in an abstract way. Those incomplete shared ideas are viewed favorably by highly achieved participants, as they often already know the dataset and methods very well before they look at what is shared. Thoughtful participants can understand the problem that the abstract solutions try to tackle and know how to leverage them:

> "The snippets are better because they are focused into one topic… they are very important and informative. I don't want end-to-end solutions because it's kind of a copy paste, which I don't like." (E4)

While abstract solutions might be understandable by experts, beginners may experience such solutions differently: "*for someone who did not fully invest in a competition, they probably won't have a deep understanding from the post.*" (E2) The educational value of incomplete solutions to the general community is therefore discounted — a point we will further discuss in section 5.3.3.

5.2.3 Experts Prioritize Competition over Contribution. While experts may not share full solutions during a competition, what about after a competition is over when sharing full results will not undermine the fairness of a competition? Will the advanced competitors organize their snippets and present them in a cohesive and comprehensive way? The answer is "no" from our investigation.

Making notebooks well-explained and organized is considered time-consuming. We found expert participants believe the scripts that they use for competitions are not clearly structured so that they are not suitable for public consumption. For example, E3 commented on the reason that he rarely posts his solutions as notebooks:

> "I don't share my code. Not because I think it's a bad idea or because I'm competitive, but I'm just embarrassed by my code… It's not very clean" (E3).

Further, some expert participants (e.g., E2), as data science professionals, write and run most of their programs on local machines instead of on notebooks which live on the Kaggle server. Therefore, sharing their full code as notebooks would require extra work to migrate code from their machine to Kaggle on top of all the effort put into competitions, which is not viewed favorably by busy professionals.

Since the cost of organizing code is high, and there are always competitions being launched in the community, expert participants prioritize budgeting their time for diving into the next competition:

> "There are a lot of competitions running concurrently. I mean I would rather spend my time in the next competition, rather than [a] competition that is already passed." (E5)

As abstract shared solutions are understandable by other experts and can potentially bring them reputation, and the next competition can potentially bring them more opportunities for winning, expert participants are not motivated to look back and assemble their partial results into a cohesive solution once a competition is over.

5.3 Beginners' Engagement with Community Knowledge Building: Challenges

5.3.1 Beginners Feel Vulnerable Exposing Their Newbie Status. While expert participants contribute notebooks and discussions that are appreciated by other experts in the community, we found that beginner participants seldom make public contributions. As achievement and experience level are clearly quantified and externalized in the community, beginners live with a newbie's identity — whenever they participate in public activities, their profile will indicate that they are less experienced. Because of this, beginners feel vulnerable when exposing their ideas, opinions or questions to the public, worrying that more experienced members will not take their contributions seriously:

> "If I were to have a question, I don't think I would have been comfortable posting. There's always some anxiety that, you know, everyone else is more experienced at this than you, and they're gonna think my question is stupid sort of thing. I think that was the impression that I got is that the people who were posting and asking questions really knew what they were doing." (B2)

Apart from sharing original knowledge and ideas, beginner participants also reported difficulty in joining in "expert niches" and adding to their knowledge building activities. For example, participants shared that it is hard to add to an ongoing discussion thread because discussions are usually driven by certain exclusive groups of experienced users: "*it's more like a conversation... it's basically quite nested*" (B4); "*They seem to already know each other, [so] there is no point for me to cut in*" (B9). They also perceive the discussion on Kaggle to be "*more formal than Stackoverflow*" (B5) and other programming or statistics help channels because Kaggle discussions include many highly visible, established people from the community. As a result, newcomers feel hesitant to contribute potentially superficial knowledge and indeed seldom post in general, driven away from presenting themselves to and getting input from the community.

5.3.2 Beginners Have Trouble Gaining Visibility. For beginners who do publicly share their contributions, they do so to communicate with other community members and to learn from their advice: "*when I share the notebooks, I want other people to see my code then help me improve it*" (B9). However, as the ranking system gives those high achievers more visibility in the community, notebooks or discussion posts written by beginners seldom gain attention and support:

> "We posted one or two of the good modeling step we made. Didn't really get any attention [be]cause like we were not that good or anything. But definitely I'd be happy if people could take a look at it and have questions about it." (B1)

We also learned that an important criterion for participants to determine whether a notebook is valuable is whether the author is a highly ranked competitor (B4, B5, B7 and B9). Therefore, beginners' solutions rarely receive views or encouragement. As reflected by an expert participant, it is rare to encounter shared solutions written by beginners:

> "there are so many novices who write excellent notebooks… there are so many people who are doing it well, but they're not able to show up (in rankings)." (E4)

Furthermore, because high achievers' notebooks are more visible in the community, beginners believe that only notebooks or discussions resulting in good performance are valuable to the community. Therefore, they feel discouraged from contributing:

> "The thing with notebooks on Kaggle for you to get feedback is that everything has to be properly and sophisticatedly written, like with headings and everything, in order for it to be run higher on the list and to gain visibility and then people actually apply it, which I was maybe lazy enough to not do that." (B6)

One result of the low visibility of beginners' contributions is that it is hard for them to build reputation in the community and thereby to find collaborators. While the ranking system is beneficial for high achievers to showcase their skills and find teammates, it becomes a barrier for lower rankers to find teammates on Kaggle as they cannot build up a portfolio with highly ranked solutions:

> "I just sent quite a few proposals, but people haven't accepted… My profiling is not that good, so people just stick with high ranking people" (B7).

As a result, beginners are separated from experts. Having no chance of working with experts means having no chance to directly learn from the top minds on the platform. Therefore, it is very difficult for someone who is new to data science with the purpose of learning to rapidly develop their ranking in the community.

5.3.3 Beginners Face Difficulty Using Abstract Solutions. Although beginner participants in general appreciate solutions that are publicly shared by experts, they are often confused by notebooks that only contain partial solutions, because they do not have the technical skills and knowledge to figure out all the dependencies and next steps independently:

> "It is common that they have some very unfriendly connections between codes. You have to design the modules yourself, and you need to add your own tool packages… Just sometimes, you can't see their whole pipeline [of code]." (B9)

Given the norm that experts mainly share notebooks that contain only a snippet of code or partial solution, it is very difficult to understand the notebook author's completed approach to the problem. Most experts' sharing thus does not help with either their problem solving or general learning — instead, such sharing creates barriers for beginners to utilize and offer insights to the notebooks.

Beginners' difficulty in comprehending experts' abstract solution may further drive them away from participating in the community. B8 shared a story about her first time participating in a competition: in the beginning she was motivated, so she looked up multiple highly scored notebooks and discussion posts, but could not get any of them to integrate with her own solution:

> "some people only put a snippet, then sometimes I don't exactly understand what they did after that; and when I want to do the same thing, it just doesn't work." (B8)

This experience made her feel unconfident about her ability to solve the problem. As a result, she dropped out from the competition without submitting a solution. While beginners need a directly usable solution to get on-board and feel a bit of empowerment despite their already low public status, failed attempts with abstract solutions can undermine their motivation to get involved in the competition, let alone contribute their own ideas to the community.

6 DISCUSSION

In this paper, we presented our findings from 14 in-depth interviews with both experienced and less experienced participants in Kaggle Competition. We unpacked how participants consume and contribute to community knowledge under a competitive system. Notably, although Kaggle Competitions is viewed as a successful platform that leverages competitive incentives to curate a large amount of independent submissions, our investigation indicates that the current design of Kaggle Competitions does not equally support community knowledge building activities among both experts and beginners. Competitive mechanisms incentivize experts to engage in public knowledge building activities, but present challenges for novices. Experts, who leverage notebook and discussion posts to build reputation in their own niche, tend to share abstract solutions that are hardly usable by beginners and prioritize competing over organizing their contributions. Beginners, on the other hand, face anxiety and difficulties in getting involved, gaining attention and encouragement for their contributions, as well as in comprehending knowledge artifacts contributed by experts. The challenges for beginners to contribute to public knowledge collection echo the fact that despite Kaggle's large user base, only 10% of users have contributed a notebook for public usage [48].

Our findings add to the ongoing discussion about competitive design in knowledge building systems by surfacing the different challenges and opportunities such designs introduce for expert and beginner participants. In this section, following the framework of Knowledge Building Communities [47], we analyze our findings to explore how competitive mechanisms may positively or negatively impact knowledge building activities in an online system. We specifically focused on 4 principles identified in Scardamalia's framework for characterizing a successful knowledge building community: symmetric knowledge advancement, democratization of knowledge, knowledge building discourse, and idea diversity. We specifically chose to discuss these 4 principles because we found these were the particular dimensions of knowledge building communities that could be impacted by the dynamics between experts and beginners based on our findings. We acknowledge that competitive mechanisms could be leveraged to motivate symmetric or mutually beneficial knowledge advancement particularly among experts, but not among beginners or between experts and beginners. We also surfaced resulting expert niches in the community, which could undermine democratization of knowledge and knowledge building discourses among beginners, impacting idea diversity in the community.

6.1 Symmetric Knowledge Advancement Among Experts

Competitive mechanisms enhance symmetric knowledge advancement among experts in the community. Symmetric knowledge advancement refers to equalized knowledge exchange in the community, where members both obtain knowledge from others and produce knowledge that others can use [47]. While it is not hard to understand that experts consume knowledge added by other experts in order to improve their performance in the competitions, our study unpacked why experts in turn contribute to public notebooks and discussions even though such public contributions could potentially undermine their opportunities for winning.

We found in our study that the high prize money in Kaggle Competitions, though extremely hard to win, attracts many experts in the field (who have at least the potential to win) to invest their effort. Only a small number of competitors can win a given competition, but that does not seem to dissuade experts from engaging in the challenges. In general, expert participants do not worry that their public contribution will undermine their already very small chances of winning. Though they recognize they are unlikely to win rewards, because of the highly competitive environment, scoring a high achievement in a well-known data science competition community is a sign of honor. Echoing the literature on gamified design [8, 13], our findings show that experts are motivated to earn medals and higher ranks in order to gain benefits both within the community and beyond. High reputation in the community offers a powerful extrinsic motivation [32] to join in the public contribution, as contribution can lead to reputation boosting [2, 28]. Our findings add that reputation under competitive structure leads to collaboration opportunities with other experts, aiding experts to advance even more in the competitions. The unique benefits of reputation introduced by competitive structure are effective incentives for experts to become both consumers and active contributors to public knowledge, achieving symmetric knowledge advancement.

6.2 Expert Niches Driving Away Beginners

The symmetric knowledge advancement observed among experts, however, does not generalize to the rest of the community. While we found that competitive structure incentivizes experts' knowledge building activities, we also recognize its negative impact on beginners' participation on such activities. This negative impact is primarily because the competitive mechanisms result in the formation and over-representation of expert niches in the community. Beginners are excluded from experts' knowledge because it is not comprehensible and usable for them. Further, in niches of expert participation, beginners feel vulnerable about exposing themselves.

6.2.1 Undemocratized Knowledge. Democratizing knowledge is an important principle in a healthy knowledge building community, according to Scardamalia's framework. The democratization of knowledge in a community requires all participants, regardless of their levels and background, to be able to contribute and consume knowledge. Our study shows that a competitive system design, however, prevents the equalized consumption of knowledge among beginners.

Under a competitive structure, it is easier for experts to form a niche with other experts, resulting in the produced knowledge being consumed only by those within the niche. First, while moderators in online competitions generally support participants and enforce competition rules [18, 39], we found that in order to maintain fairness in the ranking system, they would also intervene in the community's public knowledge sharing. For instance, Kaggle moderators would prevent participants from sharing detailed solutions — only ideational and abstract level sharing is allowed. Such abstract information does not impact experts' performance in the competitions, nor does it impact their advancement in the community, since experts have the background to understand abstract solutions and it is easier for them to find collaboration teams. However, the policy against sharing detailed information creates challenges for beginners who come in to learn and have not developed the skills to comprehend and leverage abstract solutions. It leaves less scaffolding for them to reverse engineer good approaches thus undercutting their opportunity to improve their own rankings.

Second, although experts recognize the value of community learning, because of the competitive nature of the community and the potential benefits of high achievement in a competition (financial benefits and reputation), they tend to prioritize competing over activities that further knowledge building. For example, because it takes extra time and effort to organize their code and solution to make them detailed and comprehensible for the public, it is natural for experts to choose to move forward to another competition instead. This finding is supported by literature where extrinsic rewards can distract contributors from producing public goods, even if they would feel interested in doing so [7].

Their contributions, consequently, are concentrated during the competitions for the purpose of interacting with other experts. Their perceived audience are other experts like them, who do not mind abstract or partial solutions or nested discussion conversations. As experts usually only collaborate with experts and therefore are largely disconnected from beginners, they usually do not fully recognize the beginners' desires to learn from their contributions.

6.2.2 Exclusive Knowledge Building Discourse. The niches of expert knowledge further result in patterns of expert only participation. In Scardamalia's framework, Knowledge Building Discourse is an important pathway to the sharing, refinement, and advancement of community knowledge. According to this principle, participants with all backgrounds and levels should participate in discussion and critique that leads to knowledge advances achieved by the group. While previous research recognizes that competitors participate in and benefit from idea and feedback exchanges [29, 44], we took a closer look at participants of different experience levels. We found that, on the contrary, the competitive mechanisms in Kaggle may undermine such discourse among beginners.

When experts create niches in which they share knowledge that can only be used by other experts, beginners are excluded from effectively leveraging the information and joining the discussion and exchanges, an essential activity for collective knowledge sharing. This disjuncture feeds back to the community's activities, where beginners do not see a lot of other beginners in the discussion. As the ranking system externalizes and quantifies individual competitors' skill levels, it is clear at a glance who is experienced and who is a beginner. Therefore, beginners feel more pressure when they desire to share their solutions and ideas. This echos previous literature that in professional online communities with reputation systems, beginners feel apprehensive about contributing content because they feel stress about being perceived as a "rookie" by the experts [40]. In a competition community, the mechanisms of prize money, medals and public ranking exacerbate the experience level hierarchy, leading beginners to experience more social pressure and reasons to doubt their worthiness in the community.

Another aspect of niches populated by experts is the distortion of collaboration. As indicated in literature, competing in teams leads to better solutions than joining a competition as individuals [9, 57]. However, as experts form teams in their niches, beginners, who are buried in the leaderboard, and who do not actively participate in public knowledge building activities (e.g., discussions), become more distant from those in the expert niches. While the system provides experts with a convenient way to identify similarly ranked potential teammates from whom they are likely to learn new skills, beginners do not have a chance to directly collaborate and work with high ranking contributors. This differentiated experience can result in polarization of levels in the community; experts become more closely collaborative, while beginners struggle because they receive too little direct feedback on their work and lack the courage to ask for it.

6.3 Impact on Idea Diversity

One potential negative impact from the expert niches to the community is less inspiration. Idea Diversity, as stated in Scardamalia's framework, is an important dimension of a good knowledge building community: "idea diversity is essential to the development of knowledge advancement, just as biodiversity is essential to the success of an ecosystem."[47]. Echoing this principle, participants in our interviews expressed a desire for diverse contributions. We learned that expert participants read through a large number of notebooks, especially when first approaching a problem, because they want to be exposed to different ideas. Notebooks from both beginners and experts, notebooks with high scores and low scores—all can potentially be valuable to all levels of participants in the community. Diversity, to some extent, stems from the variety of experience levels among contributors. Beginners are experience-less, so they are not afraid to try non-traditional but innovative methods. Experts can build upon and revise those innovative approaches to further the performance of their solution. However, a lack of beginners' contributions in the community can hinder experts' learning, and in the end further negatively impact the advancement of shared community knowledge. Because the experts cluster in niches, beginners are discouraged from sharing their bold ideas. Experts thus lose out on opportunities to be inspired by beginners, which impacts the overall knowledge advancement in the community.

7 DESIGN IMPLICATIONS

7.1 Using Competitive Mechanisms to Motivate Experts' Contribution

We offer design implications for the developers of new online knowledge building systems. In general, the inclusion of competitive mechanisms will motivate experienced users to contribute to sites designed for knowledge building. For a crowd-sourcing system that lacks experienced input, prizes that are hard to achieve can be used as incentives. Such incentives can attract more experienced contributors, who can potentially contribute valuable knowledge, spurring more general participation among the community. To facilitate knowledge building behaviors such as idea sharing and feedback exchange, reputation-based incentives (e.g., rankings and medals for sharing behaviors) can also be embedded in the design.

7.2 Encouraging Experts to Produce Beginner-Friendly Contribution

On the other hand, designers of open knowledge building systems should also recognize the limitation of expert contributions resulting from competitive designs. As they are motivating experienced participants to provide input, designers should also consider ways to guide them to contribute knowledge that could benefit all levels of users in the system. As detailed sharing may harm the fairness of a competition, future systems could find ways to motivate experts to produce well-organized solutions after competitions are over. For example, a new reputation system could be designed to specifically acknowledge authors whose contributions are favored by beginners. Further, future competition designs should emphasize the completeness and understandability of what is shared, and award such shared content with more tangible credits (e.g., monetary prizes) even after the competition is completed.

Apart from innovating on motivators, designers could also consider including more scaffolding mechanisms that can guide experts to generate complete and comprehensible knowledge artifacts. For example, a feedback system could be implemented into the input interface, highlighting potions of the knowledge artifact that needs more elaboration. Systems could also include rubrics and examples following principles such as cognitive apprenticeship [16], prompting experts to share accessible contributions.

7.3 Removing Barriers for Beginners' Engagement

In addition, designers need to pay more attention to beginners when introducing competitive mechanisms to any knowledge building system. Our study shows that compared to experts, beginners tend to feel apprehensive about contributing their ideas to the community, because ranking systems can externalize their newbie status. Future system design should be able to help beginners overcome such pressures. The design of future systems could explore new ways of presenting reputation status, perhaps through the choice of anonymity, so that beginners share their knowledge artifacts without the concern about their social image in the community. Also, systems could innovate on their reputation system, allowing beginners to gain special credits that can boost their status when contributing to community knowledge.

7.4 Increasing Beginner-Expert Interactions

Last, but not least, we also find that experienced participants and beginners are largely disconnected. To some extent, beginners are invisible to the experts as experts have their own closed connection networks. However, as we found, beginners use experts' contribution as scaffolding for starting to compete, and experts leverage beginners' sharing as a source of novel ideas. We thus suggest that system designers develop mechanisms to connect beginners with high status participants in the community. For example, matching mechanisms could be embedded in the team formation process so that beginners will have a greater chance of working directly with experts; new reputation systems, for instance, a "mentoring badge," could be established to motivate experts to team up with beginners and invest extra effort in guiding them. At the same time, designers could also design for collaborative activities within teams formed by both beginners and experts, building on models of legitimate peripheral participation and situated learning [35]. The teamwork should be designed so that the beginners can participate in experts' problem solving in a way that empowers them with some tasks that they are competent with, but also will not distract from experts' work.

8 LIMITATIONS AND FUTURE WORK

While we believe that our study contributes several empirical insights about knowledge building communities and competitive designs, we admit that our study could be extended in a few ways. In this paper, we mainly discussed how competitive mechanisms might impact two types of participants, the highly achieving (i.e., experts) and the less experienced (i.e., beginners). Due to our limited time and funding, we were not able to recruit a lager set of participants with more diversity in their experience (e.g., intermediate data scientists). We are aware that this dichotomous way of classifying participants may collapse some nuance in both categories — for example, within the bracket of experts or beginners, there might be differences in experience and expertise. Although we were able to discover many common patterns using this simple classification, this is a limitation.

In this study, we chose interviewing as our method because we would like to investigate participants' motivation, practices, and challenges in knowledge building under a competitive system in-depth. Due to the nature of our recruitment strategy, all participants were self-selected for this study, which may threat on the validity of our insights. In addition, while we were able to inductively identify many prominent themes, we did not study the prevalence of these themes, nor do we offer causal or statistical evidence on the relationship between participants' experience levels and their knowledge building activities. Future research could build on this work to carry out a quantitative analysis on the user profile and public contributions (e.g., notebooks, discussion posts) in Kaggle, testing if our qualitative insights hold in a larger scale.

Last, but not least, our study focuses on a single platform, Kaggle Competition. While Kaggle is the most prominent and frequently used online data science contest platform, other smaller examples exist and should be considered in future research. Our work is further constrained in that we closely invested how specific competitive design on Kaggle (prizes, medals and rankings) affect the usage of specific knowledge building affordances (notebook and discussion forum). Future work should explore whether similar patterns exist in many of other competitive systems with different implementations of competitive and knowledge building features.

9 CONCLUSION

In this study, we interviewed 14 participants of Kaggle Competitions, and produced new knowledge on how expert and beginner participants consume and contribute to community knowledge under a competitive system. We found that although experts and beginners appreciate each other's contribution to community knowledge, competitive mechanisms impact experts and beginners' knowledge building behaviors very differently. Experts contribute to shared knowledge in order to build reputation in their own niche, and often produce knowledge artifacts that are not comprehensible or directly usable by beginners. Beginners struggle with vulnerable self-images and low visibility resulting from competitive mechanisms, so they rarely share their ideas and solutions. Based on the framework of Knowledge Building Community, we discussed how competitive mechanisms can enhance symmetric knowledge advancement among experts, while negatively impacting knowledge democratization, knowledge building discourse, and idea diversity when introduced in a community.

Our findings provide implications for effectively implementing competitive mechanisms that could benefit both expert and beginner participants in future knowledge building systems.

ACKNOWLEDGMENTS

We thank Benjamin Mako Hill, Yihan Yu, Spencer Williams, Hanzi Zhang and Yinan Xuan for providing feedback to this study. We thank all the interview participants who participated in our study.

REFERENCES

[1] Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. 2008. Finding High-Quality Content in Social Media. In *Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM '08)*. Association for Computing Machinery, New York, NY, USA, 183–194. https://doi.org/10.1145/1341531.1341557

[2] Maria J Antikainen and Heli K Vaataja. 2010. Rewarding in open innovation communities–how to motivate members. *International Journal of Entrepreneurship and Innovation Management* 11, 4 (2010), 440–456.

[3] Nikolay Archak. 2010. Money, Glory and Cheap Talk: Analyzing Strategic Behavior of Contestants in Simultaneous Crowdsourcing Contests on TopCoder.Com. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. Association for Computing Machinery, New York, NY, USA, 21–30. https://doi.org/10.1145/1772690.1772694

[4] Alexandre Ardichvili. 2008. Learning and knowledge sharing in virtual communities of practice: Motivators, barriers, and enablers. *Advances in developing human resources* 10, 4 (2008), 541–554.

[5] Barry L Bayus. 2013. Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management science* 59, 1 (2013), 226–244.

[6] Yochai Benkler. 2002. Coase's Penguin, or, Linux and" The Nature of the Firm". *Yale law journal* (2002), 369–446.

[7] Pavol Bielik. 2012. Integration and adaptation of motivational factors into software systems. In *Personalized Web-Science, Technologies and Engineering: 11th Spring 2012 PeWe Workshop Modra-Piesok, Slovakia April 1, 2012 Proceedings*. 31–32.

[8] Sanat Kumar Bista, Surya Nepal, Nathalie Colineau, and Cecile Paris. 2012. Using gamification in an online community. In *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*. IEEE, 611–618.

[9] Kevin J Boudreau and Karim R Lakhani. 2015. "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy* 44, 1 (2015), 4–19.

[10] Angelika C Bullinger, Anne-Katrin Neyer, Matthias Rass, and Kathrin M Moeslein. 2010. Community-based innovation contests: Where competition meets cooperation. *Creativity and innovation management* 19, 3 (2010), 290–303.

[11] Inc Bunchball. 2010. Gamification 101: An introduction to the use of game dynamics to influence behavior. *White paper* 9 (2010).

[12] Julie Campbell, Cecilia Aragon, Katie Davis, Sarah Evans, Abigail Evans, and David Randall. 2016. Thousands of Positive Reviews: Distributed Mentoring in Online Fan Communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 691–704. https://doi.org/10.1145/2818048.2819934

[13] Huseyin Cavusoglu, Zhuolun Li, and Ke-Wei Huang. 2015. Can Gamification Motivate Voluntary Contributions? The Case of StackOverflow Q&A Community. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW'15 Companion)*. Association for Computing Machinery, New York, NY, USA, 171–174. https://doi.org/10.1145/2685553.2698999

[14] Shuchi Chawla, Jason D. Hartline, and Balasubramanian Sivan. 2012. Optimal Crowdsourcing Contests. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '12)*. Society for Industrial and Applied Mathematics, USA, 856–868.

[15] Henry William Chesbrough. 2003. *Open innovation: The new imperative for creating and profiting from technology.* Harvard Business Press.

[16] Allan Collins, John Seely Brown, and Ann Holum. 1991. Cognitive apprenticeship: Making thinking visible. *American educator* 15, 3 (1991), 6–11.

[17] Sayamindu Dasgupta, William Hale, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Remixing as a Pathway to Computational Thinking. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1438–1449. https://doi.org/10.1145/2818048.2819984

[18] Cleidson R. B. de Souza, Leticia S. Machado, and Ricardo Rodrigo M. Melo. 2020. On Moderating Software Crowdsourcing Challenges. *Proc. ACM Hum.-Comput. Interact.* 4, GROUP, Article 14 (Jan. 2020), 22 pages. https://doi.org/10.1145/3375194
[19] Sebastian Deterding, Rilla Khaled, Lenard E Nacke, and Dan Dixon. 2011. Gamification: Toward a Definition In CHI 2011 Gamification Workshop Proceedings, Vancouver, BC, Canada. *Online: http://gamification-research. org/wp-content/uploads/2011/04/02-Deterding-Khaled-Nacke-Dixon. pdf [14.4. 2017]* (2011).
[20] Casey Fiesler, Shannon Morrison, R. Benjamin Shapiro, and Amy S. Bruckman. 2017. Growing Their Own: Legitimate Peripheral Participation for Computational Learning in an Online Fandom Community. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1375–1386. https://doi.org/10.1145/2998181.2998210
[21] Eureka Foong, Steven P. Dow, Brian P. Bailey, and Elizabeth M. Gerber. 2017. Online Feedback Exchange: A Framework for Understanding the Socio-Psychological Factors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4454–4467. https://doi.org/10.1145/3025453.3025791
[22] Johann Füller, Katja Hutter, and Rita Faullant. 2011. Why co-creation experience matters? Creative experience and its impact on the quantity and quality of creative contributions. *R&D Management* 41, 3 (2011), 259–273.
[23] Bette Gray. 2004. Informal learning in an online community of practice. *Journal of Distance Education* 19, 1 (2004), 20–35.
[24] Mary L. Gray, Siddharth Suri, Syed Shoaib Ali, and Deepti Kulkarni. 2016. The Crowd is a Collaborative Network. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 134–147. https://doi.org/10.1145/2818048.2819942
[25] Christine Greenhow and Beth Robelia. 2009. Informal learning and identity formation in online social networks. *Learning, Media and Technology* 34, 2 (2009), 119–140. https://doi.org/10.1080/17439880902923580 arXiv:https://doi.org/10.1080/17439880902923580
[26] Greg Guest, Kathleen M MacQueen, and Emily E Namey. 2011. *Applied thematic analysis.* Sage Publications.
[27] John Joseph Horton and Lydia B. Chilton. 2010. The Labor Economics of Paid Crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC '10)*. Association for Computing Machinery, New York, NY, USA, 209–218. https://doi.org/10.1145/1807342.1807376
[28] Bernardo A Huberman, Daniel M Romero, and Fang Wu. 2009. Crowdsourcing, attention and productivity. *Journal of Information Science* 35, 6 (2009), 758–765.
[29] Katja Hutter, Julia Hautz, Johann Füller, Julia Mueller, and Kurt Matzler. 2011. Communitition: The tension between competition and collaboration in community-based design contests. *Creativity and innovation management* 20, 1 (2011), 3–21.
[30] Joy Kim, Maneesh Agrawala, and Michael S. Bernstein. 2017. Mosaic: Designing Online Creative Communities for Sharing Works-in-Progress. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 246–258. https://doi.org/10.1145/2998181.2998195
[31] Yubo Kou and Colin M Gray. 2017. Supporting distributed critique through interpretation and sense-making in an online creative community. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 60.
[32] Robert E Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design.* Mit Press.
[33] Karim R Lakhani, David A Garvin, and Eric Lonstein. 2010. Topcoder (a): Developing software through crowdsourcing. *Harvard Business School General Management Unit Case* 610-032 (2010).
[34] Thomas D. LaToza, Micky Chen, Luxi Jiang, Mengyao Zhao, and André van der Hoek. 2015. Borrowing from the Crowd: A Study of Recombination in Software Design Competitions. In *Proceedings of the 37th International Conference on Software Engineering - Volume 1 (ICSE '15)*. IEEE Press, 551–562.
[35] Jean Lave, Etienne Wenger, et al. 1991. *Situated learning: Legitimate peripheral participation.* Cambridge university press.
[36] Jan Marco Leimeister, Helmut Krcmar, Ulrich Bretschneider, and Winfried Ebner. 2008. Leveraging the wisdom of crowds: Designing an IT-supported ideas competition for an ERP software company. (2008).
[37] Guo Li, Haiyi Zhu, Tun Lu, Xianghua Ding, and Ning Gu. 2015. Is It Good to Be Like Wikipedia? Exploring the Trade-Offs of Introducing Collaborative Editing Model to Q&A Sites. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1080–1091. https://doi.org/10.1145/2675133.2675155
[38] Kai Lu, Wenjun Zhou, and Xuehua Wang. 2014. Social network of the competing crowd. In *2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2014)*. IEEE, 1–7.
[39] Leticia Santos Machado, Ricardo R. M. Melo, and Cleidson R. B. de Souza. 2019. The Role of Platform Moderators in Software Crowdsourcing Projects. In *Proceedings of the 12th International Workshop on Cooperative and Human Aspects*

of Software Engineering (CHASE '19). IEEE Press, 119–122. https://doi.org/10.1109/CHASE.2019.00036

[40] Jennifer Marlow and Laura Dabbish. 2014. From Rookie to All-Star: Professional Development in a Graphic Design Social Networking Site. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 922–933. https://doi.org/10.1145/2531602.2531651

[41] Brian McInnis, Xiaotong Tone Xu, and Steven P Dow. 2018. How Features of a Civic Design Competition Influences the Collective Understanding of a Problem. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 120.

[42] Benny Moldovanu and Aner Sela. 2001. The optimal allocation of prizes in contests. *American Economic Review* 91, 3 (2001), 542–558.

[43] John Morgan and Richard Wang. 2010. Tournaments for ideas. *California management review* 52, 2 (2010), 77–97.

[44] S. Nag, I. Heffan, A. Saenz-Otero, and M. Lydon. 2012. SPHERES Zero Robotics software development: Lessons on crowdsourcing and collaborative competition. In *2012 IEEE Aerospace Conference*. 1–17.

[45] Jakob Nielsen. 2006. Participation inequality: Encouraging more users to contribute. *http://www. useit. com/alertbox/participation_inequality. html* (2006).

[46] Ganit Richter, Daphne R Raban, and Sheizaf Rafaeli. 2015. Studying gamification: the effect of rewards and incentives on motivation. In *Gamification in education and business*. Springer, 21–46.

[47] Marlene Scardamalia. 2002. Collective cognitive responsibility for the advancement of knowledge. *Liberal education in a knowledge society* 97 (2002), 67–98.

[48] Yla Tausczik and Ping Wang. 2017. To Share, or Not to Share?: Community-Level Collaboration in Open Innovation Contests. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 100.

[49] Yla R. Tausczik, Aniket Kittur, and Robert E. Kraut. 2014. Collaborative Problem Solving: A Study of MathOverflow. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 355–367. https://doi.org/10.1145/2531602.2531690

[50] Eric von Hippel and Georg von Krogh. 2003. Exploring the open source software phenomenon: issues for organization science. (2003).

[51] Georg Von Krogh and Eric Von Hippel. 2006. The promise of research on open source software. *Management science* 52, 7 (2006), 975–983.

[52] Anbang Xu and Brian Bailey. 2012. What Do You Think? A Case Study of Benefit, Expectation, and Interaction in a Large Online Critique Community. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 295–304. https://doi.org/10.1145/2145204.2145252

[53] Yang Yang, Pei-Yu Chen, and Paul Pavlou. 2009. Open innovation: An empirical study of online contests. *ICIS 2009 Proceedings* (2009), 13.

[54] Yu-Chun (Grace) Yen, Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. 2016. Social Network, Web Forum, or Task Market? Comparing Different Crowd Genres for Design Feedback Exchange. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. Association for Computing Machinery, New York, NY, USA, 773–784. https://doi.org/10.1145/2901790.2901820

[55] Lixiu Yu and Jeffrey V. Nickerson. 2011. Cooks or Cobblers? Crowd Creativity through Combination. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1393–1402. https://doi.org/10.1145/1978942.1979147

[56] Alexey Zagalsky, Carlos Gómez Teshima, Daniel M. German, Margaret-Anne Storey, and Germán Poo-Caamaño. 2016. How the R Community Creates and Curates Knowledge: A Comparative Study of Stack Overflow and Mailing Lists. In *Proceedings of the 13th International Conference on Mining Software Repositories (MSR '16)*. Association for Computing Machinery, New York, NY, USA, 441–451. https://doi.org/10.1145/2901739.2901772

[57] Wenjun Zhou, Wangcheng Yan, and Xi Zhang. 2017. Collaboration for success in crowdsourced innovation projects: Knowledge creation, team diversity, and tacit coordination. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Received January 2020; revised June 2020; accepted July 2020