Enhancing Creativity as Innovation via Asynchronous Crowdwork

Pradeep K. Murukannaiah
Delft University of Technology
Delft, The Netherlands
p.k.murukannaiah@tudelft.nl

Nirav Ajmeri
University of Bristol
Bristol, UK
nirav.ajmeri@bristol.ac.uk

Munindar P. Singh
North Carolina State University
Raleigh, NC, USA
singh@ncsu.edu

ABSTRACT

Synchronous, face-to-face interactions such as brainstorming are considered essential for creative tasks (the *old* normal). However, face-to-face interactions are difficult to arrange because of the diverse locations and conflicting availability of people—a challenge made more prominent by work-from-home practices during the COVID-19 pandemic (the *new* normal). In addition, face-to-face interactions are susceptible to cognitive interference.

We employ crowdsourcing as an avenue to investigate creativity in asynchronous, online interactions. We choose product ideation, a natural task for the crowd since it requires human insight and creativity into what product features would be novel and useful. We compare the performance of solo crowd workers with asynchronous teams of crowd workers formed without prior coordination. Our findings suggest that, first, crowd teamwork yields fewer but more creative ideas than solo crowdwork. The enhanced team creativity results when (1) team workers reflect on each other's ideas, and (2) teams are composed of workers of reflective, as opposed to active or mixed, personality types. Second, cognitive interference, known to inhibit creativity in face-to-face teams, may not be significant in crowd teams. Third, teamwork promotes better achievement emotions for crowd workers. These findings provide a basis for trading off creativity, quantity, and worker happiness in setting up crowdsourcing workflows for product ideation.

CCS CONCEPTS

• Information systems → Crowdsourcing; **• Human-centered computing → Empirical studies in collaborative and social computing**.

KEYWORDS

Crowdsourcing, teamwork, creativity, personality, innovation

ACM Reference Format:
Pradeep K. Murukannaiah, Nirav Ajmeri, and Munindar P. Singh. 2022. Enhancing Creativity as Innovation via Asynchronous Crowdwork. In *14th ACM Web Science Conference 2022 (WebSci '22), June 26–29, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3501247.3531555

WebSci '22, June 26–29, 2022, Barcelona, Spain

ACM ISBN 978-1-4503-9191-7/22/06…$15.00
https://doi.org/10.1145/3501247.3531555

1 INTRODUCTION

We study the problem of how workers can asynchronously interact to jointly solve problems that require creativity. Addressing this problem is important to realize the vision of *social machines* [4, 9]. By capturing user interaction, we can imagine services that bring forth not just the semantics of the Web but its pragmatics [44], i.e., how it is applied. A major motivation behind social machines is to enable using computers on the Web to provide bookkeeping support while humans do more interesting and innovative work.

We focus on creativity as innovation (e.g., designing a product to solve perceived user needs) as opposed to expressive creativity (e.g., producing a work of art). Boden [7] refers to these two types of creativity as psychological (P-creativity) and historical (H-creativity) and argues that the explanation of the former also covers the latter.

A key motivation behind crowdsourcing is to provide a means to elicit human knowledge and insight. Typically, however, crowdsourcing is applied to tasks such as photo and video captioning and solving complex scientific problems such as protein folding. In contrast, we consider creativity, which is arguably one of the most crucial and challenging aspects of human intelligence [6]. In creative tasks, there may be no objective truth, and techniques such as selecting the majority answers would not work.

Previous research on crowdwork, including that by Kulkarni et al. [20], Oppenlaender et al. [31], and Yu and Nickerson [52], talks of creativity. However, creativity-as-innovation has largely been omitted from prior work. Questions on factors that support or inhibit crowd creativity, the tradeoffs involved, and how humans experience a creative crowdsourcing task remain largely unexplored.

We focus on the problem of product ideation, which seeks novel and useful product ideas. We adopt the framework of crowdsourcing, which involves members of a so-called "crowd" providing knowledge and insight that are combined to produce effective solutions. Product ideation is a natural crowdsourcing task—it is

- *meaningful* since crowd workers ideate on products they will likely use or consume in the future; and
- potentially *enjoyable* since it requires innovative thinking.

Both meaningfulness [49] and enjoyment [19] are shown to be intrinsic motivations for humans to participate in crowdwork.

In traditional organizational settings, (face-to-face) teamwork is known to stimulate creativity [35]. However, such teamwork is not always feasible. The COVID-19 pandemic situation makes clear that it is not always trivial for workers to be in the same physical space. Moreover, due to different timezones and family duties, it is not necessarily easy for workers to connect synchronously online. These challenges are even more pertinent for crowdwork. Whereas crowd teamwork is feasible [40], working in a team can be inconvenient and reduce a worker's flexibility because crowd workers may work from diverse locations, not know each other, multitask aggressively, and have disparate expertise and expectations.

We propose a crowd workflow, where teams are assembled on the fly. That is, when a crowd worker selects a task, the worker is assigned to a team based on predefined criteria such as team size and composition. Once in a team, workers can interact asynchronously and observe each other's work. Thus, our workflow does not require prior coordination between task requesters and workers.

Research Questions

We investigate whether asynchronous crowd teamwork promotes creativity and other positive outcomes via four research questions.

Creativity: Does asynchronous teamwork promote or inhibit crowd creativity?

Personality: Do crowd workers' personality traits influence their performance in asynchronous teamwork?

Efficiency: Does asynchronous teamwork influence crowd workers' efficiency (time, effort, and quantity)?

Emotions: Does asynchronous teamwork influence crowd workers' achievement emotions?

Although questions about creativity, personality, and efficiency have been explored, most of the existing works involve face-to-face settings. Crowdwork is different from these settings in that factors such as social inhibition and cognitive interference that inhibit creativity in face-to-face settings [33, 43] may not apply to crowdwork. Further, unlike face-to-face teams, crowd team members are usually unfamiliar with each other. Finally, we ask a question about emotions since facilitating positive achievement emotions [13, 25] can help attract and engage crowd workers. Intrinsic motivations such as enjoyment go beyond money [38] for sustaining crowd participation in a recurring task like product ideation.

We answer our research questions via an empirical study of 323 Amazon Mechanical Turk (MTurk) workers, who worked solo or in asynchronous teams. Our study is novel for two reasons. (1) Prior research has not explored the tradeoff between solo and team-based asynchronous crowdwork with respect to creativity, efficiency, and emotions. (2) Research on personality influences on crowd creativity, e.g., by Lykourentzou et al. [23] and by Yu and Nickerson [52], does not consider creativity as innovation. Answering our research questions provides insights on creating a crowdsourcing workflow that improves the quality of the work product as well as the quality of the workers' experience.

2 RELATED WORKS

Previous research on creativity has tackled related but not the same challenges as our work. Research on face-to-face teamwork suggests that group brainstorming stimulates new thinking [32, 35], but other research indicates that group brainstorming yields fewer and lower quality ideas than individual work [11]. Paulus [33] identifies two key creativity inhibitors: (1) cognitive interference (e.g., production blocking and task-irrelevant behaviors), and (2) social inhibition (e.g., evaluation apprehension and social loafing). However, these creativity inhibitors may not apply (at least, not as strongly) in crowdwork. For example, crowd workers may not be production blocked because of asynchronous teamwork, and they may not care about how strangers evaluate their work. Thus, findings in face-to-face teams may not generalize to crowdwork. This motivates us to investigate creativity afresh from the perspective of crowdwork.

Research on creativity in crowdwork has gained interest as crowdsourcing platforms have been increasingly used for studying creativity and executing creative tasks [31]. Although collaboration is a major theme in creativity research [14], factors determining successful collaboration can be quite different in face-to-face and crowd settings. For example, André et al. [2] advocate a sequential work structure, with clearly defined roles, over a simultaneous work structure for distributed collaboration. Yu and Nickerson [52] and Murukannaiah et al. [29, 30] employ a workflow in which one generation of crowd workers is stimulated by the works of a previous generation, improving the creativity of outcomes.

Since teamwork affects creativity, creating optimal crowd teams is a significant challenge. Building on insights from face-to-face teams, Lykourentzou et al. [23] employ team personality composition as a strategy for enhancing creativity. However, the types of creative tasks in Lykourentzou et al.'s study (generating an ad campaign) and our study (product ideation) are different.

Creativity in product ideation is a well-researched direction [5, 8, 48]. However, these works facilitate creativity as innovation in face-to-face workshops and focus groups with a limited number of participants. In contrast, we exploit crowdsourcing to democratize product ideation by involving a large stakeholder group.

Since creativity is task-specific [3], we investigate the influence of crowd team personality on creativity from the standpoint of product ideation. To the best of our knowledge, none of the existing works study the implications of individual (solo) and team-based crowdwork on the creative performance of the workers.

3 METHOD

We conducted a user study on the Amazon MTurk platform to answer our research questions. The study was approved by the Institutional Review Board for the Protection of Human Subjects at the institute where one of the authors was employed when the study was conducted. Informed consent was obtained from each participant. All methods were carried out in accordance with relevant guidelines and regulations.

The subjects in our study participated in a workflow shown in Figure 1. We developed a web application to configure the crowd workflow as solo-work or teamwork and to facilitate the study.

3.1 Pre Surveys

We obtained a worker's DISC personality trait and demographics. We adopted the DISC [26] personality model since it is well-suited to study personality influence in teamwork [23].

Figure 2 shows how the DISC model categorizes individuals based on task and people orientation as dominant (D), influential (I), conscientious (C), or steady (S).

We employed a free online questionnaire [1] for measuring the DISC personality [26] of participants. The questionnaire includes 28 groups of four items each. For each group, a participant must indicate one item they are *most like* and one item they are *least like*. There are commercial questionnaires, e.g., Everything DiSC [42], for measuring DISC personality. However, administering a commercial

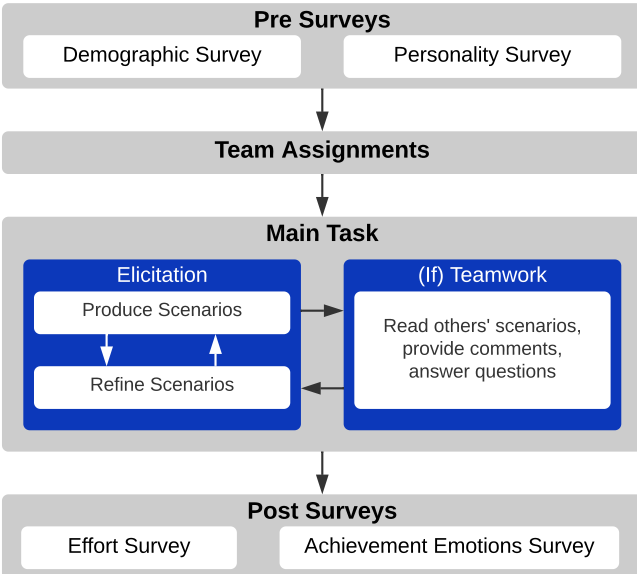


Figure 1: Workflow structure showing configurable tasks.

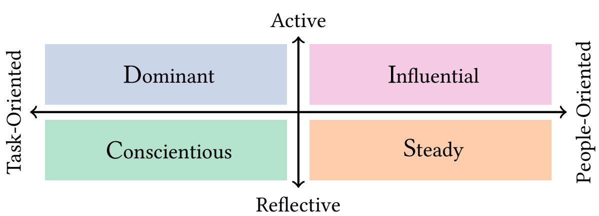


Figure 2: Marston's [26] DISC personality model.

questionnaire in our crowd study was not feasible financially and logistically (we would have to pay crowd workers upfront to take the personality test and then participate in our study). Table 1 shows the distribution of DISC traits and Spearman's rank correlations between each other in our dataset.

Table 1: Distribution, correlations of DISC traits in our data

	D	I	S	C	Distribution
D	1	0.17	−0.77	−0.51	16.2%
I	–	1	−0.36	−0.77	20.9%
S	–	–	1	0.31	28.3%
C	–	–	–	1	34.7%

3.2 Team Assignment

We employed a lightweight mechanism for assembling crowd teams dynamically, similar to *flash teams* [37]. An analyst specifies team size and composition criteria. To avoid prior coordination and risk of cancellation, we added crowd workers incrementally (as they accepted our task on MTurk) to an open team whose composition criteria they matched.

We employed solo workers and teams of four, small enough to form teams on the fly so that the earlier members would not have left when the later members arrived. For balance between the solo and group treatments, we assigned the first four workers to a solo workflow, the next four workers to a team workflow, and so on.

To analyze the influence of team personality composition, we created *active*, *reflective*, and *mixed* personality teams based on our initial analysis from a set of teams not controlled for personality. First, 95 participants worked solo, 108 participants worked in teams not controlled for personality composition (27 teams). We investigated research questions pertaining to creativity, efficiency, and emotions based on these workers' responses. Then, an additional 120 participants worked in teams controlled for personality composition (10 teams each of active, reflective, and mixed personality). We investigated a research question pertaining to personality based on these workers' responses.

3.3 Main Task

The main task assigned to the crowd workers was to identify usage scenarios for four smarthome (product) use cases. As an emerging application domain, smarthome use cases provide room for exercising creativity in identifying usage scenarios. Further, crowd workers do not need specialized knowledge to think about homes. In particular, the task was for a crowd worker to envision specific capabilities (or features) for a smarthome product. The worker was to express each product capability using a natural language template (provided by us) that indicated "upon this trigger event, under this condition, the product must respond as follows." For example, when a *resident leaves a room* provided *when no motion is detected*, *fade down the lights in the room*. Table 2 lists the number of usage scenarios crowd workers produced for each use case.

Table 2: Number of scenarios produced for each use case.

Use Case	#
If there is no movement or noise in a room for a preconfigured duration, the smarthome should automatically fade off the lights in that room, so as to save energy	507
If a person walks into a dark room, the smarthome should automatically turn on the lights in that room or brighten them if they were faded down so as to enhance convenience and safety	431
If there is movement outside the home when it is dark outside, the smarthome should automatically turn on some lights so that any potential intruders are scared off	428
A smarthome should optimize the lighting (e.g., brightness and color) in a room based on the activity of the people in that room so as to enhance user experience (assume that the desired lighting for an activity is preconfigured, e.g., dim lighting for watching TV)	457

We offered crowd workers a base reward of USD 3 for producing a minimum of two usage scenarios. We incentivized creativity by offering a bonus reward of up to USD 3, for a total of USD 6, for producing creative usage scenarios.

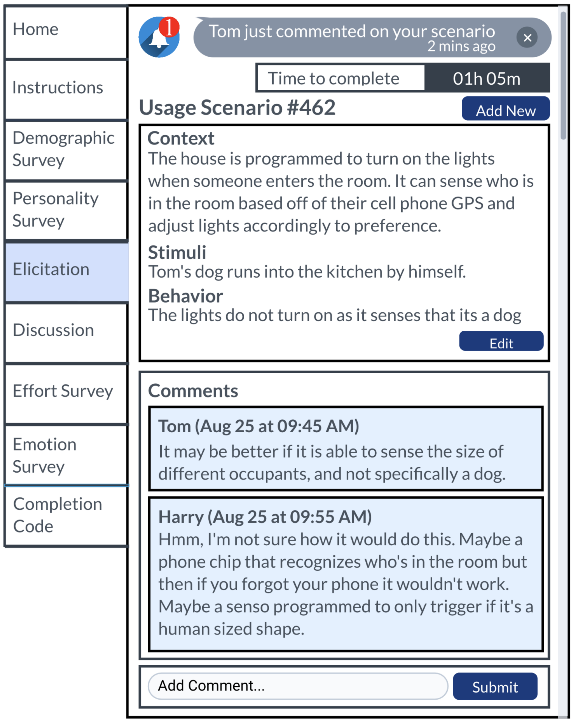


Figure 3: A wireframe showing a worker proposing a scenario and engaging in collaboration with others.

Crowd team workers collaborate by viewing and commenting on others' scenarios, requesting and providing clarifications, and refining the scenarios. The workers are notified each time a team member adds a scenario or responds to a member. Figure 3 shows a wireframe of our user interface. We include a wireframe (instead of a screenshot from the web app) to enhance readability.

Since not all team members join at once, to ensure they overlap, we required workers to wait at least three hours before submitting their task. We encouraged them to visit the web application frequently, collaborate, and refine scenarios in the three-hour window.

3.4 Post Surveys

Each worker reported the total time spent on the main task (across multiple sessions) in minutes and the effort perceived on a Likert scale of 1 (easy) to 5 (difficult).

We measured workers' achievement emotions to understand their intrinsic motivation for participating via Pekrun et al.'s [36] model (Table 3), which includes both positive and negative emotions about the *process* of producing scenarios and the *outcome* that a worker produced creative scenarios.

The model of achievement emotions that Pekrun et al. [36] developed is general, but the questionnaire (AEQ) they developed is specific to educational settings. Thus, we asked our study participants to self-report achievement emotions by answering direct questions about enjoyment and boredom (of the process) and confidence and anxiety (about the outcome). Each worker reported achievement emotions perceived while working (process) and upon completion of the task (outcome), each on a Likert scale of 1 (low) to 5 (high).

Table 3: Pekrun et al.'s taxonomy of achievement emotions.

Focus \ Valence	Positive	Negative
Process	Enjoyment	Boredom
Outcome	Confidence	Anxiety

3.5 Creativity Evaluation

Following the bipartite definition, we measure the creativity of a scenario as its novelty and usefulness. Novelty refers to the extent to which an end-user is likely to find a smarthome feature based on a scenario unique, surprising, and innovative among other smarthome features on a Likert scale of 1 (not at all novel) to 5 (very novel). Usefulness refers to the extent to which an end-user is likely to find a smarthome feature based on a scenario to be effective in achieving one or more (explicit or implicit) goals underlying the use case on a Likert scale of 1 (not at all useful) to 5 (very useful).

We developed a *rubric* for rating novelty and usefulness. We started from an initial rubric and a sample of 50 scenarios. We, iteratively, (1) rated the sample scenarios; (2) computed agreement scores; (3) discussed disagreements; and (4) refined the rubric.

Following this rubric, we rated a random sample of 700 usage scenarios (out of 1,823) on novelty and usefulness. Two annotators rated each scenario in batches of 50 and resolved conflicts in multiple iterations until the interrater reliability measured via intraclass correlation coefficient (ICC) [15] was at least 75% (which is considered excellent). The raters did not know whether a scenario was produced by a solo or team crowdworker. The raters discarded 61 scenarios that were unclear or irrelevant. Our results are based on creativity ratings of 639 scenarios.

3.6 Analyses

We employ (1) Wilcoxon's rank-sum test (nonparametric) to compare the distributions of two ordinal samples (5% significance level); (2) Kruskal-Wallis test (nonparametric extension of ANOVA) to compare the distributions of more than two ordinal samples (5% significance level); (3) Dunn's [12] multiple comparison test with the Holm-Bonferroni correction to compare pairs of samples when the Kruskal-Wallis test rejects the null hypothesis; (4) Cliff's [10] Delta to measure the effect sizes (the amount of difference). The magnitude of the delta is estimated according to the suggested thresholds: $\delta < 0.147$ is negligible (N); $\delta < 0.33$ is small (S); $\delta < 0.474$ is medium (M); and large (L), otherwise.

Data Availability. The data is shared publicly [28]. It includes all usage scenarios crowd workers produced and data necessary to reproduce our results.

4 RESULTS

We examine the relationships between crowd teamwork and (1) creativity, (2) team personality composition, (3) time and effort, and

(4) achievement emotions. We present results, key findings, and takeaways about each of these relationships.

4.1 Creativity

We treat creativity as combining *novelty* and *usefulness* [39]. Thus, a creative product feature is both original and appropriate in a specific context of use.

As Figure 4 shows, we did not observe a significant difference in novelty or usefulness of the scenarios produced by solo and team workers. However, since teams are formed on the fly, a team worker entering a team early may have fewer opportunities to work with teammates joining the team later. Thus, in Figure 5, we compare the creativity of solo workers with those team workers entering their team first or last.

We make the following observations. (1) Team workers who join a team last produce scenarios that are significantly more novel than scenarios produced by solo workers. (2) Solo and team workers do not differ in terms of the usefulness of the scenarios they produce. (3) The number of interactions (comments, questions, and answers) in teams was typically low (often none): mean=3.05, median=1, SD=3.9. Thus, we attribute the enhanced novelty of teamwork to workers' ability to observe and learn from each other, which does not require active communication.

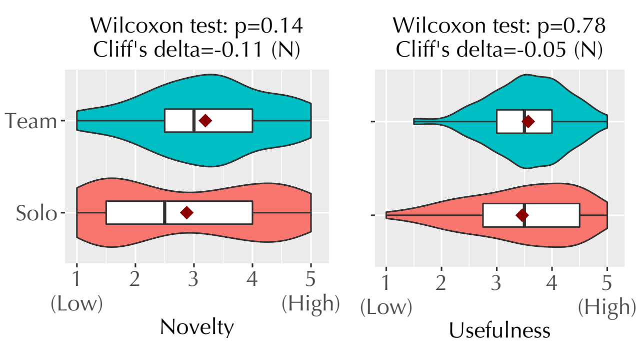
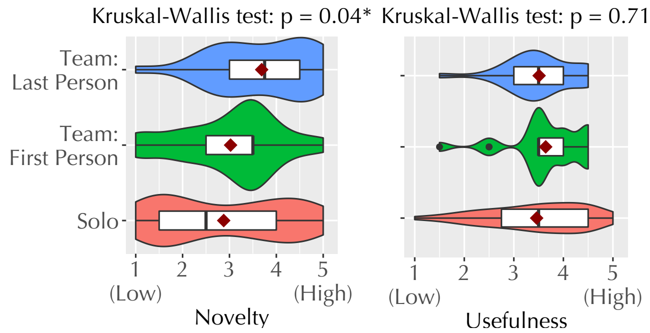


Figure 4: Creativity: solo-work vs. teamwork.

Takeaway from the research question
Crowd teamwork yields novel outcomes. However, for full benefit, teams must be formed swiftly, or team workers must be incentivized to wait for teams to form.

4.2 Personality

Figures 6 and 7 compare the creativity of solo and team workers of different DISC personality types. Subsequently, we create active, reflective, and mixed teams by controlling team personality composition. Figure 8 compares the creativity of active, reflective, and mixed teams, where (1) each member of an active team is of type D or I, the active personality types, (2) each member of a reflective team is of type S or C, the reflective personality types, and (3) a mixed team consists of two members, respectively, of the active and reflective personality types.



(Novelty)	H-B *p*-value		Cliff's Delta	
	Solo	Team (First)	Solo	Team (First)
Team (First)	0.77		−0.05 (N)	
Team (Last)	$\mathbf{0.03}^*$	0.14	−0.35 (**M**)	−0.09 (N)

Figure 5: Creativity: solo and team (first and last joiners)

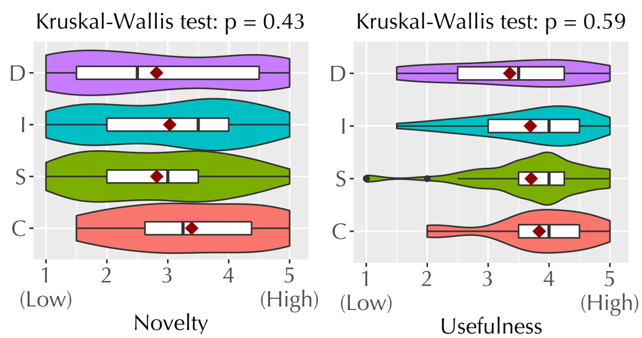
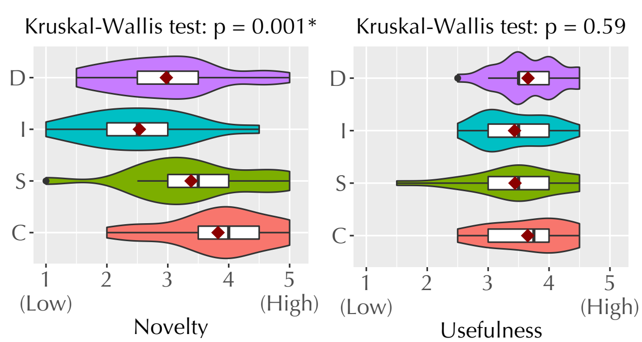


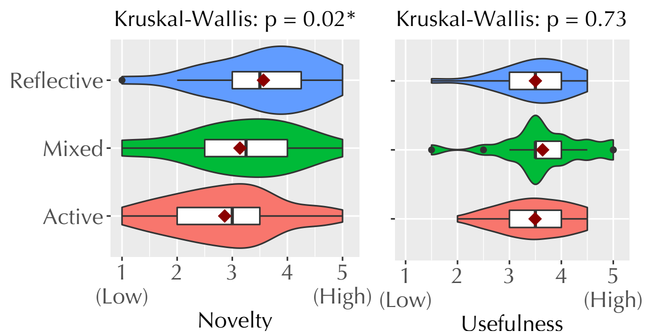
Figure 6: Creativity vis à vis personality types in solo work.



(Novelty)	H-B p-value			Cliff's Delta		
	C	D	I	C	D	I
D	$\mathbf{0.02}^*$			−0.51 (**L**)		
I	$\mathbf{0.001}^*$	0.23		−0.70 (**L**)	−0.26 (**S**)	
S	0.53	0.38	0.07	−0.28 (**S**)	0.25 (**S**)	0.50 (**L**)

Figure 7: Creativity vis à vis personality types in teamwork.

We make the following observations. (1) There was no significant difference in the novelty or usefulness of the scenarios among solo



(Novelty)	H-B p-value		Cliff's Delta	
	Active	Mixed	Active	Mixed
Mixed	0.27		−0.18 (**S**)	
Reflective	**0.02***	0.40	−0.41 (**M**)	−0.23 (**S**)

Figure 8: Effect of personality composition on teams.

workers of different personalities. (2) The small peak of high novelty in Figure 5 for those who join a team first suggests that early joiners can also benefit from teamwork if they wait for the team to fully form, but the large peak of medium novelty suggests that many first joiners did not wait for the team to fully form. (3) Team workers of the reflective (S and C) personality produced scenarios that are significantly more novel than the scenarios produced by workers of active (D and I) personality. (4) Reflective teams produced scenarios of a significantly greater novelty than scenarios produced by active teams. (5) Personality types and team personality composition did not influence the usefulness of the scenarios produced.

Takeaway from the research question

Controlling team personality composition to include reflective (Steady and Conscientious) workers in teams is an effective mechanism for enhancing novel outcomes.

4.3 Efficiency

Figure 9 compares the time and effort expended and the fluency (number of scenarios produced) of solo and team workers. Table 4 compares the Pearson's correlation between the number of all scenarios and the number of novel and useful scenarios (rated 4 or 5) produced by solo and team workers.

Table 4: Quantity vs. Creativity: Correlations between the number of all vs. the number of novel or useful scenarios

	Pearson's correlation			
	Solo		Team	
Scenarios	Novel	Useful	Novel	Useful
All	0.49 ($p < 0.01$)	0.46 ($p < 0.01$)	0.38 ($p < 0.01$)	0.32 ($p = 0.04$)

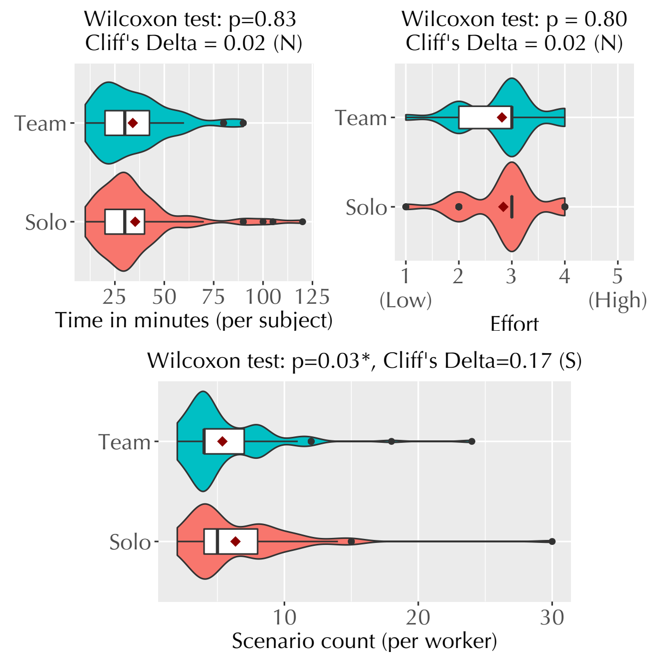


Figure 9: Time, effort, and quantity: solo vs. teamwork.

We make the following observations. (1) The time and effort expended by solo and team workers do not differ significantly. (2) Solo workers produce more scenarios than team workers. (3) The correlation between the number of scenarios and the number of novel and useful scenarios produced by solo and team workers is moderate (positive).

Takeaway from the research question

Solo workers are more fluent than team workers. However, the correlation between fluency and creativity is moderate but stronger for solo than team workers.

4.4 Emotions

Figure 10 compares the achievement emotions (process and outcome) of solo and team workers.

We make the following observations. (1) Team workers are more confident and less anxious about the scenarios they produce than solo workers. (2) Neither solo nor team workers are bored, but team workers enjoy the process more than solo workers.

Takeaway from the research question

Team workers perceive better achievement emotions than solo workers. Thus, teamwork can enhance workers' intrinsic motivation to participate in crowdwork.

5 DISCUSSION

Designers and developers need new methods to elicit ideas for applications with a consumer focus [18]. Such methods should provide insights into what prospective users want in a variety

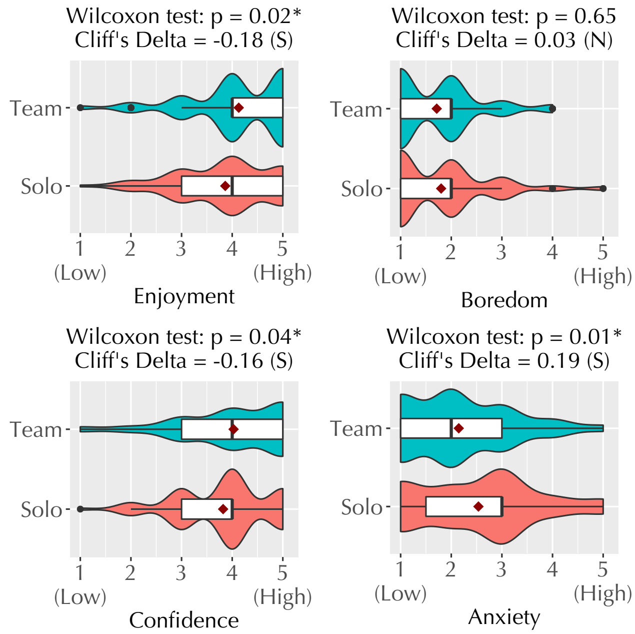


Figure 10: Achievement emotions in solo and team work.

of contexts. First, these objectives are best served by engaging outsiders—something crowdwork offers naturally. Second, developers' objectives are best served by eliciting creative ideas. We demonstrate how creative ideas can be crowdsourced.

A charm of crowdwork is that it does not require face-to-face interaction, so workers can work whenever and wherever convenient. In addition, Zeng et al. [53] find that the freshness of a team (i.e., absence of prior collaboration among team members) is associated with greater originality. Freshness is also an attribute of crowd teams since the members of a crowd team, especially those formed on the fly, are unlikely to have prior collaboration.

Our investigation reveals that subtle characteristics of how crowdwork, such as whether it is carried out solo or in teams, how the teams are composed, and how team members interact with each other, are important. Our approach yields guidance on how product ideation teams ought to be composed on the fly and how their work ought to be structured to produce the best outcomes in terms of both the project collateral and the joy perceived by the workers.

5.1 Lightweight Distributed Teamwork

Teams formed on the fly provide flexibility for both task requesters and workers. Such teams of strangers are best for tasks such as product ideation that do not require active coordination between workers.

However, Salehi et al. [40] find that familiar teams are more efficient since members know each other's strengths and weaknesses and experience better psychological acceptance and safety, although the levels of cooperation and conflict vary. Thus, for tasks requiring cooperation between workers, approaches such as team dating can be used, where workers interact ("date") first and then are assigned to teams based on compatibility [24].

5.2 Cognitive Interference

We did not find cognitive interference, a creativity inhibitor in face-to-face teams, to be a major concern for our crowd teams. Team workers do not engage in long conversations. Thus, their creativity is not influenced by the number of interactions between them. Yet, team workers benefit from exposure to each other's ideas. Workers who joined late and thus saw previous work by others produced ideas that were more novel than other workers. Thus, to facilitate novel outcomes, teams must be formed swiftly, with workers who join early incentivized to wait for the team to fully form.

5.3 Team Personality Composition

We found a simple but effective means of enhancing novel outcomes by controlling a team's personality composition. Specifically, reflective (Steady and Conscientious) workers yield higher novelty than other workers in the product ideation task. This result refines Lykourentzous et al.'s [23] finding that mixed teams are more creative. We attribute this observation to the difference in task type between the two studies. Lykourentzous et al.'s task of creating an ad campaign requires active coordination, but product ideation does not. We conjecture that active (Dominant and Influential) workers serve as team coordinators in such tasks.

5.4 Fluency and Creativity

In face-to-face teamwork, whereas Osborn [32] claims that group brainstorming promotes creativity, Diehl and Stroebe [11] and Lamm and Trommsdorff [21] find that group brainstorming can hurt both the quality and the quantity of ideas. We refine Osborn's and Diehl and Stroebe's observations to crowd teamwork. Whereas solo workers produce more ideas than crowd workers, the correlation between fluency and creativity is moderate for solo workers and low for team workers. However, team workers produce ideas that are more novel than solo workers, demonstrating the difference between teamwork in face-to-face and crowd settings.

5.5 Novelty and Usefulness

Existing works treat creativity as a unitary measure. We understand creativity as *novelty* and *usefulness* [39]. We observe that what applies to novelty does not necessarily apply to usefulness and vice versa. We posit that future work on creativity should consider both novelty and usefulness as dimensions to measure creativity.

5.6 Worker Happiness

Paulus et al. [34] find that most people perceive their ideas and productivity as more favorable when working in a team than working solo. In the same spirit, Sutton and Hargadon [47] describe that organizations adopt group brainstorming, despite evidence for its inefficiency in generating ideas because it yields positive outcomes such as developing organizational memory, supporting peer respect, and establishing organizational norms. Along similar lines, we find that crowd team workers are less anxious and perceive better process-related and outcome-related achievement emotions than solo crowd workers. Thus, we conjecture that supporting crowd teamwork is necessary for establishing a sustainable crowd community that contributes ideas to the development and evolution of a product.

5.7 Threats to Validity

We identify four threats to the validity of our findings. First, we employed a free online test to measure the DISC personality traits of workers and validate the hypotheses corresponding to personality. However, the reliability and validity of this test are unknown.

Second, to the best of our knowledge, there is no well-established questionnaire for measuring achievement emotions in a creative task, such as the one adopted in our study. Thus, we developed a simple questionnaire of our own. Since we ask direct questions about outcome-related and process-related achievement emotions, we expect the reported emotions to be valid and reliable.

Third, we asked participants to self-report their time spent on the assigned task, excluding interruptions. However, for any participants who may have been multitasking (i.e., splitting their attention between the assigned task and other activities), the time estimate they reported for the assigned task may not be accurate. However, since all participants estimated the time they spent, we expect the time comparison between the various treatments to be fair.

Fourth, crowd workers may have joined a team at different times. As a result, the amount of overlap between teammates may vary, both within a team and across teams. To ensure that each team member overlaps with at least some of the other team members, we required team workers to wait at least three hours before submitting their tasks. It is possible (and indeed quite likely) that workers switched to other tasks during this wait time. Such multitasking by workers may explain the low extent of team interactions we observed (e.g., when a team member produced a scenario, the other team members were not readily available to comment on it). To make switching back to our task easy for team workers, we sent a notification (within our web application) to each team member when a comment was added.

6 CONCLUSIONS AND DIRECTIONS

The true potential of crowdsourcing lies in realizing sociotechnical systems [45], where humans contribute creative work and machines perform mundane or cognitively difficult tasks for humans [16].

Realizing sociotechnical systems requires addressing not only technical challenges (e.g., reward mechanisms) but also human-centered challenges such as enhancing creativity, engagement, peer learning, and enjoyment. Creativity in the design of sociotechnical systems can help by generating new solutions to capture stakeholders' requirements and their attitudes regarding acceptable solutions given their values. In particular, we need ways to discover potential solutions that trade off between their values and other preferences to ensure fair and ethical outcomes [27, 46, 51].

Our results can guide the creation of hybrid intelligence systems [17, 22, 50], where crowd workers and computational techniques work in tandem. Directions for enhancing our crowd workflow with computational techniques include automatically composing teams of crowd workers on the fly, intelligently triggering workers to be creative, rewarding workers to promote creative effort, and detecting creative ideas automatically.

Finally, the ethics of crowdsourcing is garnering attention [41]. This paper offers two ethical benefits. First, it identifies ways that could reduce a worker's anxiety. Second, it promotes having a lower quantity of higher quality work.

ACKNOWLEDGMENTS

MPS thanks the NSF (grant IIS-1908374) for partial support.

REFERENCES

[1] 123test BV. 2018. DISC Personality Test. https://www.123test.com/disc-personality-test/.

[2] Paul André, Robert E. Kraut, and Aniket Kittur. 2014. Effects of Simultaneous and Sequential Work Structures on Distributed Collaborative Interdependent Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto, 139–148. https://doi.org/10.1145/2556288.2557158

[3] John Baer. 2012. Domain Specificity and the Limits of Creativity Theory. *The Journal of Creative Behavior* 46, 1 (2012), 16–29.

[4] Tim Berners-Lee. 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. Harper Business, New York.

[5] Tanmay Bhowmik, Nan Niu, Anas Mahmoud, and Juha Savolainen. 2014. Automated support for combinational creativity in requirements engineering. In *Proceedings of the 22nd IEEE International Requirements Engineering Conference*. IEEE, Karlskrona, Sweden, 243–252. https://doi.org/10.1109/RE.2014.6912266

[6] Margaret A. Boden. 1998. Creativity and Artificial Intelligence. *Artificial Intelligence* 103, 1–2 (Aug. 1998), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

[7] Margaret A. Boden. 2005. What is creativity? In *Creativity in human evolution and prehistory*, Steven Mithen (Ed.). Routledge, London, 27–55.

[8] Corentin Burnay, Jennifer Horkoff, and Neil Maiden. 2016. Stimulating Stakeholders' Imagination: New Creativity Triggers for Eliciting Novel Requirements. In *Proceedings of the 24th IEEE International Requirements Engineering Conference*. IEEE, Beijing, 36–45. https://doi.org/10.1109/RE.2016.36

[9] Amit K. Chopra and Munindar P. Singh. 2016. From Social Machines to Social Protocols: Software Engineering Foundations for Sociotechnical Systems. In *Proceedings of the 25th International World Wide Web Conference*. ACM, Montréal, 903–914. https://doi.org/10.1145/2872427.2883018

[10] Norman Cliff. 2014. *Ordinal Methods for Behavioral Data Analysis*. Psychology Press, New York.

[11] Michael Diehl and Wolfgang Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology* 53, 3 (1987), 497–509.

[12] Olive Jean Dunn. 1964. Multiple comparisons using rank sums. *Technometrics* 6, 3 (1964), 241–252.

[13] Tom Edixhoven, Sihang Qiu, Lucie Kuiper, Olivier Dikken, Gwennan Smitskamp, and Ujwal Gadiraju. 2021. Improving Reactions to Rejection in Crowdsourcing Through Self-Reflection. In *Proceedings of the 13th ACM Web Science Conference (WebSci '21)*. ACM, Virtual Event, United Kingdom, 74–83. https://doi.org/10.1145/3447535.3462482

[14] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018. Twenty Years of Creativity Research in Human-Computer Interaction: Current State and Future Directions. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, Hong Kong, 1235–1257. https://doi.org/10.1145/3196709.3196732

[15] Kevin A. Hallgren. 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology* 8, 1 (2012), 23–34.

[16] Jiuchuan Jiang, Bo An, Yichuan Jiang, Donghui Lin, Zhan Bu, Jie Cao, and Zhifeng Hao. 2018. Understanding Crowdsourcing Systems from a Multiagent Perspective and Approach. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 13, 2 (2018), 8:1–8:32. https://doi.org/10.1145/3226028

[17] Ece Kamar. 2016. Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, New York, 4070–4073.

[18] Georgi M. Kanchev, Pradeep K. Murukannaiah, Amit K. Chopra, and Pete Sawyer. 2017. Canary: Extracting requirements-related information from online discussions. In *Proceedings of the 25th IEEE International Requirements Engineering Conference*. IEEE, Lisbon, 31–40.

[19] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. Worker motivation in crowdsourcing: A study on Mechanical Turk. In *Proceedings of the Seventeenth Americas Conference on Information Systems*. Association for Information Systems, Detroit, 1–11.

[20] Anand Kulkarni, Prayag Narula, David Rolnitzky, and Nathan Kontny. 2014. Wish: Amplifying Creative Ability with Expert Crowds. In *Proceedings of the 2nd AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*. AAAI, Pittsburgh, 112–120. http://www.aaai.org/ocs/index.php/HCOMP/HCOMP14/paper/view/8977

[21] Helmut Lamm and Gisela Trommsdorff. 1973. Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. *European Journal of Social Psychology* 3, 4 (1973), 361–388.

[22] Enrico Liscio, Michiel van der Meer, Luciano C. Siebert, Catholijn M. Jonker, Niek Mouter, and Pradeep K. Murukannaiah. 2021. Axies: Identifying and Evaluating Context-Specific Values. In *Proceedings of the 20th Conference on Autonomous*

Agents and MultiAgent Systems (AAMAS '21). IFAAMAS, London, 799–808.

[23] Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P. Dow. 2016. Personality Matters: Balancing for Personality Types Leads to Better Outcomes for Crowd Teams. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, San Francisco, 260–273.

[24] Ioanna Lykourentzou, Robert E. Kraut, and Steven P. Dow. 2017. Team Dating Leads to Better Online Ad Hoc Collaborations. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, Portland, Oregon, 2330–2343. https://doi.org/10.1145/2998181.2998322

[25] Stacy Marsella and Jonathan Gratch. 2014. Computationally modeling human emotion. *Communications of the ACM (CACM)* 57, 12 (2014), 56–67. https://doi.org/10.1145/2631912

[26] William M. Marston. 2013. *Emotions of Normal People*. Taylor & Francis, London.

[27] Pradeep K. Murukannaiah, Nirav Ajmeri, Catholijn M. Jonker, and Munindar P. Singh. 2020. New Foundations of Ethical Multiagent Systems. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. IFAAMAS, Auckland, 1706–1710. https://doi.org/10.5555/3398761.3398958 Blue Sky Ideas Track.

[28] Pradeep K. Murukannaiah, Nirav Ajmeri, Muhammad Fazalul Rahman, and Munindar P. Singh. 2020. Enhancing Crowd Creativity as Innovation via Teamwork: Dataset. https://doi.org/10.5281/zenodo.4308066. https://doi.org/10.5281/zenodo.4308066

[29] Pradeep K. Murukannaiah, Nirav Ajmeri, and Munindar P. Singh. 2016. Acquiring Creative Requirements from the Crowd: Understanding the Influences of Personality and Creative Potential in Crowd RE. In *Proceedings of the 24th IEEE International Requirements Engineering Conference*. IEEE, Beijing, 176–185.

[30] Pradeep K. Murukannaiah, Nirav Ajmeri, and Munindar P. Singh. 2017. Toward Automating Crowd RE. In *Proceedings of the 25th IEEE International Requirements Engineering Conference (RE)*. IEEE Computer Society, Lisbon, 512–515. https://doi.org/10.1109/RE.2017.74 Data Track.

[31] Jonas Oppenlaender, Kristy Milland, Aku Visuri, Panos Ipeirotis, and Simo Hosio. 2020. Creativity on Paid Crowdsourcing Platforms. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, 1–14.

[32] Alex F. Osborn. 1957. *Applied Imagination: Principles and Procedures of Creative Thinking*. Scribners, New York.

[33] Paul B. Paulus. 2000. Groups, teams, and creativity: The creative potential of idea-generating groups. *Applied Psychology* 49, 2 (2000), 237–262.

[34] Paul B. Paulus, Mary T. Dzindolet, George Poletes, and L. Mabel Camacho. 1993. Perception of Performance in Group Brainstorming: The Illusion of Group Productivity. *Personality and Social Psychology Bulletin* 19, 1 (1993), 78–89. https://doi.org/10.1177/0146167293191009

[35] Paul B. Paulus and Huei-Chuan Yang. 2000. Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes* 82, 1 (2000), 76–87. https://doi.org/10.1006/obhd.2000.2888

[36] Reinhard Pekrun, Andrew J. Elliot, and Markus A. Maier. 2006. Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology* 98, 3 (Aug. 2006), 583–597.

[37] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert Crowdsourcing with Flash Teams. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. ACM, Honolulu, 75–85. https://doi.org/10.1145/2642918.2647409

[38] Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. 2011. An Assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *Proceedings of the International AAAI Conference on Web and Social Media*. AAAI Press, Barcelona, 321–328.

[39] Mark A. Runco and Garrett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24, 1 (2012), 92–96. https://doi.org/10.1080/10400419.2012.650092 arXiv:http://dx.doi.org/10.1080/10400419.2012.650092

[40] Niloufar Salehi, Andrew McCabe, Melissa Valentine, and Michael Bernstein. 2017. Huddler: Convening Stable and Familiar Crowd Teams Despite Unpredictable Availability. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Portland, OR, 1700–1713. https://doi.org/10.1145/2998181.2998300

[41] Daniel Schlagwein, Dubravka Cecez-Kecmanovic, and Benjamin Hanckel. 2019. Ethical norms and issues in crowdsourcing practices: A Habermasian analysis. *Information Systems Journal* 29, 4 (2019), 811–837. https://doi.org/10.1111/isj.12227 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/isj.12227

[42] Mark Scullard and Dabney Baum. 2015. *Everything DiSC Manual*. John Wiley & Sons, Hoboken, NJ.

[43] Dean Keith Simonton. 2000. Creativity: Cognitive, personal, developmental, and social aspects. *American Psychologist* 55, 1 (2000), 151–158. https://doi.org/10.1037/0003-066X.55.1.151

[44] Munindar P. Singh. 2002. The Pragmatic Web. *IEEE Internet Computing (IC)* 6, 3 (May 2002), 4–5. https://doi.org/10.1109/MIC.2002.1003124 Instance of the column *Being Interactive*.

[45] Munindar P. Singh. 2013. Norms as a Basis for Governing Sociotechnical Systems. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 1 (Dec. 2013), 21:1–21:23. https://doi.org/10.1145/2542182.2542203

[46] Munindar P. Singh. 2022. Consent as a Foundation for Responsible Autonomy. *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)* 36 (Feb. 2022), 6 pages. Blue Sky Track.

[47] Robert I. Sutton and Andrew Hargadon. 1996. Brainstorming groups in context: Effectiveness in a product design firm. *Administrative Science Quarterly* 41, 4 (1996), 685–718.

[48] Richard Berntsson Svensson and Maryam Taghavianfar. 2015. Selecting creativity techniques for creative requirements: An evaluation of four techniques using creativity workshops. In *Proceedings of the 23rd IEEE International Requirements Engineering Conference*. IEEE, Ottawa, 66–75. https://doi.org/10.1109/RE.2015.7320409

[49] Yuushi Toyoda, Gale Lucas, and Jonathan Gratch. 2020. The Effects of Autonomy and Task Meaning in Algorithmic Management of Crowdwork. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '20)*. IFAAMAS, Auckland, New Zealand, 1404–1412.

[50] Michiel van der Meer, Enrico Liscio, Catholijn M. Jonker, Aske Plaat, Piek Vossen, and Pradeep K. Murukannaiah. 2022. HyEnA: A Hybrid Method for Extracting Arguments from Opinions. In *Proceedings of the First International Conference on Hybrid Human-Artificial Intelligence (HHAI)*. IOS Press, Amsterdam, 1–16. To appear.

[51] Jessica Woodgate and Nirav Ajmeri. 2022. Macro Ethics for Governing Equitable Sociotechnical Systems. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. IFAAMAS, Online, 1824–1828. https://doi.org/10.5555/3535850.3536118 Blue Sky Ideas Track.

[52] 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*. ACM, Vancouver, 1393–1402. https://doi.org/10.1145/1978942.1979147

[53] An Zeng, Ying Fan, Zengru Di, Yougui Wang, and Shlomo Havlin. 2021. Fresh teams are associated with original and multidisciplinary research. *Nature Human Behaviour* 5, 10 (2021), 1314–1322. https://doi.org/10.1038/s41562-021-01084-x