Effects of Simultaneous and Sequential Work Structures on Distributed Collaborative Interdependent Tasks

Paul André, Robert E. Kraut, Aniket Kittur
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA
{pandre,robert.kraut,nkittur}@cs.cmu.edu

ABSTRACT
Distributed online groups have great potential for generating interdependent and complex products like encyclopedia articles or product design. However, coordinating multiple group members to work together effectively while minimizing process losses remains an open challenge. We conducted an experiment comparing the effectiveness of two coordination strategies (simultaneous vs. sequential work) on a complex creative task as the number of group members increased. Our results indicate that, contrary to prior work, a sequential work structure was more effective than a simultaneous work structure as the size of the group increased. A mediation analysis suggests that social processes such as territoriality partially accounts for these results. A follow up experiment giving workers specific roles mitigated the detrimental effects of the simultaneous work structure. These results have implications for small group theory and crowdsourcing research.

Author Keywords
group size; coordination; interdependence; small groups.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI).

INTRODUCTION

Distributed groups routinely collaborate on complex, creative tasks like designing products, writing encyclopedia articles, and developing software. The number of people to devote to these creative tasks and the way to organize their effort are crucial questions for any distributed work group. Aiming for optimal output and the best use of each worker, factors such as group size and coordination mechanisms have been studied both experimentally [20, 23, 29] and in organizational settings [14, 42].

Structural contingency theory [13] suggests there is no single effective communication structure for all organizations or tasks. Rather, the optimal structure or coordination method varies according to factors such as group size and task uncertainty. Routinized assembly-line work is at one extreme of a task-uncertainty spectrum while complex, creative tasks are at the other. In many creative tasks, the work process does not follow a rigid sequence and is unable to be defined in advance, the task is interdependent where work done by one person has unpredictable effects on what a coworker will need to do next, and the output cannot be defined in advance.

Prior research indicates when people are collaborating on uncertain and interdependent tasks, they require 'organic coordination' or coordination by 'mutual adjustment.' That is, they need rich, fast, responsive communication to help them understand the current state of a task and others' activities, so that they can adapt their own work in response [13]. Thus, groups are most successful and better able to take advantage of a larger workforce—e.g., the workers' effort, diverse talents and perspectives, cognitive stimulation, and increased error-checking [23, 28, 29]—when they work simultaneously [40, 13].

On the other hand, simultaneous work can lead to process losses from production blocking, social influence, and social loafing [23, 35]. Even for relatively independent tasks, such as brainstorming, the presence of more people can reduce each person's quality and productivity, undermining the benefit of adding more people [11, 23]. As complexity, uncertainty and interdependence of the work increases, communication and coordination difficulties can increase even more, undermining the benefits of extra personnel [5, 39, 40].

Because adding more workers to a project brings both benefits and challenges, understanding how to best structure interdependent tasks is an important and open question [26]. Typically, researchers have examined the impact of group size by constructing groups of people *simultaneously* working on a task [23, 25, 29]. Yet in many cases those workers could have potentially performed their tasks *sequentially*, in an iterative process. Research in design has largely focused on the benefits an individual designer receives from iteration, showing that the refinement through iteration improves the design artifact [6, 10]. However, iteration on a single design also runs the risk of the designer persevering, ignoring alternatives [10], and achieving local rather than global optima [6]. Recent crowdsourcing re-

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

http://dx.doi.org/10.1145/2556288.2557158

search has also shown benefits from iterative work processes in independent tasks, e.g., handwriting recognition [30], proof-reading [4], and even chair design [44]. However, the benefits and drawbacks of sequential work processes on more interdependent tasks remains unknown.

This paper investigates the benefits and drawbacks of sequential vs. simultaneous work structures on interdependent tasks as the number of workers increases. We explore these tradeoffs using a creative, interdependent task in the context of a microtask market, Amazon's Mechanical Turk. Our results aim to contribute to theories of small group research as well as the more practical understanding of how to use crowdsourcing for creative and complex tasks [25, 27]. Specifically, we make the following contributions:

- We believe this paper is the first to explore tradeoffs in group size and coordination method for crowd workers, filling a need for research on the size of virtual teams [31], as well as extending crowd work to rigorously examine interdependent tasks.
- Our findings show that groups perform uncertain, interdependent tasks better by working sequentially rather than simultaneously, contrary to prior proposals by structural contingency theories [13].
- Through mediation analyses and a follow-up experiment we show analytically and experimentally the role of social processes (i.e., territoriality) and cognitive processes (shared mental model) in mitigating the drawbacks of the simultaneous condition.

RELATED WORK

A recent Science article noted, "if research in psychology had a Dr. Jekyll and Mr. Hyde Award, it would go to … the group as a decision-making instrument" [19], emphasizing the nuanced and often contradictory findings of group performance.

Group Size and Process Losses and Gains

Groups perform better than independent individuals on a range of tasks, though often fail to meet "a reasonable potential productivity baseline" [23] such as the group's most capable member [20, 35, 29].

Adding members to work groups can offer a variety of process gains: additional workers can offer new knowledge, cognitive stimulation and error detection [11, 20, 23, 28, 29], although the extent of the benefits depends on the task and environment.

Despite the potential benefits of group work for complex, interdependent and uncertain tasks, motivation and coordination problems can cause groups to fail, or at least fail to live up to their potential, because of *social process losses*—motivational losses and coordination problems in combining contributions. *Motivation losses* include *social loafing, sucker effects* and *evaluation apprehension* that cause people to work less hard in groups than when working by themselves. *Coordination losses* represent the variety of ways that people fail to optimally combine their work products when working in groups, such as *production blocking* when individuals cannot talk at the same time, and *thought derailment,* when group discussion interferes with one's ability to start or continue a train of thought; for reviews see [20, 23, 29, 35].

Our research examines how to maximize the benefits of additional workers, while minimizing the process losses. Two factors moderate these processes: task type and coordination mechanism.

Task Type

Performance of groups and individuals differs depending upon type of task. In tasks with objective answers, where it is easier to combine individual contributions, larger groups do better than smaller ones [23], for example, in numeric estimation tasks the central limit theory as implemented via "wisdom of the crowd" suggest more estimators lead to more accurate estimates [41]. In "Eureka" tasks, where everyone will recognize a correct answer once it is offered, larger groups are more likely to contain an individual who knows the answer. However, for problem-solving tasks, where the answer is more difficult to derive, Laughlin [28] demonstrated that groups of two and three workers performed better than individuals, as well as larger groups, potentially because of the process losses previously mentioned.

Recent examples in crowdsourcing draw on multiple task aspects. For example: iterative chair design – a creative production task, though one that is also judged on practical considerations [44]; trip planning – a task that both requires generation of ideas and plans yet adherence to global constraints [34]; and collaborative poetry translation – a creative intellectual task that requires negotiation between workers [25]. While these projects have been successful with groups, they have not focused on the effect of group size or coordination method. Additionally, there has been little work examining group size in virtual teams [31].

Coordination Mechanism

Task performance is further affected by coordination mechanism. A group of workers can be utilized in different ways, and their interactions defined by their environment. Coordination theorists have proposed that different coordination styles—pooled coordination, where the group product is simply the aggregation of individually performed tasks, sequential coordination, where one worker builds on another's output, and team coordination, which requires interaction among members—are appropriate respectively for work with increasing degrees of interdependence [39, 40].

Thus, groups working on uncertain, interdependent work, such as writing a creative document, should be best able to coordinate by working synchronously, in the 'team coordination' style described above. Synchronous work supports the mutual adjustment (seeing each others work and adjust-

ing), ad hoc arrangements, and rich, fast communication necessary for this type of task [40, 13].

However, task uncertainty changes the relationship between group size and performance. Synchronous work imposes high coordination costs, and adding workers to complex and interdependent tasks may increase process losses beyond any benefit the extra workers bring [28, 39, 40].

In this work, we argue that sequential work may mitigate those process losses, while maintaining the benefit of additional workers. There is little research on sequential work and interdependent tasks in the small group literature. We posit that sequential interaction can allow groups to coordinate the interdependent tasks without having to talk to one another; that the work artifact contains all the information needed for mutual adjustment. Unlike prior sequential coordination for assembly line work, the task and artifact are not divisible, the output is novel, and changes by one person have unpredictable effects on the next person's task. At a large-scale, this may be similar to Wikipedia, where groups of people are seemingly able to coordinate sequential editing through the artifact (although with the addition of Talk pages), though such interdependent coordination remains an open question [26]. We conduct controlled empirical experiments to test these arguments more rigorously.

To investigate optimal ways to deploy workers for uncertain and integrated production work, we propose extending prior work by studying group size in (a) virtual teams, (b) with a creative interdependent task, and (c) comparing simultaneous and sequential coordination methods. We then compare the two methods to identify processes that may aid or hinder work in different coordination methods.

EXPERIMENT 1: GROUP SIZE IN SIMULTANEOUS & SEQUENTIAL WORK

Drawing from small group research, we test groups of size 1, 2, or 3 workers simultaneously or sequentially working on a task.

Task

We are interested in complex, creative, collaborative work (i.e., work that has a high degree of uncertainty). We required a task that:

- produces a single artifact, interdependent enough that producing it requires coordination from a group of workers;
- produces a work product whose quality can be assessed;
- leaves traces, so that researchers can analyze the detailed coordination processes associated with quality.

We chose group limerick writing as the creative task. The creative and interdependent nature of writing a limerick means the work is highly uncertain: each limerick is unique, even though it conforms to a specific structure. Its components are highly interdependent, e.g., both the content and rhyme in line 5 depend upon line 1. The task of limerick writing also combines a number of attractive task elements seen in related work: *production*—creative writing task requiring generation of ideas and a final product; *discussion*—highly interdependent and may require group negotiation; and *problem-solving*—though a creative task, there are constraints such as specific rhyme and rhythm. Thus, coordination is likely desirable.

Study Design

We conducted a 2 X 2 experiment, in which groups worked using a simultaneous or sequential workflow. We treat individuals (1 worker) as a baseline condition, and compare to groups of size 2 or 3 workers. Simultaneous groups collaboratively worked online with 2 or 3 workers at once, while in sequential work we add 1 worker in each iteration, i.e., 1 worker, 1+1 workers, and 1+1+1 workers. Groups wrote and communicated using a single etherpad,[1] an online collaborative editing space (see Figure 1).

To analyze the results, we test two main effects: effect of condition (simultaneous or sequential), and number of workers, along with an interaction effect: differential effects on limerick quality, as a 2 X 2 ANOVA.

Hypothesis

As we saw in previous work, the potential benefits and losses from groups depend on the task and environment. Broadly, limericks are a creative, integrated artifact requiring an interdependent process. Prior work suggests simultaneous work is likely to be most effective, and most able to take advantage of extra workers and diverse perspectives, cognitive stimulation, and error-checking [23, 29]. However, adding workers also increases process loss due to coordination and communication issues [39]. Sequential work, where each worker works alone, does not incur those process losses, but on the other hand may incur other types of process loss, e.g., not knowing a predecessor's intention, or being left with a difficult story or rhyme.

We hypothesize that iterative work will remove the potential process loss from simultaneous work—no overhead of communication or attempting to coordinate edits to an interdependent task, but retain some of the process gains—full effort and attention of each individual, and multiple eyes correcting and refining the limerick.

Figure 1. The Etherpad collaborative text editor. A shared editing space with colored contributions on the left, and a chat area on the right.

[1] http://etherpad.org

Procedure

From the point of view of a worker, the flow of work was thus: click on the task in Mechanical Turk, be taken to a 'waiting room', perform the task, then take a survey. Participants could leave at any time. Workers were filtered to US only, greater than 95% approval rate, and could only take part in one task.

In the simultaneous condition, workers could sign up for the task labeled "Help write a limerick (a short rhyme)", and be taken to a waiting room (similar to previous real-time crowdsourcing experiments [3]). Payment was $0.70 per limerick, a wage in line with MTurk hourly rates. Workers could leave this page open in the background, and take other tasks, until two or three people were ready to take the task. The workers would then be alerted and taken to a page with instructions (described below), and shown an etherpad to write the limerick.[2] No time limit was imposed. Workers who waited 10 minutes without another worker joining made up our pool of one worker contributions. The instructions of the task explained the collaborative nature of the task (if more than one person), the assessment criteria, the theme of the limerick, and details about rhyme and meter of a limerick.

In the sequential condition, the task title and payment were changed to reflect the nature and length of the task: "Improve/edit a limerick (short rhyme)", with payment of $0.50. Workers could take the task only once. The task began straight away. The same instructions and description of the original task was given as above, with the addition of editing instructions: "We previously asked turkers to write a short rhyme—a limerick—based on the instructions and assessment criteria below. We are now asking you to edit and improve one of those limericks. You should improve the limerick as much as is needed to fulfill the criteria below. (There is no limit to the editing you can do)."

Simultaneous and sequential groups used the same collaborative etherpad interface, saw similar instructions, and the same task artifacts—the limerick itself. While groups had the same interface, the conditions could have led the groups to use the interface in different ways, but we saw no evidence of this. For example, we initially manipulated whether simultaneous workers could use a chat box or not, but no difference in results was found.[2] In the sequential condition, workers were not aware that others would edit the limerick after them, and thus were not more aware of the 'process' of the experiment [16], and did not leave, e.g., comments or to-dos for later workers.

[2] To further investigate the factors affecting interdependent work, we originally also manipulated whether the group could directly communicate using a chat box. While some artefacts are created without explicit real-time coordination (e.g., Wikipedia, shared codebases), prior work suggests direct communication is key to interdependent work. However, the results show no difference between workers allowed to directly chat or not, and we devote the rest of the paper to group size and subsequent experiments.

We are able to use data from the 1-worker condition in the simultaneous condition as our 1-worker condition in the sequential condition, as the interface and instructions are the same. Because later stages of the sequential condition depend on having the earlier stages complete, this part of the experiment was run at a later time, and subjects were thus not randomly assigned to condition. We revisit this in the Limitations section.

Measures

Measuring the quality of a creative artifact like a limerick is challenging, but Amabile [1] has suggested a product or idea is creative to the extent that expert observers agree it is creative, when making judgments independently, relative to the corpus, and on other dimensions such as technical quality.

We calculate an outcome measure defined as the average of two scores: overall quality, and technical quality (is the limerick the right number of lines, the correct rhyme structure and meter). We initially split assessment into four criteria to capture both objective and subjective aspects (technical, story coherence, creativity, and overall quality) but initial testing found all to be correlated and for the remainder of the experiment we use the two clearest features. While codebooks or extensive training can be important when criteria are unclear, Hak & Bernts [17] warn of codebooks merely codifying biases, and Hennessey & Amabile suggest independent 'creative' raters should not confer or be trained prior to judgment [18]. Other techniques such as pairwise comparison are valuable but prohibitive in scale.

Three raters independently rated all limericks blind to condition; two were undergraduate or master's students in creative writing (with additional experience in editing and judging creative pieces), along with one of this paper's authors. Quality and technical dimensions were rated on a 1—7 likert scale. The technical dimension was intended to assess objective elements of the limerick: "Consider the number of lines, the rhyme scheme, and meter." The quality dimension was intended to assess the more subjective elements of the limerick: "A holistic rating, consider the story, the coherence, the creativity or interestingness, more generally: do you like it?" We provide examples to illustrate these measures in the Results section. We also capture a number of process measures to understand the coordination process: total time spent, characters written / deleted, distribution of edits, and number of author switches.

We calculated Cronbach's alpha for the two dimensions, to test whether each dimension was independent or whether we might combine the items to create a more reliable scale. The resulting alpha was 0.87, i.e., each dimension measures a similar underlying concept. For this reason, and to simplify analysis, we combine (average) the ratings into a single rating. Agreement between the three raters (Cronbach's alpha) was 0.84, indicating high reliability.

Results & Analysis

Table 1 shows the number of groups and the combined quality score for each experimental condition. Overall 716 participants wrote a total of 431 limericks working in a total of 79 1-person groups, 185 2-person groups and 167 3-person groups. Note, although Table 1 shows 79 1-person groups as controls in both the simultaneous and sequential conditions, these are actually the same groups; groups of one can be neither simultaneous nor sequential.

The limericks below exemplify excellent, average, and poor output.

There once was a squirrel in a tree
who wanted to dance like a bee
he bumbled around
but fell to the ground
and said "dear, this task's not for me."
(Overall: 7; Technical: 6; Combined: 6.5)

There once was a wolf in the forest
who howled at the moon like a chorus
lonesome seemed he
to never taste tea
his kidneys became wasted and porous
(Overall: 3; Technical: 5; Combined: 4)

A squirrel lives, lives, lives, gives out side
And talks to a mouse, moose, mole, downtown.
the mouse lives in a house
Where he likes to eat, sleep, but be neat
he saw a cat who looked like a rat
But he still wants to play all day!
(Overall: 2; Technical: 1; Combined; 1.5)

Effect of Adding Workers in Simultaneous vs Sequential

We measure the relationship between coordination method (i.e., simultaneous vs sequential organization), number of workers, and an interaction effect: differential effects on limerick quality.

We performed a 2 X 2 ANOVA, removing the individual (1-worker) condition as a baseline for comparison. The ANOVA showed a significant effect of coordination method (simultaneous or sequential), $F(1,352) = 23.15$, $p<.0001$; no significant difference between groups with two or three workers, $F(1,352) = 1.53$, $p=.217$; and the interaction between coordination and number of workers was not significant, $F(1,352) = 0.00$, $p=.998$. This suggests that the mean quality for simultaneous and sequential conditions differ, and the effect is similar in groups of two or three workers. Larger groups (or groups with more steps) don't differ from smaller groups.

We next contrast the 1-worker conditions to the conditions wit more than 1-worker. Note we use the same set of subjects in the 1-worker condition when comparing multi-person groups in both simultaneous and sequential conditions. *Simultaneous* groups of 2 and 3 workers (M=4.08, SD=0.99) did not produce higher quality limericks than individual workers (M=4.05, SD=.93), $t(273) = .22$, $p=.827$.

Number of Workers	Simultaneous			Sequential		
	Mean	*SE*	*n*	*Mean*	*SE*	*n*
1	4.05	.10	79	4.05	.10	79
2	4.03	.09	107	4.48	.07	78
3	4.15	.11	89	4.59	.08	78
Combined	4.08	.06	275	4.37	.05	235

Table 1. Overall score of final limerick. Simultaneous workers (group sizes of 1, 2, 3 workers), and Sequential work (iterations of 1, 1+1, 1+1+1 workers).

In contrast, *sequential* groups of 2 and 3 workers (M=4.53, SD=.11) produced higher quality limericks than individual workers (M=4.05, SD=.93). That is, the improvement in work quality that resulted from working in groups of two or three rather than working individually was reliably greater when the work was organized sequentially rather than simultaneously, $t(233) = 4.53$, $p<.0001$.

The prior literature, which suggests that simultaneous coordination is necessary for interdependent, creative work [39, 40], would have led to a prediction that the greatest benefit from additional workers would have occurred when work was organized simultaneously. However, this expectation was disconfirmed in this experiment. Rather, additional workers in the sequential, iterative condition led to an increase in quality.

Although we designed the environment to minimize some known process losses, we may not have been entirely successful. We used etherpad, an editor for simultaneous work, which should eliminate one form of production blocking, since workers were able to write and edit at the same time. However, the presence of others may have led to a virtual production blocking, in which workers felt reluctant to write and especially edit a partner's work when the partner was working. Workers were anonymous (or pseudonymous, referenced by a nickname of their choosing), which should minimize evaluation apprehension; though their contributions were colored, increasing the identifiability of their edits, which should have minimized the social loafing common in groups [23]. However, we may not have fully eliminated these problems. The sequential process, on the other hand, seemed to be able to take advantage of the extra workers, allowing process gains with the extra eyes and effort. Each iterative worker worked alone, and so there was no process loss from collaboration or communication.

ANALYSIS: COMPARISON OF SEQUENTIAL AND SIMULTANEOUS

Our results showed that additional workers improved production quality when workers wrote sequentially rather than simultaneously. However, these results do not explain why sequential workers were able to increase the limerick quality while simultaneous workers were not, given that the same number of people worked on the limerick.

Above we speculated that workers in the simultaneous condition suffered more from some specific types of process losses than those in the sequential condition. To investigate why additional workers improved quality with a sequential organization of the work process, but not a simultaneous one, we looked at factors that may mediate the differential benefits from adding workers. Specifically, we conducted a mediation analysis [2] to test whether territoriality (a form of production blocking) and/or social loafing could account for why adding workers improved quality in the sequential condition but not the simultaneous one.

Possible Mediators

1. Territoriality. Previous research shows that production blocking – unwillingness or inability to contribute when working with others – can explain why adding workers does not improve work performance in real groups compared to nominal groups where workers do not interact [12]. Although the etherpad editor eliminated physical production blocking by allowing workers to simultaneously edit, the presence of co-workers in the simultaneous condition may have led to social production blocking. In the simultaneous condition, workers may have avoided editing or deleting others' work due to territoriality, in which people are uncomfortable editing another person's work. Territoriality has been noted in other collaborative authoring such as Wikipedia [38]. Territoriality could result in missed opportunities to improve the limerick, failure to fix errors, and a lack of creative tension. In contrast, in the sequential condition, workers could freely edit previous work without fear of 'stepping on someone else's toes' because the previous worker had departed and would not know about changes to their text.

To test whether territoriality, a form of social production blocking, can account for difference in quality, we measure the amount of deletions to someone else's work versus one's own work. (Note: in the sequential case this measure is mostly constant, e.g., in the case of the second worker in the sequential condition, edits are mostly to someone else's work. The mediation analysis will allow us to determine if this measure can account for output quality in either condition.)

2. Social loafing. Simultaneous workers may put forth less effort because they believe other group members will pick up the slack, or they feel their ideas are dispensable [22, 35]. However, in the sequential condition, because participants worked alone, they would be less susceptible to these sources of social loafing.

To evaluate whether increasing group size reduced effort from workers, we measure the edits per person in each condition. It is likely not all edits are equal (some changes might have more of an impact than others), but prior research (e.g., in Wikipedia [26]) has shown that number of edits correlates highly with other converging measures of effort (e.g., number of characters changed), and so we believe is a reasonable proxy for effort.

Explaining the Differential Effects of Group Size

The results of Experiment 1 suggest that groups of two or three outperform individuals when working sequentially, but not simultaneously. An alternative analysis of whether additional workers improved quality, that will allow us to simply perform a mediation analysis, is a regression analysis with coordination method (sequential vs simultaneous), number of workers (one to three), and an interaction effect: differential effects on limerick quality. The analysis shows that the experimental manipulations influenced quality: $R^2 = .06$, $F(3, 506) = 9.83$; $p<.001$. Neither coordination method (b=.08, p=.53) nor number of workers (b=.05, p=.49) alone significantly predicted quality, but the interaction did (b=.22, p=.02). Interpreting the coefficients shows that each additional worker added .22 points to the score in the *sequential* condition, but only .05 in the *simultaneous* condition. Similar to the analysis of variance combined with t-tests in Experiment 1 results, this suggests that additional workers outperformed a single worker in the sequential condition, but not the simultaneous one.

In an attempt to explain these effects, we performed two mediation analyses with the variables described earlier—territoriality and social loafing—as mediation variables. Territoriality and social loafing are not related to each other, $r = 0.014$, indicating suitability for use in mediation. We use standardized coefficients from now on, to account for different units in mediation variables. The standardized coefficient of the interaction (previously 0.22) is 0.18. The goal of the mediation analysis is to determine whether adding one of the proposed mediation variables lowers that interaction effect. We walk through one analysis, described graphically in Figure 2.

Territoriality. We previously showed that the interaction variable (the differential effect of number of workers depending on coordination method) predicted limerick quality; this is the total effect shown in Figure 2. Path *a* shows how the interaction variable is correlated with the mediator, territoriality (beta=.46). Path *b* shows how the mediator affects the limerick quality score (beta=.12).

The total effect of the interaction variable on quality can be broken into a direct effect and an indirect effect through territoriality. The direct effect of the interaction variable accounting for territoriality is the total effect (beta=.18) minus the effect through the mediator (beta=.06, calculated as product of coefficients in the paths). The effect of interaction on score drops from a total effect of 0.18 to a direct effect of 0.12 once territoriality is controlled.

A Sobel-Goodman test indicates that the mediation effect of territoriality is marginally significant, p=.07, with approximately 32% of the total effect being mediated. In sum, the evidence weakly suggests that territoriality partially mediated the interaction effect of number of workers and coordination method on output quality.

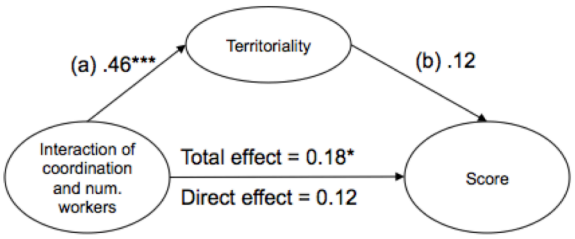

Figure 2. Mediation analysis for interaction of coordination method and number of workers. Mediating variable is territoriality (the amount of deletion of others' work). Each path is quantified with standardized regression coefficients. The direct effect of interaction on score is calculated as total effect minus effect through the mediator, which is calculated as the product of coefficients in the paths.

Social loafing. We perform a parallel analysis to test whether social loafing mediates the interaction effect of number of workers and coordination method on output quality. We first test if social loafing is correlated with an increase in workers (in sequential work: beta=.08, p=.28, and in simultaneous work: beta=-.23, p<.001). This implies that adding people increases social loafing in simultaneous substantially more than in the sequential condition. To test similar to path *a* as previous, the interaction variable is correlated with the mediator, social loafing, beta=.189. Path *b* shows how social loafing affects the limerick quality, beta=.186. The direct effect of the interaction variable accounting for social loafing is 0.15. Sobel-Goodman tests indicate this effect is not significant. This suggests that while social loafing does affect simultaneous workers more than sequential, it does not predict quality and thus cannot account for our interaction.

In summary, we proposed two reasons why adding workers to sequential work may increase output quality while adding in simultaneous work does not. The social loafing variable implies that while adding workers decreased their individual effort, doing so does not affect quality. The territoriality variable was able to partially mediate the differential effects of number of workers, demonstrating that increased edits to others' work accounted for some quality difference.

EXPERIMENT 2: IMPROVING SIMULTANEOUS WORK

Results from Experiment 1 indicate that additional sequential workers outperformed additional simultaneous workers at a creative interdependent task—limerick writing. Mediation analysis suggests that a territoriality phenomenon whereby workers are reluctant to edit another's work in the presence of that person contributed to this difference in quality.

To test this idea, we conducted a follow-up experiment to encourage people in the simultaneous condition to edit each other's work, by assigning an editor role (as well as a writer role, often used in collaborative writing [32]). Having an editor role gives a person explicit license to edit someone else's work. A writer is tasked to create an initial limerick, while the editor is asked to focus on improving that limerick. This is similar to the sequential method, but the editor is there for the creation of the limerick, and the writer stays to allow back-and-forth once the limerick is written. The assigning of specific roles may also reduce coordination and communication issues [9]. In this experiment, we hold group size constant (n=2), to focus on the effect of two non-conflicting roles. Future work may examine the effect of roles and group size, as well as multiple workers given complementary or conflicting roles. Given two simultaneous workers, we hypothesize that assigned roles will increase task performance compared to a control condition similar to prior experiments.

Condition (two workers)	Simultaneous score		
	Mean	*SE*	*n*
Control	3.70	.24	37
Roles	4.33	.15	40

Table 2. Overall score from Experiment 2: two simultaneous workers in control condition, or given assigned roles of writer and editor.

Method

A similar procedure to Experiment 1 was followed. On MTurk, workers signed up for the task labeled "Help write a limerick (a short rhyme)" and were taken to a waiting room. When two workers were ready, they were alerted and taken to a page with instructions and an etherpad. Workers were randomly assigned to a control condition (the same as in Experiment 1), or the role manipulation which added role instructions: "You and your partner have different roles: a writer and an editor. You are the writer. The writer should focus on creating new content. Your partner is the editor. The editor should focus on revising the content. These should be happening at the same time."

Experiment 2: Results & Analysis

Given the high inter-rater reliability in Experiment 1, one rater (blind to condition) rated all limericks. Descriptive statistics are presented in Table 2. An independent samples t-test was conducted to compare score in control and role conditions. There was a significant difference in the scores for control (M=3.70, SE=.24), and role (M=4.33, SE=.15) conditions; t(75)=-2.23, p=.029. These results suggest that assigning writer and editor roles in simultaneous work is able to mitigate process losses, increasing the limerick quality above that of a control condition.

DISCUSSION AND CONCLUSIONS

Online groups have great potential for creating interdependent and complex work, but coordinating members and minimizing process loss is an ongoing challenge [23, 26, 29]. In this paper we examined the effect of group size on two methods of coordination: simultaneous and sequential work. The results demonstrate that for an interdependent creative task like limerick writing, in the context of an

online environment, adding workers sequentially has more benefit than adding workers simultaneously.

The evidence suggests that this effect occurs because when people work sequentially, the initial person creates, while the second does editing. This leads the editor to feel freer to edit the author's text, i.e., less territorial. This is not happening in the simultaneous condition, where workers create and edit at the same time. These conclusions are supported by two types of evidence:

1. Mediation analyses suggest that social loafing, a key concept for process loss in groups, does not account for the difference. Although simultaneous group workers did exert less effort (number of edits) than a lone worker, the mediation analysis suggests this did not account for the difference in quality. However, editing others' work partially accounts for the difference in quality.
2. A follow-up experiment demonstrates that workers in a simultaneous condition can work better than a control condition if they are given writer/editor roles. The explicit assigning of roles lends authoritative weight to the ability to edit another's work.

Our results suggest that one can get the benefits of collaboration with fewer process losses if people work sequentially. Given the relative lack of research on sequential coordination and interdependent tasks, these results suggest iteration should be further examined. It may be the case that iteration is beneficial under some conditions but not others, for example where the artifact does not contain enough information or context for a worker to effectively contribute, and in particular other factors of the task and environment should be tested in future work.

Isolating factors to test specific hypotheses, e.g., which element of creative writing or poetry is beneficial to collaborate on – the ideation, the production, the fitting to constraints? – is one promising area for further work. Future work may also consider ways to optimize workers' attention, to focus on poor artifacts or highlight artifacts that need different types of work, e.g., a rewrite compared to light editing.

Limitations

Territoriality may explain why sequential groups were able to perform better than simultaneous groups. However, another explanation, that we were unable to test because of correlation with territoriality, is that of 'shared mental models.' Subsequent workers in the sequential condition see an entire limerick, and may have a clearer idea of the direction of the work before adding their contributions. This is not happening in the simultaneous condition, where workers create and edit at the same time.

Simultaneous and sequential experiments were not conducted at the same time, and so comparisons between them may introduce a confound. Within condition, workers were not randomly assigned: the waiting room procedure meant that our 1-person pool generally waited for 10 minutes (though during this time were able to browse to other pages, complete other tasks) without other workers joining, whereas the 2 and 3-person conditions started when a group was all present. As a result, subjects in the 1-person conditions may have been more conscientious than those in the multi-person conditions. Despite their potentially greater conscientiousness, subjects in the 1-person groups performed more poorly than those in multi-person groups. Moreover, while workers were allocated non-randomly to this condition, we use the same individual condition for both simultaneous and sequential conditions, so any pre-existing differences should apply to both. All experiments ran for multiple days (or even weeks), and some worker effects may average out.

Payment differed between conditions ($0.70 in simultaneous, and $0.50 in sequential). Prior work has shown workers expect pay to reflect the length of the task [8]. Since workers must wait for a period of time before starting in the simultaneous task, we opted to advertise the task at a higher payment. Prior work has found that increasing payment attracts more workers faster, but has no consistent effects on quality of work [e.g., 33]. Future work should further examine the relationship of task pay, uptake, and output quality.

Since we did not run both experiments concurrently, it is not a fully randomized experiment design. In the sequential condition, since later stages obviously depend on the prior stages, such a randomization is not possible. Future work may investigate how day or time of day affects participant pool (prior work has found some seasonality effects with completion rate depending on day of the week [21], but to our knowledge quality has not been examined) and attempt to control for quality effects, as well as broaden beyond paid marketplaces.

We used limericks as an example of the complex, creative work that has been undertaken in crowdsourcing recently (e.g., [7, 24, 25, 27]), and extend beyond the single task types commonly seen in small group research [29]. Future work may focus on particular aspects such as ideation or editing, as well as extend to different tasks.

Implications

We began with a question of optimal resource deployment: how might we get the most benefits from multiple workers working on a complex, interdependent task, where it is not clear what the answer is beforehand.

Prior work in organizational behavior has focused on small, synchronous groups, and suggested that for highly interdependent tasks such as limerick writing, coordination using team interdependence is the most valuable [39, 40]. Such coordination allows the fast communication and mutual adjustment necessary for such interdependent work. In contrast, while recent crowd work has examined iterative work

[44, 27, 30], there has been little examination of workers directly interacting and communicating with each other synchronously, nor comparison between the two methods.

We believe our paper is the first to explore the tradeoffs between crowd workers working together simultaneously or iteratively. We focus on limerick writing as an instance of a complex, creative task that has uncertain and interdependent task aspects. Our findings suggest that, contrary to prior work, sequential coordination is actually a more beneficial use of multiple workers. Simultaneous work was not able to capture the benefits of additional workers, while iteration was able to do so, and removed communication and coordination overheads.

It seems that rather than collaborate on an artifact directly, sequential workers were able to effectively collaborate *through* the artifact. This may be similar to a process such as Wikipedia that effectively manages to combine collaborative yet sequential editing. This paper has been able to demonstrate and compare in a more rigorous way the potential of such coordination, a process that has to date largely been ignored in the organizational behavior and small group research areas.

Further, we have demonstrated that territoriality partially accounts for the difference in quality between simultaneous and iterative work. By assigning writer/editor roles to two-person simultaneous groups, they were able to outperform control groups. With the desire for crowdsourcing services to create increasingly complex or creative work, these findings may have immediate practical implications. With interest in CHI and CSCW specifically in collaborative crowd writing, even very recently, e.g., [7, 24], these results suggest possible sources of quality increase, as well as savings in time and effort (along with traditional factors that should be further explored, such as coordination method, instructions, and planning stages).

Finally, these findings suggest that interdependent tasks requiring team interdependence is an assumption that should be challenged and explored more thoroughly, in the context of particular tasks and environments. This research appears at a time when the very notion of a team and group research paradigm is under discussion by organizational psychologists, as workers and collaboration are increasingly solely virtually mediated [15, 37, 31]. As teamwork and virtual teams change and grow, understanding how to optimally use workers could have significant impact on organizations.

ACKNOWLEDGMENTS

This work was partially supported by Carnegie Mellon's Center for the Future of Work, Bosch, Google, and NSF OCI-09-43148, IIS-0968484, IIS-1111124, IIS-1149797.

REFERENCES

1. Amabile, T.M. (1982). Social psychology of creativity: A consensual assessment technique. *J Pers Soc Psychol, 43*, 997—1013.
2. Baron, R. M. & Kenny, D. A. (1986). The moderator--mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*, *51* (6), 1173.
3. Bernstein, M. S., Brandt, J., Miller, R. C., & Karger, D. R. (2011). Crowds in two seconds: Enabling realtime crowd-powered interfaces. *Proc. UIST 2011*, 33—42.
4. Bernstein, M., Little, G., Miller, R.C., Hartmann, B., Ackerman, M., Karger, D.R., Crowell, D., & Panovich, K. Soylent: A word processor with a crowd inside. *Proc. UIST 2010*. ACM.
5. Brooks, F. (1975). *The Mythical Man Month: Essays on Software Engineering.* Addison-Wesley, Mass.
6. Buxton, B. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann.
7. Cheng, J., Kang, L. & Cosley, D. (2013). Storeys – Designing Collaborative Storytelling Interfaces. *Proc. CHI 2013 EA* (Interactivity).
8. Chilton, L., Horton, J., Miller, R., Azenkot, S. (2010) Task Search in a Human Computation Market. *KDD-HCOMP 2010*.
9. Cohen, E. G. (1994). *Designing groupwork: Strategies for the heterogeneous classroom* (2nd ed.). New York: Teachers College Press.
10. Cross, N. 2004. Expertise in design: An overview. *Design Studies*, 25, 427–441.
11. Dennis, A. R. & Valacich, J. S. (1993). Computer brainstorms: More heads are better than one. *J Appl Psychol, 78* (4), 531.
12. Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53*(3), 497-509.
13. Donaldson, L. (1999) The normal science of structural contingency theory. *Studying organizations: theory and method.* Sage, pp.51—70.
14. Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. *Admin Sci Quart*, 499—517.
15. Hackman, J. R. (2012). From causes to conditions in group research. *J Organ Behav*, *33* (3), 428—444.
16. Hackman, J. R., Brousseau, K. R., & Weiss, J. A. (1976). The interaction of task design and group performance strategies in determining group effectiveness. *Organizational Behavior and Human Performance*, *16*(2), 350—365.
17. Hak, T. & Bernts, T. (1996). Coder training: Theoretical training or practical socialization? *Qual. Sociology*, 19(2).
18. Hennessey, B.A., & Amabile, T.T. (1999). Consensual assessment. *Encyclopedia of Creativity*, 1.

19. Hertwig, R. (2012). Tapping into the Wisdom of the Crowd—with Confidence. *Science*, 336(6079), 303-304.
20. Hill, G. W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin*, 91(3), 517.
21. Ipeirotis, P. G. (2010). Analyzing the Amazon Mechanical Turk marketplace. *XRDS*, 17(2), 16—21.
22. Kerr, N. L. & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *J Pers Soc Psychol*, 44(1), 78.
23. Kerr, N. L. & Tindale, R. S. (2004). Group performance and decision making. *Annu Rev Psychol*, *55*, 623—655.
24. Kim, J., Cheng, J. & Bernstein, M. (2014). Exploring Complementary Strengths of Leaders and Crowds in Creative Collaboration. *Proc. CSCW 2014*.
25. Kittur, A. (2010). Crowdsourcing, collaboration and creativity. *XRDS*, 17(2), 22—26.
26. Kittur, A., & Kraut, R.E. Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *Proc. CSCW 2008.* ACM.
27. Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. (2011). CrowdForge: crowdsourcing complex work. *Proc. UIST 2011*.
28. Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *J Pers Soc Psychol*, *90* (4), 644.
29. Levine, J. M. & Moreland, R. L. (1990). Progress in small group research. *Annu Rev Psychol*, *41* (1), 585—634.
30. Little, G., Chilton, L. B., Goldman, M., & Miller, R. C. (2010). TurKit: human computation algorithms on mechanical turk. *Proc. UIST 2010*.
31. Martins, L. L., Gilson, L. L., & Maynard, M. T. (2004). Virtual teams: What do we know and where do we go from here? *J Manage*, *30* (6), 805—835.
32. Posner, I.R. & Baecker, R.M (1993). How People Write Together. In R.M. Baecker (ed.): *Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*. San Mateo, CA: Morgan Kaufmann, pp. 239–250.
33. Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. *Proc. ICWSM 2010*.
34. Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation analysis. *Psychol Bull*, *113*(1), 67.
35. Steiner, I. D. (1972). *Group process and productivity*. Academic Press.
36. Straus, S. G. & McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *J Appl Psychol*, *79* (1), 87.
37. Straus, S. G., Parker, A. M., Bruce, J. B., & Dembosky, J. W. (2009). The Group Matters: A Review of the Effects of Group Interaction on Processes and Outcomes in Analytic Teams. *RAND Corporation Working Paper*.
38. Thom-Santelli, J., Cosley, D., Gay, G. (2009). What's Mine is Mine: Territoriality in Collaborative Authoring. *Proc. CHI 2009*.
39. Thompson, J. D. (2003). *Organizations in action: Social science bases of administrative theory*. Transaction Pub.
40. Van de Ven, A. H., Delbecq, A. L., & Koenig Jr, R. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, 322—338.
41. Vul, E. & Pashler, H. (2008). Measuring the Crowd Within Probabilistic Representations Within Individuals. *Psychological Science*, *19* (7), 645—647.
42. Wageman, R. (1995). Interdependence and group effectiveness. *Admin Sci Quart*, 145—180.
43. Williams, K., Harkins, S. G., & Latané, B. (1981). Identifiability as a deterrant to social loafing: Two cheering experiments. *J Pers Soc Psychol*, *40* (2), 303.
44. Yu, L. & Nickerson, J. V. (2011). Cooks or cobblers?: crowd creativity through combination. *Proc. CHI 2011*.
45. Zhang, H., Law, E., Miller, R., Gajos, K., Parkes, D., & Horvitz, E. (2012). Human computation tasks with global constraints. *Proc. CHI 2012*.