Parallel Worlds: Repeated Initializations of the Same Team To Improve Team Viability

MARK E. WHITING, University of Pennsylvania & Stanford University, USA
IRENA GAO", Stanford University, USA
MICHELLE XING", Stanford University, USA
N'GODJIGUI JUNIOR DIARRASSOUBA", Texas Tech University, USA
TONYA NGUYEN", University of California, Berkeley, USA
MICHAEL S. BERNSTEIN, Stanford University, USA

A team's early interactions are influential: small behaviors cascade, driving the team either toward successful collaboration or toward fracture. Would a team be more viable if it could undo initial interactional missteps and try again? We introduce a technique that supports online and remote teams in creating multiple parallel worlds: the same team meets many times, led to believe that each convening is with a new team due to pseudonym masking while actual membership remains static. Afterward, the team moves forward with the parallel world with the highest viability by using the same pseudonyms and conversation history from that instance. In two experiments, we find that this technique improves team viability: teams that are reconvened from the highest-viability parallel world are significantly more viable than the same group meeting in a new parallel world. Our work suggests parallel worlds can help teams start off on the right foot —and stay there.

CCS Concepts: • Human-centered computing → Empirical studies in collaborative and social computing; Computer supported cooperative work.

Additional Key Words and Phrases: Team dynamics; team viability; team intervention; online teams

ACM Reference Format:
Mark E. Whiting, Irena Gao, Michelle Xing, N'godjigui Junior Diarrassouba, Tonya Nguyen, and Michael S. Bernstein. 2020. Parallel Worlds: Repeated Initializations of the Same Team To Improve Team Viability. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 67 (May 2020), 22 pages. https://doi.org/10.1145/3392877

1 INTRODUCTION

Early interactions between members of online and remote teams powerfully shape team norms and interpersonal dynamics [19, 33]. Because even teams composed of the same individuals can have unpredictable outcomes [63], teams may — by sheer bad luck —establish unproductive norms and structures, setting the team on a path to frustration and fracture. Unfortunately, once a team is on

Contributed equally to the research; alphabetical by first name.

Authors' addresses: Mark E. Whiting, mwhiting@cs.stanford.edu, University of Pennsylvania & Stanford University, Philadelphia, Pennsylvania, USA; Irena Gao, igao@stanford.edu, Stanford University, Stanford, California, USA; Michelle Xing, mxing621@stanford.edu, Stanford University, Stanford, California, USA; N'godjigui Junior Diarrassouba, ngodjigui.diarrassouba@ttu.edu, Texas Tech University, Lubbock, Texas, USA; Tonya Nguyen, tonyanguyen@berkeley.edu, University of California, Berkeley, Berkeley, California, USA; Michael S. Bernstein, msb@cs.stanford.edu, Stanford University, Stanford, California, USA.

2573-0142/2020/5-ART67 $15.00
https://doi.org/10.1145/3392877

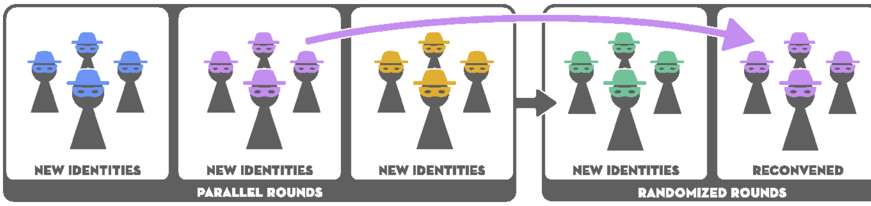

Fig. 1. Mask colors indicate pseudonym identities; each round actually consists of the same people. Three rounds initialize new parallel identities. These are followed by two rounds in random order: (1) an additional parallel round with new identities, and (2) a reconvening of the highest-viability world from the first three rounds, where participants have the same identities from the first interaction in that world. In the depicted case, the second of the first three rounds was the most viable and is reconvened with the same pseudonyms at a later point.

a downward path, it may never recover [17]. What if a fractured team could rewind time and start again on the right foot? Could a fresh start and the emergent team properties that come with it cause future collaboration to be more positive and viable? In this paper, we introduce a technique that enables computationally-mediated teams to temporarily reset their interpersonal dynamics, allowing more than one attempt to establish positive dynamics.

Our goal in this paper is to improve online and remote teams' *viability*: the capacity of a team to sustain their collaboration [26]. Intuitively, team viability reflects members' continued interest in working together. Team viability is a property of the emergent processes, inputs, and outcomes of team collaboration: it entails satisfaction, cohesion, adaptability, and a willingness to cooperate in the future [6, 11, 26]. Even highly performing teams can be non-viable due to toxic norms [18], bullying, or unequal contribution [15]. Non-viable teams will fracture [63], opting not to work together again. Online platforms increasingly convene teams for project-based work [27, 37, 41, 52, 58]. However, such teams are particularly under threat of low viability [8]: they suffer from high levels of conflict and misunderstanding leading to resentment, unwillingness to cooperate, and undermined team goals [9, 12, 30, 49, 56]. The forming and storming phases at the early part of a team's collaboration are particularly impactful to its lasting behavioral success [57] so this context is our focus.

To help online teams achieve higher initial viability, we introduce a *parallel worlds* technique: repeatedly reconvening a single team while temporarily hiding their interaction history, allowing the team to restart their collaboration many times, and then keeping the most viable world as the starting point for future interactions. We develop an online chatroom messaging system that hosts teams for multiple rounds of collaboration while hiding teams' interaction history by assigning new pseudonyms for each team member between each round. While prior studies of team reconvening brought together multiple teams per round, allowing participants to work with a previous team once, our system reconvenes the same team through all rounds [63]. While this system masks the original pseudonyms for all participants, the rewiring of pseudonyms causes each person to see that their own pseudonym remains static. The result of this one-way pseudonym masking is that team members are led to believe that they are working with new teams every round, while in reality team membership is unchanged —the original team has started their new collaboration from scratch. Consequently, the *parallel worlds* technique implemented in our online chatroom system allows for changes in participant identity through the assignment of different pseudonyms without a change in the actual team composition.

In effect, each reconvening of the team creates a parallel world in which the team collaborates on tasks with actual team composition remaining unchanged but perceived team composition being changed through the rounds. After several rounds, the team returns to the parallel world with the highest self-reported team viability, re-establishing the team using that world's pseudonyms and drawing upon its past conversation history. We imagine this technique being used with early team member interactions [41], as team members are trying out various membership configurations and determine the one most suited for success.

Prior work attempts to remediate low team viability by trying to repair the problem, for example by targeting personal behavior or repairing relationships [55]. Instead, we recognize that the initial missteps that resulted in low viability are not only challenging to repair completely [7, 17] but are also, in many cases, simply unlikely accidents that can be learned from or avoided if the team tried again without the influence of predisposed bias from team interaction history [63]. Thus, this paper proposes an alternative approach: resetting some of these early, impactful dynamics instead of repairing them at a later point in the timeline. We situate our work within the CSCW research drawing on crowd work platforms for team-based experiments [14, 27, 41, 42, 52, 53, 67]. This research argues for a growth of team-based crowdsourcing work, a model already deployed in industry (e.g, Gigster, B12.io, and Upwork). We focus especially on early phases in teamwork: forming and storming [57].

We perform two online field experiments using synchronous teams of 4–8 Amazon Mechanical Turk (MTurk) workers to test the effects of the parallel worlds technique. We confirm through a manipulation check that participants cannot accurately re-identify their recurring teammates, indicating that our system succeeds at temporarily masking individuals' identities and team history. In the first experiment ($n = 143$ in 24 teams), we initially construct three parallel worlds for each team, measuring each world's team viability through a self-report survey. We then compare the viability of the team in a reconvening of their most viable parallel world to a fresh convening in a new parallel world (Fig. 1). We find that the reconvened high-viability world retains its high viability and is significantly more viable than the median of all experienced parallel worlds (Cohen's $d = 0.87$). In the second experiment ($n = 109$ in 18 teams), we test the bounds of this effect by reconvening not just the most viable parallel world but also the least viable parallel world. The effect replicates —the reconvened high-viability world remains high viability and higher viability than the low-viability world. Additionally, we observe that the effect is asymmetric: while the high viability world retains high viability, the low-viability world instead regresses to an average level.

Our findings indicate that computationally-mediated processes can help teams start out on the right foot, and stay there. While these experiments focus on online teams in relatively brief interactions, like those found in the early stages of many remote team collaborations today, prior work suggests that such brief slices of early interactions are predictive of long-term team viability [33, 41]. This paper contributes a new class of interventions based on temporarily and selectively resetting a group's interaction history rather than trying to fix a broken history, and a series of two experiments that demonstrate its effectiveness and robustness. Moreover, the nature of online work causes collaborators to be more likely to re-engage with the same individuals across different capacities such as popular crowdsourcing platforms like MTurk and Upwork, or even within the same organization [50, 58]. We next survey related work in team fracture and viability, then describe our two experiments and discuss how these findings can be used in practice.

2 TEAM VIABILITY AND ITS ANTECEDENTS

Team viability is defined as a team's capacity for the sustainability and growth required for future performance [26]. Within organizational behavioral research, team viability is understood as an emergent construct that captures both the satisfaction of teammates with their membership and their behavioral intent to remain in their team [1]. In general, viability depends on team context, inputs, processes and outcomes that ensure that the team is composed of members who have the knowledge and skills to meet future demands, the motivation to continue making an effort toward tasks, and strategies to adapt for effective future performance [6].

Because viable teams are more adaptable and motivated, such teams will require less intervention and experience less failure in both long term and short term performance episodes [6]. However, even highly performing teams maybe non-viable [15, 18] —they may develop toxic norms, bullying, or unbalanced contribution. As a result, non-viable teams suffer from instability and eventually fail [63]. In particular, online, remote, and computer-mediated teams are noted to be rife with conflict and miscommunication [8, 10, 12, 30, 56]. While online teams are becoming increasingly prevalent, they are at great risk of low viability and fracture. In the long run, these teams likely require more intervention and support [11]. Ultimately, team viability serves as a cornerstone feature that informs the future success of repeated collaborations.

2.1 Shifts and Path Dependency in Team Viability

Shifts in viability may occur during performance episodes [5]. Performance episodes describe a distinguishable period of time over which work is evaluated, and most teams engage in multiple performance episodes over time [43]. As one performance episode ends, an understanding of whether or not a team has the capacity to sustain itself and grow emerges through the team's processes, inputs, and outcomes. Thus, every time a team repeats collaborations, they have a chance at shifting their viability for better or worse [48].

Several antecedents are known to influence team viability. One prominent one is perceived performance [23]: in essence, can team members complete the task or perform to the bare minimum? Perceptions of team efficacy are shifted during spirals, which are consecutive increases (or decreases) in perceived potency and performance over task attempts [40]. Additionally, changes in task demands will also influence team viability. This may occur because viability is predicated on team adaptability —teams that can adapt when they encounter dynamic situations are more likely to perform better in the long run, increasing a team's subsequent sustainability and satisfaction. However, team viability is distinct from team performance; even high performing teams may practice toxic norms [18], bullying, and suffer from unequal contributions [11, 15].

A second central antecedent is interpersonal friction [3, 13, 62]: interpersonal troubles undermine the viability, performance, and perceived efficacy of teams [40]. Ultimately, the affective reactions of team members moderate a strong relationship between interpersonal processes and team viability [59]. Although computer-mediated teams face the challenge of sparse social interactions devoid of in-person cues, they can achieve cohesion and viability through team members' online interactions [4]. For instance, cohesive online teams may use similar language, bursty communication patterns, and emotive terms to convey complex emotional states [22, 51]: similar language use demonstrates a sense of solidarity and shared experience, while frequent bursts of messages indicate a willingness to participate in group tasks.

A team's early interactions can set the tone for future collaboration [57]. Positive team history tends towards positive collaborations, while negative attributions and harmful interaction patterns can sow conflict. Yet prior studies also demonstrate that interaction quality and team viability is path-dependent —the same group of people can have both positive or negative interactions depending on their initial interactions [63]. This result is intriguing because it suggests that if the same team were given multiple chances to meet anew, their outcomes might vary each time. Though much research has established the historical basis of team viability and the potential antecedents for positive shifts in viability, we do not know whether multiple chances for variation in path dependency can help a team achieve high viability. Therefore, this study is guided by the following research question:

RESEARCH QUESTION. *Will teams achieve higher viability if they have multiple chances to initiate positive dynamics?*

2.2 Interventions

Interventions are designed behavior protocols that aim to relieve interpersonal friction or improve task-based problem solving processes. Most interventions in the organizational behavior literature are offline. These interventions might, for example, improve members' skills or task knowledge via feedback [7, 36, 65]. In this sense, task-based interventions have an effect of improving viability by improving performance. On the other hand, interpersonal interventions emphasize soft skills [21], clarify roles within the team [47] and improve social processes [54]. In particular, these interventions aim to diffuse built-up negative emotional tension and alleviate negative feelings between teammates [7, 36, 60, 65]. Most of these interventions repair some form of established behaviors that have developed as a result of team history.

As computer-mediated collaborations have become more prevalent in literature and in practice, human-computer interaction researchers have built a long line of system interventions aimed at improving online collaborations. These interventions span a wide scope including interpersonal feedback [20], recoloring perception of team efficacy [24], assigning leadership [29], real-time feedback [39], and membership algorithms [53, 55]. For instance, visualizations of teammates' rankings in leadership, efficacy, and participation volume effect how teammates behave during collaboration — encouraging some to speak up or let others voice their ideas [39]. In addition, researchers have used spoken repair by robots to de-escalate personal attacks and negative emotions [34].

Ultimately, it is important to recognize that the core of traditional and computer-mediated team interventions are of the same fundamental quality —they serve to correct and transform negative behaviors that have been established in the past. However, for past teams with high viability, researchers have found that teams composed of members that were familiar with one another worked faster and created a higher quality product than ad hoc teams without familiarity [28]. Rather than shifting established team processes and negative historical attributions, we build off this prior work by instrumenting team history and path dependence as a means for intervention. Our intervention allows us to introduce parallel worlds —situations in which the same team members collaborate repeatedly, while being led to believe they are working with new groups each time. With this mechanism we hypothesize:

HYPOTHESIS 1 (H1). *Reconvening the best parallel world will result in higher team viability than the median of experienced parallel worlds.*

While familiar teams carry a compositional and interpersonal history, they also accumulate collective knowledge over repeated collaborations. As a result of repeated collaboration, this team learning [35] enables the group to develop more effective collaboration strategies over time. Team learning is in part due to a transactive memory of past interactions shared among team members [61], which involves the operation of memory systems of the individuals and the processes of communication that occur within a group. Therefore, familiar teams with prior high viability are likely to improve over repeated collaboration [53]. However, accumulation of task knowledge [16] and team learning can happen even across teams [44], making members better collaborators even without a shared transactive memory or familiarity. Team learning theories would suggest that viability will improve over time, even if the team is not aware that they are working together again. Thus, we hypothesize:

HYPOTHESIS 2 (H2). *Reconvening the best parallel world will result in higher team viability than the original meeting in that best parallel world.*

The rest of this paper tests these two hypotheses. In our first study, we create a series of team reconvenings to answer: could a team achieve high viability if they had multiple chances to reset their dynamics? As a result, we aim to use team history as an instrument for a new type of intervention. Instead of fixing a broken history, we intend to leverage variations of teams' first interactions to set effective dynamics. In our second study, we aim to understand if team learning occurs when high-viability and low-viability parallel worlds reconvene. Specifically, would teams with high team viability improve their initial viability further? Would low-viability teams inherit their negative interaction patterns or improve over time?

3 STUDY 1: RECONVENING HIGH-VIABILITY WORLDS

In this study, we ask: if we allow a team multiple first convenings under different pseudonyms, temporarily hiding the collaboration's history between these convenings and then persisting the memory of only the best convening, would the resulting team be more viable? To answer this question, we implement an intervention in which an online team works together three times in succession, each time believing they are a newly convened team, giving the team three chances at an initial interaction. We refer to each of these meetings as a *parallel world*: the team has not forgotten their prior interaction, but because they believe each world to be a new set of people, they do not carry over their transactive memory and history to this new meeting. From these three parallel worlds, we carry on the memory of only the most viable meeting by calling upon a specific team memory.

3.1 Method

It is, of course, impossible — or at least impractical without Infinity Stones — to create a multiverse of parallel worlds or to erase a team's transactive memory. Consequently, we seek an intervention that models a similar effect. We cannot remove task learning over time, nor team learning that would occur as people develop better strategies for collaboration when a given situation arises. Nevertheless, we believe that it is possible to temporarily mask the transactive memory surrounding participants' past conceptions of the team overall and of fellow team members, enabling a team to meet again under the assumption that they are working with a new team. To achieve this effect, we develop a pseudonym masking system to rewire actual identities with reassigned identities.

3.1.1 Pseudonym masking. We draw on a method developed to study team fracture [63], which performs a one-way masking of pseudonyms and rotates pseudonyms between each team meeting. In this method, participants arrive to an online, synchronous discussion platform (Fig. 2) and participate in a series of rounds of a short task. Participants operate under system-assigned pseudonyms of a concatenated random sample of an adjective and animal (e.g., conventionalPanda, littleHorse, culturedBear). New pseudonyms may be assigned in a round, such that even if the same partners reconvened from a previous round, they would appear as new collaborators to each other. However, for any given participant, the system makes their view such that their own personal pseudonym appears consistent from round to round regardless of whether it has changed in others' views. The system automatically replaces pseudonyms and near-misspellings of pseudonyms in each participant's view to ensure that participants have a consistent view: that one participant always sees themselves as conventionalPanda, for instance, and sees others referring to them as conventionalPanda, even if others saw them as engagedLemur in the last round and see them as niceDonkey and refer to them as niceDonkey in the current round. We call this one-way pseudonym *masking*. This experimental scaffold allows researchers to temporarily erase collaboration history by changing participants' perceived identities.

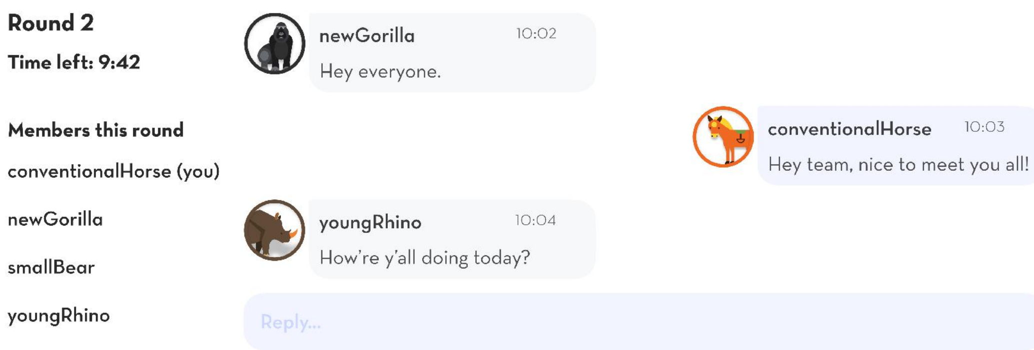

Fig. 2. The chat interface displays the current round number, a countdown timer for the round, a list of the team members' pseudonyms, and the chat log itself.

Prior work [63] demonstrated several results about pseudonym based methods for repeated team interaction that are useful to us in our design. First, a manipulation check found that, across many different task types, masking ensures that participants are unaware of which teams are reconvened and which teams are entirely new. In particular, participants asked to re-identify which pair of their teams were the same team reconvened were only able to re-identify correctly at a random chance rate. Second, the prior study demonstrated that creative tasks such as ad writing result in high variance in viability across reconvened teams. Third, the prior study confirmed that neither task learning nor team learning had an effect across rounds of the experiment. Finally, the prior study analyzed discourse patterns such as sentiment, Gini coefficient, and turn taking in teams' chatlogs and found no significant differences across time in the experiment, between conditions, or between high and low viability teams. We leverage these results in our design, enabling us to be confident in the tasks, experimental manipulation, and other learning measures in our experiments. We also reproduce many of these measures to check if our results reinforce the previous findings.

In this paper, we build upon the prior method so that we can re-establish a transactive memory by restoring particular identities. However, instead of always assigning new pseudonyms, we change the experimental design so that the system can reuse old sets of pseudonyms from another round and display the previous conversation history from that set of pseudonyms. Consequently, participants possess and are aware of the same identities drawn from a previous teamwork interaction.

In addition, we modify the prior method to operate with a single team rather than many parallel teams, enabling experiments with larger team sizes than had previously been used. The previous method required recruiting many simultaneous participants and swapping participants into different teams per round. This complex assignment guaranteed participants meeting a new set of people whenever it was not a reconvening round. Now, the new and simpler single-team experimental design allows for the analysis of larger teams. To guarantee that each person could work with an entirely new team each time, the previous method required N^2 simultaneous participants for teams of size N, which placed practical constraints on team size. Instead, we adjust the approach to recruit one team and repeatedly reconvene just that team. This streamlined design enables recruitment of larger teams of 6–8 participants (which would have required 36–64 simultaneous people in the previous approach), reconvening them every round. Additionally, it affords flexibility in choosing multiple reconvening rounds and isolates team history as the main target studied.

In the experiment's first three rounds, we create three successive parallel worlds for the team. Each round, the team is re-masked such that they see new pseudonyms that suggest a first interaction with a new team. Between rounds, participants score their most recent team by completing a Likert 14-question viability survey about that team (e.g., “Most of the members of this team would welcome the opportunity to work as a group again in the future”) [11]. Question scores were summed to comprise an individual team viability scale with a minimum of 14 points and a maximum of 70 points, with more points indicating higher individual team viability. We averaged all team members’ viability scores to calculate our main dependent variable, the team viability score. Team viability thus reflected how positively participants remembered each interaction.

In our new design, two final rounds followed the three initial ones (Fig. 1): one standard and one experimental round. The order of the standard and experimental rounds was randomized. In the standard round, the team is again masked, and members experience a supposedly-new team just as in the initial three rounds. In the reconvened condition round, however, the team is reconvened and unmasked. Immediately prior to starting this round, the platform notifies members that they are being reconvened with a previous team. The platform then compares the team viability scores of the first three rounds and selects the highest-scoring round. When participants begin this unmasked task, they see the same pseudonyms as in the highest-scoring round, and the team’s previous chat log is displayed at the top of the screen. The best of the three parallel universes is thus selected to persist into a second interaction. We measure the viability of these final two rounds using the same method as the initial three.

3.1.2 Participants. Following precedent in team-based research within CSCW, we recruited participants from Amazon Mechanical Turk [27]. We filtered workers to those who were located in the United States (to fulfill the language requirement) and had completed at least 100 tasks. Each round, teams worked on an ad-writing task drawn from prior work [14, 53, 63]. We incentivized team performance by telling teams that we would run the ads and provide bonus payments to the teams with the highest performing advertisements. In this task, teams communicate via the platform’s chatroom to generate and submit one 30-character slogan for a Kickstarter campaign. This task is a creative generative task as specified by McGrath’s circumplex [45], chosen to be in line with the goals of evaluating this system for forming and storming phases of team collaboration [57]. Teams had 10 minutes to complete the task and 5 minutes to individually answer the viability survey. The total study lasted 75 minutes. Workers were bonused to a payment rate of $15/hour for full participation through the end of the experiment, matching fair work standards [64].

To account for participants leaving the system during the study, our system removes participants who have not typed any message during a round. We warned participants of this need for participation at the start of every round. We discarded data where a team had fewer than four active participants at any point. This ensured that teams had robust team histories to reference. In addition, we only analyzed data where team membership had been static across the final standard round, reconvened round, and the initial highest-viability round. This check protects against teams rating viability of different membership compositions.

3.1.3 Manipulation Check. It is a key assumption of this design that most participants not recognize that they are working with the same team across all rounds, and thus do not associate previous team histories with their present round, unless the system explicitly reveals this information. Prior work demonstrated that this masking-based experimental design successfully prevents participants from recognizing this [63]; however, to test this in our context, we also perform a manipulation check at the end of the experiment. Intuitively, if information is bleeding across rounds, participants should be able to match pseudonyms between a pair of rounds — to identify that two pseudonyms represent the same person. So, when debriefing the study and sharing that the rounds were in fact the same people, we select a random two of the initial three rounds, present the pseudonym of a teammate from one of the rounds, and ask the participant to match the corresponding pseudonym in the other

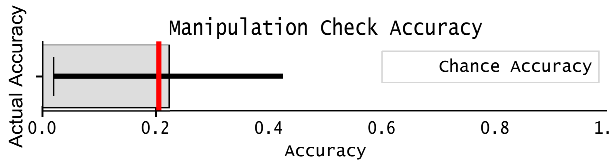

Fig. 3. In study 1, participants responded to the manipulation check question with chance accuracy, indicating that they were unable to identify repeated collaborators, and that the manipulation was successful.

round. Participants selected the corresponding name from a dropdown that auto-populated with the teammate pseudonyms of that round. Chat transcripts of the two rounds were not displayed. A roster listed each team and its members' pseudonyms. The surveys were randomized such that different participants were asked about different teammates in different rounds. If a participant worked in a team of N members, they had a $\frac{1}{N-1}$ chance of guessing the correct pseudonym randomly. A team's chance level is the average of each member's random chance rate, since members may have been shown a different pairing. A team's manipulation check accuracy is the percentage of team members who answered correctly. We compare these two values in our manipulation check, considering the manipulation to have failed if participants guess correctly at a rate substantially above random chance. Additionally, chat logs were also manually reviewed. In 3 situations, teams that explicitly recognized each other across rounds were excluded from analysis.

3.2 Results

N = 143 participants completed all five rounds of the experiment, across 24 teams. Teams had 4–8 members. The interquartile range of team viability was 11.37 points (37.73–49.10) and the average difference between highest and lowest viability rounds was 10.1.

Participants' accuracy in identifying prior collaborators and correctly answering the manipulation check question ($\mu = 0.22, \sigma = 0.20$) was not significantly different from chance accuracy ($\mu = 0.21, \sigma = 0.07$): $t(23) = 0.49, p = 0.63$ (Fig. 3). As a consequence, we consider the experiment to have successfully temporarily hidden teams' interaction history, at least at a level that they could not reconstruct who is who.

3.2.1 Reconvening a high-viability world results in high-viability collaborations.

Reconvened rounds ($\mu = 47, \sigma = 4.4$) exhibited significantly higher viability than median rounds ($\mu = 43, \sigma = 4.0$) via a paired t-test: $t(23) = 4.25, p = 0.0003$ (Fig. 4). On average, the reconvened round's viability score was 3.6 ($\sigma = 4.1$) points higher than the median round's score. Cohen's d = 0.87, indicating an effect size of over three quarters a standard deviation improvement in team viability. This result indicates that reconvening a high-viability parallel world leads to higher viability than other parallel worlds of the same people, supporting H1.

The data thus far suggest that the reconvened round significantly outperforms the median of other experienced rounds, supporting H1. We now consider how the reconvened round —the second convening of the best parallel world —compares to the initial convening of that world ($\mu = 46, \sigma = 4.4$) (Fig. 4). These two instances exhibit highly similar viabilities: the mean difference is 0.23 points, not significant: $t(23) = 0.23, p = 0.82, d = 0.05$. These results do not support H2, instead suggesting that the best parallel world maintains the same high viability score as in the first interaction.

3.2.2 Qualitative analysis: reconvened teams recognize and build on positive attributions.

We performed an inductive analysis of the team chat logs, looking for themes that arose in how teams

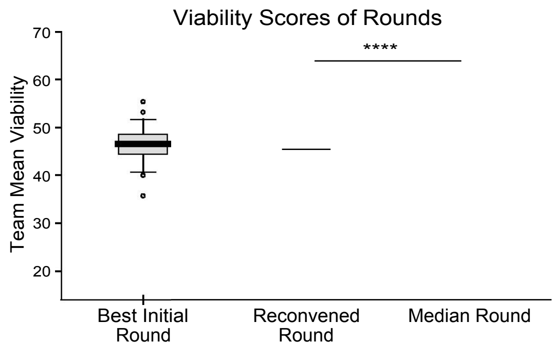

Fig. 4. Reconvened rounds exhibited significantly higher viability than median rounds, and equally high viability as the initial round it reconvened.

interacted with each other. Teams recognized and expressed positive reactions when they reconvened with high-viability worlds, as compared to what they believed to be a new team.

A substantial minority (about 20%) of all reconvened teams explicitly mentioned to each other that they were meeting again. For example:

> conventionalOrangutan: Hello!
> culturedPanda: Hi again
> spryGoat: Hey I missed you guys
> conventionalOrangutan: very sweet
> spryGoat: Best Group Hands Down

At the beginning of their reconvening, this team greeted each other with high spirits and an inside joke, referencing their previous task:

> conventionalCow: hello
> newHippo: hello again
> nicePanda: hello everyone
> likelyRabbit: funny seeing you all here
> conventionalCow: Hello cool stoolers
> newHippo: lol
> nicePanda: lol

Reinforcing the finding that reconvened high-viability worlds produced high-viability collaborations, many reconvened teams ended their second interaction on a positive note:

> conventionalRhino: bye
> likelyHorse: bye bye for now
> likelyHorse: :-D
> conventionalRabbit: Bye guys thank you!
> spryGiraffe: good working with everybody again
> likelyHorse: same here spryGiraffe

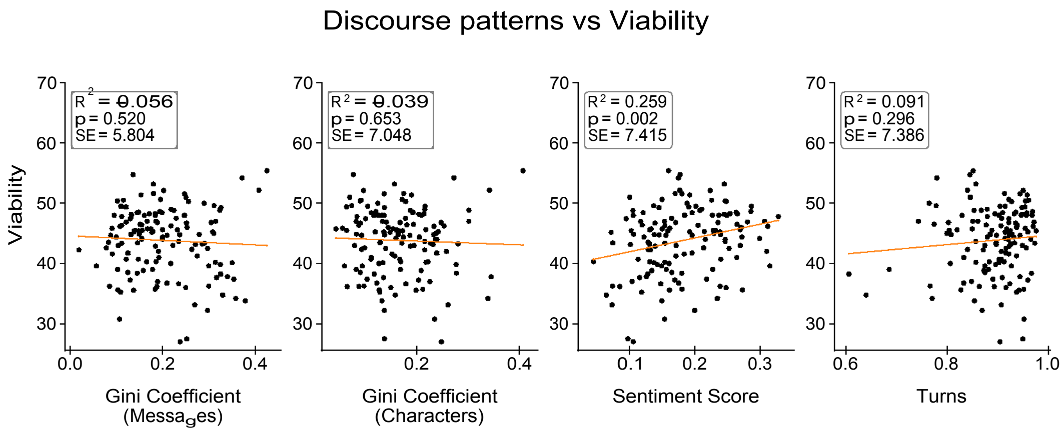

Fig. 5. None of the discourse patterns were significantly correlated with viability.

In summary, Study 1 suggests that a parallel worlds technique can successfully increase team viability for online and remote teams (H1 supported). While the teams do not further improve in viability when they reconvene (H2 not supported), they maintain their prior levels of viability.

3.3 Testing alternative explanations

We investigated four discourse patterns based on prior work [63] to test alternative explanations for the changes in viability. We computed Gini coefficients of the number of messages per chat member and the average length of messages per chat member, as well as the average sentiment, and turn-taking patterns across three conditions —the first three rounds, the median round and the reconvened round.

We calculated 2 different Gini coefficients to represent the internal consistency of the chat members' participation. The first Gini coefficient evaluates such consistency in terms of the number of messages from each participant and the second Gini coefficient evaluates it in terms of the average length of each member messages. A Gini coefficient of 1 indicates perfect inequality, which could be interpreted as a chat being dominated by only one member (either in terms of number of messages or of message length), whereas a Gini coefficient of 0 indicates a perfect distribution of member participation.

We measured the average sentiment in the chats using the Python `nltk` package's `VADER` lexicon-based sentiment analysis tool [32]. Finally, we measured turn-taking in a given chat by first counting the number of times a message had a different author than the previous message, and then divided the count by the number of messages in a chat, thereby giving a percentage. Fig. 5 shows scatter plots of viability against the different discourse patterns.

We then performed a linear regression on viability using the different discourse patterns' measures and the condition (First three rounds vs median vs reconvened), reported in Table 1. The results show that none of the discourse patterns are significantly associated with viability, which agrees with prior work. The results also confirm H1 as they show that the relationship between the condition (first three rounds vs median vs reconvened) and the viability is statistically significant.

4 STUDY 2: WHICH WORLD DO WE RECONVENE?

Our results from Study 1 indicate that reconvening a team from the best parallel world results in a more viable collaboration than the same group meeting in a new parallel world. An open question

	Dependent variable:
	Viability
Gini Coefficient (Messages)	12.143
	(7.777)
Gini Coefficient (Characters)	−7.398
	(8.433)
Turns	16.806*
	(7.785)
Sentiment	6.879
	(9.241)
Median Round	−0.040
	(1.157)
Reconvened Round	4.010**
	(1.200)
Constant	25.362**
	(7.827)
Observations	81
R^2	0.218
Adjusted R^2	0.154
Residual Std. Error	4.242 (df = 74)
F Statistic	3.431** (df = 6; 74)
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 1. Result of linear regression on viability showing lack of relationship between discourse patterns and viability across conditions (First three initial rounds, median round and the reconvened round)

remains: what caused that improvement? According to transactive memory theory, positive team history predicts positive future interactions [28, 61] —thus, the higher viability of our reconvened round is because it inherits the transactive memory of positive experiences, and reciprocally, that reconvening a lower-viability world would reproduce that lower viability. However, the reconvened round was also the only round that was reconvened in Study 1. While all other sets of pseudonyms interacted only once, the experimental set interacted twice. The literature on team familiarity would suggest that higher viability might be due simply to the team getting to know each other better through repeated interaction [28], and not because the history was positive. This theory would suggest that all parallel worlds would experience high viability when they reconvene.

With Study 2, we seek to disambiguate between these two mechanisms, and to replicate our original result to test its robustness. In this study, in addition to reconvening the best parallel world, we also reconvene each team's worst parallel world. Thus, two different parallel worlds

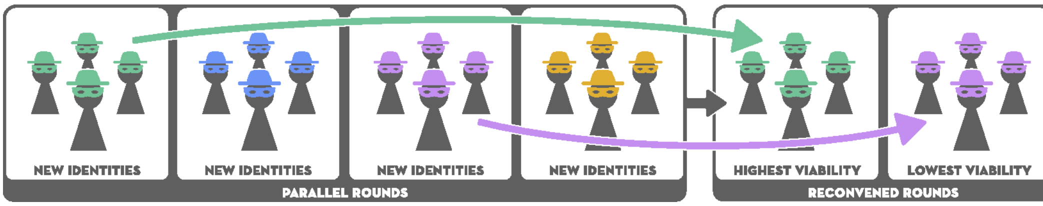

Fig. 6. Four rounds of fresh parallel worlds are followed by two experimental rounds in which the highest-viability and the lowest-viability worlds from the first rounds are reconvened with their same respective identities. The experimental rounds are counterbalanced to account for order effects. In this depiction, in the initial rounds, the first round featured the highest viability and the third round featured the lowest, so these two are reconvened.

are repeated: one the world rated most viable, and one the world rated least viable. If transactive memory is paramount, then the most viable parallel world should remain the most viable team on the second interaction and the least viable parallel world would remain the least viable when reconvened:

HYPOTHESIS 3 (H3). *Reconvening the most viable parallel world will result higher team viability than reconvening the least viable world.*

On the other hand, if team familiarity is responsible for the increase in viability, then both the best and worst configurations should improve on their second interaction:

HYPOTHESIS 4 (H4). *Reconvening the most viable parallel world will result in no difference compared to reconvening the least viable parallel world.*

4.1 Method

Our experimental design in Study 2 is similar to Study 1. Again the study consisted of two parts. First, a team of eight members underwent four initial parallel worlds across four rounds, each round followed by a team viability survey. The number of initial rounds increased by one from the first study, because the manipulation check requires asking about two initial rounds that were not later reconvened, and this study is now reconvening two of the initial rounds, requiring that we add one more round so that there are two non-reconvened rounds.

The four initial rounds are followed by two reconvened rounds, each representing a treatment conditions (Fig. 6). The first condition is the highest-viability parallel world, matching Study 1: the parallel world with the highest team viability. The system prompts participants that they are working with a previous team as before, reinstates the old pseudonyms, and the historical chat is displayed at top. The second condition is the lowest-viability world, this time reconvening the pseudonyms and chat history of the round with the lowest team viability score. Participants were not cued towards a reason for a particular reconvening; no team was labeled best or worst. The order of these two reconvened rounds was randomized. There was no additional round (masked, new parallel world) in this study.

Participants were again drawn from Amazon Mechanical Turk to complete the same ad-writing task with the same time constraints. The total study lasted 90 minutes, and workers were paid at a rate of $15/hour for full participation through the end of the experiment [64]. The same requirements for inclusion and analysis applied as in Study 1, with the addition that team membership now needed to be static, with no dropouts, across four worlds: the initial highest-viability world, the

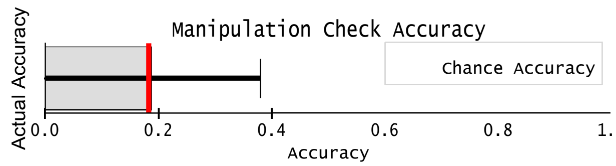

Fig. 7. In study 2, participants again responded to the manipulation check question with chance accuracy, indicating that they were unable to identify repeated collaborators, and that the manipulation was successful at masking prior history.

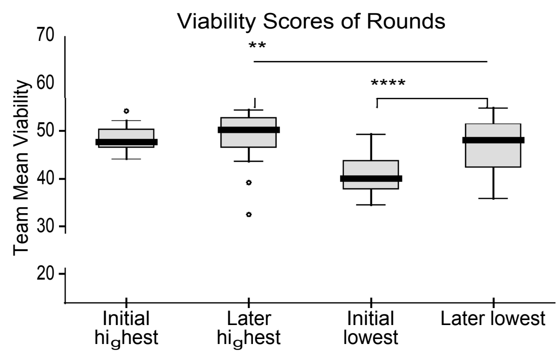

Fig. 8. In study 2, when reconvening the team from the initial rounds with highest mean viability, their viability was maintained and saw a slight increase. Comparing the team with the lowest mean viability with its respective reconvened round, a substantial increase in mean viability is apparent. However, the reconvened low mean viability parallel worlds do not outperform either the initial or reconvened high viability ones.

reconvened highest-viability world, the initial lowest-viability world, and the reconvened lowest-viability world. The minimum number of members allowed was four, so teams ranged from 4–8 members.

4.2 Results

In Study 2, $N = 109$ participants participated in all six rounds of the experiment across 18 teams. Teams had between 4–8 members. The interquartile range of team viability scores was 12.62 points (39.06–51.68).

The manipulation check again suggested that the manipulation worked. Answer accuracy (μ = 0.19, σ = 0.19) was not significantly different from chance accuracy (μ = 0.18, σ = 0.05); $t(17)$ = 0.09, p = 0.93, d = 0.02 (Fig. 7). As in Study 1, participants were not able to match prior teammates at a rate higher than chance, indicating that identity signals are not accidentally communicated between parallel worlds.

4.2.1 Reconvened high-viability worlds outperform reconvened low-viability worlds. We first replicate our test of H2. In Study 1, we saw that the first and second instances of the best parallel world showed no significant difference in viability. Our data for Study 2 replicates this result. The mean viability of the initial interaction of the best parallel world (μ = 48, σ = 2.5) and the mean viability of that world reconvened (μ = 49, σ = 5.4) saw an increase of only 0.37 viability points, which is not significantly significant ($t(17)$ = 0.37, p = 0.72, d = 0.09, Fig. 8).

With respect to the lowest-viability worlds: the viability of the worst parallel world ($\mu = 41, \sigma = 4.0$) saw an increase of 5.7 points when it was reconvened ($\mu = 47, \sigma = 5.7$) — a significant difference with a paired t-test ($t(17) = 4.90, p = 0.0001, d = 1.16$, Fig. 8). So, low-viability worlds recovered viability when they were reconvened, a trend not seen when comparing reconvened worlds of high viability.

Comparing the mean viability of the reconvening of the highest-viability parallel world and mean viability of the reconvening of the lowest-viability parallel world, the reconvening of the lowest-viability world stays lower by 2.2 points (Fig. 8). A paired t-test of these two conditions confirms a significant difference ($t(17) = 2.2, p = 0.05, d = 0.48$), and a moderate effect size. This result suggests that even when reconvened, the worst world is still significantly less viable than the best one, confirming H3 and not supporting H4. Though the lowest-viability world improves on its second attempt, it does not reach the same level as the reconvened highest-viability world.

Reconvened low-viability worlds improve over time, but they still under perform compared to reconvened high-viability worlds. Improvements in low-viability worlds can be attributed to team learning and accumulated task knowledge. For instance, team members began re-sending links in the chat:

> likelyCow: here's the link, in case it gets buried https://www.kickstarter.com/projects/letbco/letb-color-take-a-look-at-time-in-different-ways

However, although team members learned how to perform the task more efficiently, many members remain unsatisfied with their collaboration. After completing a reconvening, one team member expressed their dissatisfaction:

> newDonkey: I feel like my slogan was better that what was agreed on, but that I let the others have the one they wanted. I don't feel like I voiced my opinions loudly enough. When you're part of a team, when do you let it go for the good of the team? That's my struggle. I think I could have come up with something better on my own, but for the sake of the team, I backed down

Despite working with the same team across all parallel worlds, first impressions from the initial low-viability world lasted:

> newCow: Great group effort. Different vibe from this group, but overall it was a pleasant group.
> culturedGoat: This team was a little different. I still think that the capacity for success in working together is there but there was one person on the team that basically seemed to want to get the task done as quickly as possible regardless of the quality of the idea. Don't get me wrong, the idea was fine but I think with a little more thought put into it, it could have been better.

Similar to reconvening of the highest-viability world seen in Study 1, a qualitative analysis of the reconvened interactions of the lowest-viability world showed an overall positive effect.

> newDonkey: Hello
> culturedOrangutan: hello again
> snappyRhino: Hello again old team hehe
> youngGoat: hello again comrades

5 DISCUSSION

This paper introduces a method enabling online and remote teams multiple chances to start off on the right foot. The method convenes the same team for a short creative task multiple times under the illusion of a new convening each time, thus creating a “parallel world”, and selects the highest-viability world. Our results suggest that this parallel worlds method can be an effective technique for building viable teams online — it assumes that early interactions are both consequential and somewhat random [63], so multiple attempts can help. Moreover, we suggest that this technique is most effective for scenarios when a team is being convened for the first time: many organizations regularly convene project-based work —software consultancy, film crews, incident response teams, management consulting, or health care. Reconvening high-viability parallel worlds maintains team viability across interactions. Meanwhile, low-viability parallel worlds improve when reconvened, but not immediately to a viability level above that of a random restart. In this section, we reflect on the broader design space of our approach, its implications, and the limitations of our study.

5.1 Implications

In Study 2, we explored the theory behind the high viability observed in Study 1. We considered several possible phenomena at play: transactive memory recreating old norms and structures (H3), or team familiarity leading to improved viability.

Our results lead us to support H3 and discount H4. We did find that the highest-viability parallel world significantly outperformed the lowest-viability parallel world. However, the effect appears to be asymmetric, lending some partial credence to H4: the low-viability world did improve in its second iteration, but the high-viability world did not. Either there is a ceiling effect that prevents the high-viability worlds from doing better than already achieved, or there is aright-censored regression to the mean, where the low-viability parallel worlds benefit from trying again and the high-viability parallel worlds manage to avoid regressing to an average viability score.

We observed that persisting viable histories can have drastic behavioral effects in some cases. In a Study 1 pilot, we observed one bad actor who, though serious in the first round, consistently trolled their “new” masked teammates in rounds 2 and 3. After the second parallel world, one teammate mentioned in a post-study survey,

> newDonkey: the other members were helpful, but one member kind of dragged down the teams morale.

After the third parallel world, two other teammates reported,

> spryFox: One of the team members was fooling around and giving stupid answers.
> snappyHorse: I would work with 2 of them again and not mind, but there is one person in there that keeps making nasty comments

However, when the first parallel world reconvened, the bad actor who trolled for two subsequent parallel worlds changed behavior to match the positive initial round. In posthoc surveys, all reviews were positive. After reconvening, this team’s viability skyrocketed by 11% from its initial interaction.

> newDonkey: Everyone came up with great ideas in this round with no issues..
> culturedGorilla: Awesome team!
> spryFox: This team works well together. The ideas flow properly. There is little to no resistance from team members.
> snappyHorse: everyone was so nice and well spoken!

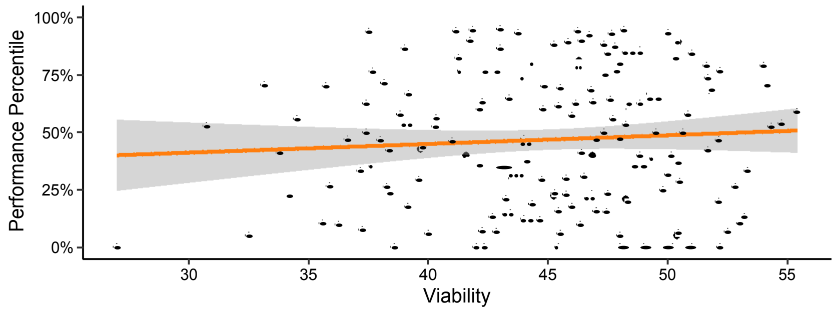

Fig. 9. Viability was not significantly correlated with a performance measure: normalized click-through rate

We cannot yet fully understand the forces that drove the team norms in different directions each time, but it is clear from examples such as this that the trolling outcome was not a foregone conclusion.

5.1.1 Performance and viability. To examine the relationship between viability and performance in our dataset, we ran the advertisements our participants created on Google Ads in a six day campaign. We smoothed the raw click-through rate (CTR) data with Laplace's Rule of Succession [38, 66], because some ads received no clicks, and normalized within each kickstarter product, because some products were much more successful than others. Our results suggest that viability is not a significant predictor of performance ($R^2 = 0.005$) in our dataset (Figure 9). The slight positive slope agrees with prior evaluation of the relationship between team viability and performance [2], however, the generally noisy nature of the ads' performance mirrors our prior findings [63]. Further, the underlying relationship was not significantly different due to teams' history.

5.2 Design Implications

As the Internet presents potential for more collaborative work [50], we anticipate that this intervention could be used to allow online teams to quickly find and set the best starting team dynamics. Like team dating [41], this intervention uses short creative tasks and a few initial rounds to output a high-viability configuration, but when accumulating many teams' worth of participants is cost-prohibitive, this intervention can be used to improve viability for one fixed group of people. Implementation of parallel worlds requires the addition of a masking mechanism and the design of the initial tasks.

Although the parallel worlds technique is generally effective for maximizing a team's potential viability, the technique should be used with care. Parallel worlds is intended to serve as a means for well-intentioned teammates to establish the emotional states and interaction patterns that are preferable for their team's success. At the level of the organization, there is the risk that the intervention is used unfairly against some of the teammates due to their behavior in one of the worlds. On the individual level, if bad actors are present, the intervention runs the risk of erasing memories of harassment, which are common to everyday, online trolls [10]. In this case, manipulated team memory may delay a team's recognition of a repeat harasser. For deployments and future work, we suggest that researchers build safety mechanisms into parallel worlds that protect parallel worlds' participants from blindly participating in unsafe online environments. One way this can be achieved is by allowing participants to report the pseudonyms of problematic individuals. Harassers could be removed from the team in future rounds. Or, if they must be retained, the system could warn the reporting individual. This flagging step would serve to prevent participants from blindly entering unsafe online environments.

In addition, outside of experimental contexts with mandatory debriefs, it is important to consider the nature of informed consent in this environment. If team members are never told of the multiple iterations, this removes autonomy from them. One approach would be to debrief after the parallel worlds exercise, allowing the team to discuss their learning's —ideally to develop norms that sustain the good interactions and avoid the bad ones. A second approach would be to draw on prior work [63] to include both repeat teams and new teams in the rounds. In this case, it would be possible to fully inform everyone of the manipulation and gather agreement that they not subvert the masking, resulting in team members willingly not knowing whether two teams were actually the same people. Nevertheless, the broader findings about team behavior may be useful in practice when companies are considering their team building and initial team forming policies.

5.3 Limitations

Our context limits our generalizability in organizational collaborations, where participants may have prior relationships and side channels may be utilized. Consequently, we suggest that this technique be is best for scenarios when a team is being convened for the first time —known as the forming and storming stages of a team's life cycle [57]. In short, this technique would best serve as a early stage norm and team forming activity to engage in membership try-outs and ensure positive viability.

Our study was performed in a short term, online, experimental context. This context limits our generalizability. Moreover, we cannot guarantee that these viability trends persist over a long period. A longer-term study would enable us to test whether these viability trends extend over longer periods of time.

Additionally, our study participants were recruited from Amazon Mechanical Turk to form ad hoc teams in a controlled setting. Although some online teams now work in an ad hoc fashion, the study does not necessarily generalize to team members who have more of a well-established history. Our controlled setting allowed us to have a particular manipulation check to ensure that the team members were not aware that they were working together in successive interactions. For teams with a well-established history, the team history may be so strong that a different parallel world implementation (and manipulation check thereof) might be required.

Finally, we experimented on one specific type of task: a creative, generative task [14, 53]. Many real-world teams work on creative tasks, and prior work has found that teams working on creative tasks are more path dependent than other tasks. Specifically, creative tasks result in inconsistent team outcomes with the same team composition [63]. However, other tasks that have a more consistent outcome may yield different results across parallel worlds. Future work should investigate the impact of the parallel worlds technique given other task types from the McGrath task circumplex [45]. Such work can provide more insight into variations between the interaction processes [46] and emergent properties that arise from different task types.

In this study, we presented reconvened teams neutrally, without hinting as to whether the team was reconvened for being a best or worst parallel world. In earlier pilots, we conducted a version of Study 1, where reconvened teams from the best parallel world were told that the configuration was previously most viable of the initial rounds. Although preliminary, the results suggested that teams signalled with this message more strictly *maintained* viability levels from their initial interaction, while teams without this signaling more freely *improved* their viability. These results agree with previous work on a team's sense of potency in agreement with previous studies that ascribed the group's actual performance (measured here by viability) to the group's sense of potency [25], where potency is the collective belief of the group in its effectiveness. Teams that were signalled may have experienced stronger belief in potency and thus did not change behavior. Further work can confirm these preliminary results and study potency's effect on reconvened worst parallel worlds.

6 CONCLUSION

In this article, we introduce a parallel world intervention to support high viability collaborations for online and remote teams. We examined how and to what extent the manipulation of transactive memory through selective reestablishment of specific team interactions impacts a team's present viability during a team-based creative task. We find that upon reconvening the highest-viability parallel world, the second interaction achieves a viability consistent with its first interaction and higher than that of a fresh start. Furthermore, when the lowest-viability parallel world is reconvened, team viability is remediated but does not improve past the level of a fresh start. Thus, by leveraging the large variation in dynamics for a single team and masking to erase unfavorable past interactions, we help teams find their best configuration and stay there even if the team had originally started off on the wrong foot.

This research offers opportunities for online teams to go beyond being there [31] — to gain opportunities that traditional, in-person teams lack. By helping teams sidestep accidental false starts early rather than pursuing a sometimes quixotic quest to undo them, we hope that this approach can help promote more pro-social teamwork and enjoyable collaborations.

ACKNOWLEDGMENTS

We are extremely grateful to Laura Fiuza, J.D. Zamfirescu-Pereira, Yu Jin Lee, and Stanford CSLI for their support on this work, and to the thousands of Amazon Mechanical Turk workers for their participation in our pilots and study. This project was supported by the Stanford Data Science Initiative, RISE Thailand Consortium, the Hasso Plattner Design Thinking Research Program and the Office of Naval Research (N00014-16-1-2894).

REFERENCES

[1] Prasad Balkundi and David A. Harrison. 2006. Ties, Leaders, and Time in Teams: Strong Inference about Network Structure's Effects on Team Viability and Performance. *The Academy ofManagement Journal* 49, 1 (2006), 49–68.
[2] Prasad Balkundi and David A Harrison. 2006. Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal* 49, 1 (2006), 49–68.
[3] Sigal G. Barsad. 2002. The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior. *Administrative Science Quarterly* 47:4 (12 2002), 644–675.
[4] Daniel J Beal, Robin R Cohen, Michael J Burke, and Christy L McLendon. 2003. Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of applied psychology* 88, 6 (2003), 989.
[5] Daniel J Beal, Howard M Weiss, Eduardo Barros, and Shelley M MacDermid. 2005. An episodic process model of affective influences on performance. *Journal of Applied psychology* 90, 6 (2005), 1054.
[6] Suzanne T Bell and Brian J Marentette. 2011. Team viability for long-term and ongoing organizational teams. *Organizational Psychology Review* 1, 4 (2011), 275–292.
[7] John Bradley, Barbara White, and Brian Mennecke. 2003. Teams and Tasks: A Temporal Framework for the Effects of Interpersonal Interventions on Team Performance. *Small Group Research* 34 (06 2003), 353–387.
[8] John R Carlson and Robert W Zmud. 1999. Channel expansion theory and the experiential nature of media richness perceptions. *Academy of management journal* 42, 2 (1999), 153–170.
[9] Jennifer A Chatman, Lindred L Greer, Eliot Sherman, and Bernadette Doerr. 2019. Blurred Lines: How the Collectivism Norm Operates Through Perceived Group Diversity to Boost or Harm Group Performance in Himalayan Mountain Climbing. *Organization Science* 30, 2 (2019), 235–259.
[10] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1217–1230.
[11] Jessica Nicole Cooperstein. 2017. *Initial Development of a Team Viability Measure*. Master's thesis. DePaul University.
[12] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.

[13] Carsten De Dreu and Laurie Weingart. 2003. Task Versus Relationship Conflict, Team Performance, and Team Member Satisfaction: A Meta-Analysis. *The Journal of applied psychology* 88 (09 2003), 741–9.

[14] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 18.

[15] James E Driskell, Eduardo Salas, and Joan Johnston. 1999. Does stress lead to a loss of team perspective? *Group dynamics: Theory, research, and practice* 3, 4 (1999), 291.

[16] Amy Edmondson, James Dillon, and Kathryn Roloff. 2007. Three Perspectives on Team Learning: Outcome Improvement, Task Mastery, and Group Process. *Academy of Management Annals* 1 (12 2007). https://doi.org/10.5465/078559811

[17] Paul R Ehrlich and Simon A Levin. 2005. The evolution of norms. *PLoS biology* 3, 6 (2005), e194.

[18] Stale Einarsen, Helge Hoel, and Cary Cooper. 2003. *Bullying and emotional abuse in the workplace: International perspectives in research and practice*. CRC Press, Boca Raton, FL.

[19] Frederich E Emery and Eric L Trist. 1969. The causal texture of organizational environments. *Systems thinking* 1 (1969), 245–262.

[20] Susanne Geister, Udo Konradt, and Guido Hertel. 2006. Effects of Process Feedback on Motivation, Satisfaction, and Performance in Virtual Teams. *Small Group Research* 37 (10 2006), 459–489.

[21] McManus J. Girling, R. 1998. The future of software development in large organizations? (a case study in the application of RAD). *Management Services* 42:4 (1998), 8–17.

[22] Amy L Gonzales, Jeffrey T Hancock, and James W Pennebaker. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research* 37, 1 (2010), 3–19.

[23] Stan Gully, Kara Incalcaterra, Aparna Joshi, and J Beauien. 2002. A Meta-Analysis of Team-Efficacy, Potency, and Performance: Interdependence and Level of Analysis as Moderators of Observed Relationships. *The Journal of applied psychology* 87 (11 2002), 819–32. https://doi.org/10.1037/0021-9010.87.5.819

[24] Stanley M Gully, Kara A Incalcaterra, Aparna Joshi, and J Matthew Beaubien. 2002. A meta-analysis of team-efficacy, potency, and performance: interdependence and level of analysis as moderators of observed relationships. *Journal of applied psychology* 87, 5 (2002), 819.

[25] Richard A Guzzo, Paul R Yost, Richard J Campbell, and Gregory P Shea. 1993. Potency in groups: Articulating a construct. *British journal of social psychology* 32, 1 (1993), 87–106.

[26] J Richard Hackman. 1980. Work redesign and motivation. *Professional Psychology* 11, 3 (1980), 445.

[27] Alexa M Harris, Diego Gómez-Zará, Leslie A DeChurch, and Noshir S Contractor. 2019. Joining together online: the trajectory of CSCW scholarship on group formation. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.

[28] David Harrison, Susan Mohammed, JOSEPH MCGRATH, ANNA FLOREY, and Scott Vanderstoep. 2006. Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality. *Personnel Psychology* 56 (12 2006), 633 – 669.

[29] HM Herman, Marie T Dasborough, and Neal M Ashkanasy. 2008. A multi-level analysis of team climate and interpersonal exchange relationships at work. *The Leadership Quarterly* 19, 2 (2008), 195–211.

[30] Pamela J. Hinds and Diane E. Bailey. 2003. Out of Sight, Out of Sync: Understanding Conflict in Distributed Teams. *Organization Science* 14, 6 (2003), 615–632.

[31] Jim Hollan and Scott Stornetta. 1992. Beyond Being There. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 119–125.

[32] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

[33] Malte F Jung. 2016. Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 3 (2016), 18.

[34] Malte F. Jung, Nikolas Martelaro, and Pamela J. Hinds. 2015. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, USA, 229–236.

[35] Elizabeth Kasl, Victoria J Marsick, and Kathleen Dechant. 1997. Teams as learners: A research-based model of team learning. *The Journal of Applied Behavioral Science* 33, 2 (1997), 227–246.

[36] J. A. Kernaghan and R. A. Cooke. 1990. Teamwork in planning innovative projects: improving group performance by rational and interpersonal interventions in group process. *IEEE Transactions on Engineering Management* 37, 2 (May 1990), 109–116.

[37] Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The Future of Crowd Work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1301–1318.

[38] Pierre-Simon Laplace. 1814. Essai philosophique sur les probabilités. 2e éd. *Paris: Mme Ve Courcier* (1814).

[39] Gilly Leshed, Jeffrey T. Hancock, Dan Cosley, Poppy L. McLeod, and Geri Gay. 2007. Feedback for Guiding Reflection on Teamwork Practices *(GROUP '07)*. ACM, New York, NY, USA, 217–220. https://doi.org/10.1145/1316624.1316655

[40] Dana H. Lindsley, Daniel J. Brass, and James B. Thomas. 1995. Efficacy-Performance Spirals: A Multilevel Perspective. *The Academy of Management Review* 20, 3 (1995), 645–678.

[41] 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, New York, NY, USA, 2330–2343.

[42] Andrew Mao, Winter Mason, Siddharth Suri, and Duncan J. Watts. 2016. An Experimental Study of Team Size and Performance on a Complex Task. *PloS one* 11, 4 (2016), e0153048.

[43] Michelle A. Marks, John E. Mathieu, and Stephen J. Zaccaro. 2001. A Temporally Based Framework and Taxonomy of Team Processes. *The Academy ofManagement Review* 26, 3 (2001), 356–376. http://www.jstor.org/stable/259182

[44] Richard McDermott. 1999. Learning across teams. *Knowledge Management Review* 8, 3 (1999), 32–36.

[45] Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Vol. 14. Prentice-Hall, Englewood Cliffs, NJ.

[46] Joseph E. Mcgrath. 1991. Time, Interaction, and Performance (TIP): A Theory of Groups. *Small Group Research* 22, 2 (1991), 147–174. https://doi.org/10.1177/1046496491222001 arXiv:https://doi.org/10.1177/1046496491222001

[47] Brian Mennecke. 1999. The impact of group process training and role assignments on the performance and perceptions of student iS project teams. *Journal of Informatics Education and Research* 1 (10 1999), 23–36.

[48] Ben B Morgan Jr, Eduardo Salas, and Albert S Glickman. 1993. An analysis of team evolution and maturation. *The Journal of General Psychology* 120, 3 (1993), 277–291.

[49] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.

[50] 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 (UIST '14)*. ACM, New York, NY, USA, 75–85.

[51] Christoph Riedl and Anita Williams Woolley. 2017. Teams vs. Crowds: A Field Test of the Relative Contribution of Incentives, Member Ability, and Emergent Collaboration to Crowd-Based Problem Solving Performance. *Academy of Management Discoveries* 3, 4 (2017), 382–403.

[52] Niloufar Salehi and Michael S Bernstein. 2018. Hive: Collective Design Through Network Rotation. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 151.

[53] Niloufar Salehi, Andrew McCabe, Melissa Valentine, and Michael Bernstein. 2017. Huddler: Convening Stable and Familiar Crowd Teams Despite Unpredictable Availability *(CSCW '17)*. ACM, New York, NY, USA, 1700–1713. https://doi.org/10.1145/2998181.2998300

[54] Steve Sawyer and Patricia Guinan. 1998. Software development: Processes and performance. *IBM Systems Journal* 37 (02 1998), 552 – 569.

[55] Marissa L Shuffler, Deborah Diazgranados, M Travis Maynard, and Eduardo Salas. 2018. Developing, sustaining, and maximizing team effectiveness: An integrative, dynamic perspective of team development interventions. *Academy of Management Annals* 12, 2 (2018), 688–724.

[56] John Suler. 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7, 3 (2004), 321–326.

[57] Bruce W Tuckman. 1965. Developmental sequence in small groups. *Psychological bulletin* 63, 6 (1965), 384.

[58] Melissa A. Valentine, Daniela Retelny, Alexandra To, Negar Rahmati, Tulsee Doshi, and Michael S. Bernstein. 2017. Flash Organizations: Crowdsourcing Complex Work by Structuring Crowds As Organizations *(CHI '17)*. ACM, New York, NY, USA, 3523–3537. https://doi.org/10.1145/3025453.3025811

[59] Annelies EM Van Vianen and Carsten KW De Dreu. 2001. Personality in teams: Its relationship to social cohesion, task cohesion, and team performance. *European Journal of Work and Organizational Psychology* 10, 2 (2001), 97–120.

[60] Femke Velden, Bianca Beersma, and Carsten De Dreu. 2007. Majority and minority influence in group negotiation: The moderating effects of social motivation and decision rules. *The Journal of applied psychology* 92 (02 2007), 259–68.

[61] Daniel M Wegner. 1987. Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior.* Springer, New York, NY, 185–208.

[62] Laurie R. Weingart, Kristin J. Behfar, Corinne Bendersky, Gergana Todorova, and Karen A. Jehn. 2015. The directness and oppositional intensity of conflict expression. *Academy ofManagement Review* 40, 2 (4 2015), 235–262.

[63] Mark E Whiting, Allie Blaising, Chloe Barreau, Laura Fiuza, Nik Marda, Melissa Valentine, and Michael S Bernstein. 2019. Did It Have To End This Way? Understanding the Consistency of Team Fracture. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[64] Mark E Whiting, Grant Hugh, and Michael S Bernstein. 2019. Fair Work: Crowd Work Minimum Wage with One Line of Code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 197–206.

[65] Anita Woolley. 1998. Effects of Intervention Content and Timing on Group Task Performance. *The Journal ofApplied Behavioral Science* 34 (03 1998), 30–46.

[66] Sandy L Zabell. 1989. The rule of succession. *Erkenntnis* 31, 2-3 (1989), 283–321.

[67] Sharon Zhou, Melissa Valentine, and Michael S. Bernstein. 2018. In Search of the Dream Team: Temporally Constrained Multi-Armed Bandits for Identifying Effective Team Structures (CHI '18). ACM, New York, NY, USA, Article 108, 13 pages. https://doi.org/10.1145/3173574.3173682

Received January 2020; accepted March 2020