Did It Have To End This Way? Understanding the Consistency of Team Fracture

MARK E. WHITING, Stanford University, USA
ALLIE BLAISING, California Polytechnic State University, USA
CHLOE BARREAU, Stanford University, USA
LAURA FIUZA, Stanford University, USA
NIK MARDA, Stanford University, USA
MELISSA VALENTINE, Stanford University, USA
MICHAEL S. BERNSTEIN, Stanford University, USA

Was a problematic team always doomed to frustration, or could it have ended another way? In this paper, we study the consistency of team fracture: a loss of team viability so severe that the team no longer wants to work together. Understanding whether team fracture is driven by the membership of the team, or by how their collaboration unfolded, motivates the design of interventions that either identify compatible teammates or ensure effective early interactions. We introduce an online experiment that reconvenes the same team without members realizing that they have worked together before, enabling us to temporarily erase previous team dynamics. Participants in our study completed a series of tasks across multiple teams, including one reconvened team, and privately blacklisted any teams that they would not want to work with again. We identify fractured teams as those blacklisted by half the members. We find that reconvened teams are strikingly polarized by task in the consistency of their fracture outcomes. On a creative task, teams might as well have been a completely different set of people: the same teams changed their fracture outcomes at a random chance rate. On a cognitive conflict and on an intellective task, the team instead replayed the same dynamics without realizing it, rarely changing their fracture outcomes. These results indicate that, for some tasks, team fracture can be strongly influenced by interactions in the first moments of a team's collaboration, and that interventions targeting these initial moments may be critical to scaffolding long-lasting teams.

CCS Concepts: • **Human-centered computing → Collaborative and social computing design and evaluation methods**; **Empirical studies in collaborative and social computing**;

Additional Key Words and Phrases: team fracture; collaboration

ACM Reference Format:
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. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 209 (November 2019), 23 pages. https://doi.org/10.1145/3359311

Authors' addresses: Mark E. Whiting, Stanford University, USA, Stanford, CA, mwhiting@cs.stanford.edu; Allie Blaising, California Polytechnic State University, USA, San Luis Obispo, CA, allieblaising@gmail.com; Chloe Barreau, Stanford University, USA, Stanford, CA, cbarreau88@stanford.edu; Laura Fiuza, Stanford University, USA, Stanford, lfiuza@stanford.edu; Nik Marda, Stanford University, USA, Stanford, CA, nmarda@stanford.edu; Melissa Valentine, Stanford University, USA, Stanford, CA, mav@stanford.edu; Michael S. Bernstein, Stanford University, USA, msb@cs.stanford.edu.

2573-0142/2019/11-ART209 $15.00
https://doi.org/10.1145/3359311

1 INTRODUCTION

Each of us can remember a team that we would rather not work with again. While teams are a relational structure that can support complex interdependence [21, 62], successful interdependence requires that team members engage in behaviors that are discretionary, pro-social, non-programmed, and at times risky, including asking questions, revealing ignorance, ceding power, putting in extra effort, offering their own ideas, monitoring each other, and holding each other accountable [1, 16, 25, 39, 42]. The result is that teams can be motivating and supportive — team members are trusted confidants all pulling together — or they can just as easily be demoralizing and unfair, when team members' fates are tied to others they cannot count on. In the latter case, people will resist, resent, or actively undermine teams [33, 43, 55, 60]. Ultimately, many members of the team may not want to work together again [17].

What causes teams to no longer want to work together? In this paper we study this phenomenon, which we call *team fracture*. We define team fracture as a loss of a team's viability [6]—one that is so dramatic that collaborators do not want to work together again. In this paper, we focus on online teams, including virtual (remote) teams as well as teams of crowd workers, as a lens whose technological mediation allows us to study team fracture. These online teams are increasingly prevalent in the literature and in practice [32, 37, 51, 59], and computer-mediated collaborations have been noted especially for their high levels of antisocial behavior and conflict [8, 10, 23, 56]. Understanding the causes of team fracture in these environments helps us diagnose what can be done to avoid negative affective fallout, and to design platforms that help prevent it. For example, teams in organizations will sometimes fall into states of low viability. Whether or not fracture is inevitable given the people on a team, which would suggest the need to reform membership, or is avoidable, which would suggest the need for mediation and attempts to improve existing team dynamics, knowing more about the underlying causes of fracture will dramatically help teams achieve their goals and minimize friction.

Investigating fracture and its causes in online teams is central to broadening our understanding of computer-supported cooperative work (CSCW) toward a fuller understanding of the affective dimensions of collaboration that can make the team experience either a joy or an albatross. This perspective informs core questions in the CSCW community, including why distance matters [44], why remote teams experience increased levels of conflict [23], and how the future of crowd work might evolve [32]. The understanding also illuminates opportunities for design to scaffold and encourage behaviors in the initial moments of a team's collaboration that minimize the chances of fracture. For example, interventions might detect early predictors of fracture, suggest how a team can improve their language or behavior, or even help fractured teams repair their dynamics.

When a team fractures, members try to make sense of the unpleasant experience and are likely to make personal attributions about teammates: "he is too critical"; "she is too authoritative"; "they do not put in equal work" — with the implication that it would be impossible to ever succeed in a team with such individuals. Indeed, some research studies do suggest that some teams will reliably perform less well than others [65]. Are some teams doomed from the start? If it were possible to rewind time, reset the social dynamics, and try again, would the team get along better, or was the fracture inevitable given the people involved?

In this paper, we perform an online experiment that reconvenes teams a second time without them realizing that they have worked together before. Our method allows us to study the consistency of fracture when prior team dynamics are masked. We developed an online discussion system that performs a two-way masking of participants' identities (Figure 1). In our experiment, participants are grouped into teams for a first task, then into new teams for a second task, a third team for a third task, and so on. Participants see their own pseudonyms as persistent across rounds in the

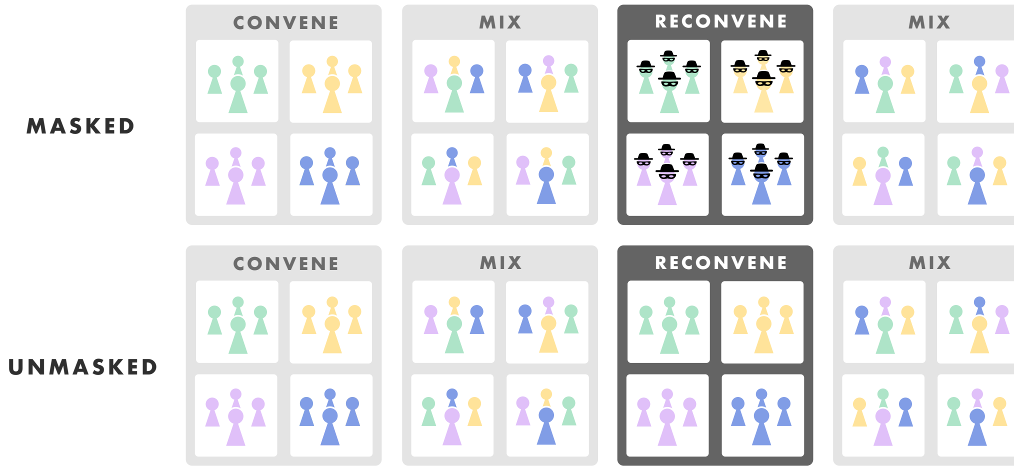

Fig. 1. To evaluate the consistency of fracture, participants repeatedly collaborate in rounds. When reconvened with the same collaborators, identities are masked or unmasked via pseudonyms. Masked teams appear as a new collaboration, while unmasked ones revisit a previous collaboration.

discussion system. However, in reality, our system changes participants' pseudonyms each round and dynamically replaces any mentions (or near misses) of each others' pseudonyms. The result is that participants see their own identity as static and unchanging — leading them to assume the same of others — when, in reality, they are interacting with old teams under new pseudonyms. With participants thinking that this was a new group, the collaboration begins anew without strong attributions carried over. When participants are debriefed and told that two of the teams they worked with were in fact composed of the same people, they correctly guess at only slightly above chance rate (21.5% accuracy vs. 17% chance). We allowed team members to privately blacklist any of their teams from future interactions; if half of the members blacklisted a team, we recorded that team as fractured. We then measured whether the fracture outcome was consistent between the team's first and second convening. We replicate this design across three different tasks drawn from separate areas of McGrath's circumplex [41]: a creative task, an intellective task, and a cognitive conflict task.

We find that the consistency of team fracture is strongly polarized by task. There appears to be little middle ground: tasks either result in extremely consistent fracture, or in extremely inconsistent fracture, even though team dynamics have been reset in all cases. For all tasks, teams in a control condition that use static pseudonyms, and thus retain a memory of their shared experience, are consistent 72–88% of the time. For the cognitive conflict task and intellective task, when the team dynamics are masked, teams almost always replay the same dynamics they developed last time without realizing it, resulting with consistent fracture outcomes 73% and 75% of the time, the same rate as unmasked teams. However, for the creative task, it might as well have been a completely different team: masked teams in the creative task only have consistent fracture outcomes 48% of the time—essentially changing their original fracture outcome at the same rate as a coin flip.

That team fracture consistency is so strongly polarized by task opens up substantial questions for CSCW. Performance can be substantially and stably predicted across tasks by team constructs such as collective intelligence [66]; however, it seems that fracture exhibits different patterns. In classic results on online collaboration [44], task type is less commonly strongly theorized; our result suggests further emphasis on exploring how these results in computer-supported collaboration vary across task types.

Our result provides opportunities for designs to support improved team interaction. For tasks where fracture becomes reversible and path dependent, our result suggests that effective process and platform design can influence the outcome, and that team collaboration platforms (e.g., Slack) can have substantial effects on team viability. For example, on tasks similar to our creative task, teams must design their collaboration to be sensitive to how their early behaviors could cascade into positive or negative outcomes. A computational platform might, for example, help them keep tabs on behaviors associated with negative attributions. Further, ad-hoc teams drawn together rapidly from crowdsourcing marketplaces have attributed teams falling apart to team composition and membership errors (e.g. [37, 53]); our results suggest that for certain tasks, giving the team a fresh start might be as beneficial or even more beneficial than selective membership. In addition, this work suggests a clear need for theorizing the mechanisms that produce affective outcomes such as fracture. For example, platforms supporting creative work would benefit from scaffolds ensuring early pro-social situations, reducing the probability of fracture. Platforms that recruit ad-hoc or crowd teams may especially benefit from more effective rapid onboarding and team-forming.

2 RELATED WORK

Team structures set up a complex interdependence [21, 62]. Members share responsibility and credit for a joint outcome, and must coordinate efforts and combine ideas to accomplish that shared work [2, 27, 50]. Such open-ended interdependence is a risky and complex social relationship, and members of various teams end up with completely polarized experiences. On one hand, when the interdependence is met with a synergistic motivation or with mutually developmental ideation, the results can be better than a sum of the parts, and the team can become an affirming, productive social group to which members are excited to return [19, 34]. On the other hand, when members' time and rewards are so closely intertwined, then unequal effort or lack of respect can be particularly unpleasant, meaning team experiences can just as often deliver demoralizing injustices that have members swearing to never work together again [7, 18, 29, 33, 58]. These experiences motivate our research question: When a team fractures, did it have to end that way?

Research Question. *How consistent is team fracture when membership is fixed but existing attributions are erased?*

Many researchers have explored the potential and the social risk of working in teams, asking when and why some groups become cohesive and viable while others are characterized by betrayals so dramatic that members will not continue working together. This paper builds on the research literature examining these specific affective experiences working in teams. In this literature, *team viability* is defined as the capacity of a team to be sustainable and continue to succeed, and measures both the satisfaction of teammates with their membership and their behavioral intent to remain in the team [6, 9, 21]. In this paper, we define *team fracture* as a related construct that operationalizes members' affective reaction to being part of a team with such low viability that members would choose not to continue collaborating. One reason that these and related variables are of interest is because they are correlated to, but not the same as, team performance [3, 38]. Team performance variables measure task outcomes, such as quality, efficiency, or productivity [11, 36]. Teams that are performing well often have high viability, but there are many conditions under which teams may perform well but still fracture, including toxic norms, seeming toxic personalities, bullying, overwork, unequal contribution, or performance pressure [14, 26]. Many of these factors are situational, but when a team begins to fall apart, members are likely to make personal attributions about teammates personalities being the underlying cause[6, 21].

Team viability and team fracture are particularly relevant for online and virtual teams, but the relationship between known antecedents and team viability or team fracture are also moderated by computer-mediated communication [44]. Antecedents of in-person teams becoming cohesive with high viability include positive shared experience, member demographic similarity, and group prestige or entry difficulty [4, 5, 40, 47]. When team members share these properties, they often develop emergent emotional states of group belonging, group pride, and task commitment, states in which they want to continue their working relationship together [9, 40, 57]. However, online teams are often working under conditions that are not characterized by these known antecedents. They often lack in-person social cues about shared demographics or shared emotions, struggle more with emotional and social entrainment, use asynchronous communication, and in many cases are working together for the first time [23, 67]. In these contexts, cohesion and viability often form through sparse social relationships, and thus emerge through members' interactions in online environments, for example as they use similar language or "bursty" communication pattern [20, 48].

This prior research provides some insight into the challenges to team viability, and the reasons why online teams might fracture. In this paper, we aim to integrate this research, which closely implicates the way that online groups interact to their ultimate viability or fracture, with team members' known propensity to make personal attributions for failed social interactions. If we hold all conditions constant, would the same online team always fracture? Is team fracture inevitable in some teams, as our tendencies toward personal attributions might suggest?

2.1 Interaction patterns, viability, and fracture in online teams

Prior research thus demonstrates that for online teams, group social outcomes such as cohesiveness, viability, and fracture emerge largely through their specific interactions, because their social relationships outside of these interactions are non-existent or limited. This general finding is intriguing because it suggests that if online team members can get their interactions right, they are likely to have an affirming social experience. But online teams are known to be rife with conflict and misunderstanding precisely because the interactions are so spare and lack so much context and social cuing [23, 31, 67]. Together these ideas begin to suggest that the trait characteristics of team members, such as emotional intelligence, social sensitivity, narcissism, or insensitivity are likely to matter for team viability, because these team member traits shape how they interact [61, 65]. Indeed, foundational work on team collective intelligence has consistently shown that team member social capabilities do predict consistent interaction patterns in teams, and consistent levels of team performance across a variety of tasks [64, 65].

Together, this literature suggests that person-specific attributions and path-dependent accretions of actions and affect can impact the outcome. The memory of these actions will carry over to the next time the team works together, which is likely to increase the consistency of the fracture outcome when the team is aware that they are reconvening. So, we hypothesize:

Hypothesis 1 (H1). *Masked teams (i.e., teams who think they are working with a new team, but are working together again) will have less consistent viability and fracture than unmasked teams (i.e., teams who know they are working together again).*

Hypothesis 2 (H2). *Masked teams will have near random odds of repeat fracture outcomes.*

This paper thus builds on prior work by contributing the construct of team fracture, a behavioral measure of it, and a study of the situations in which it is predictable. Prior methods do not attempt to answer questions of how repeatable (or inevitable) team social dynamics are. In general, once you have been burned by an arrogant or power-hungry team member, you walk into the next interaction with that group already on your guard. The social hierarchy has been set in motion, power struggles have emerged, and assessments of personalities well-formed.

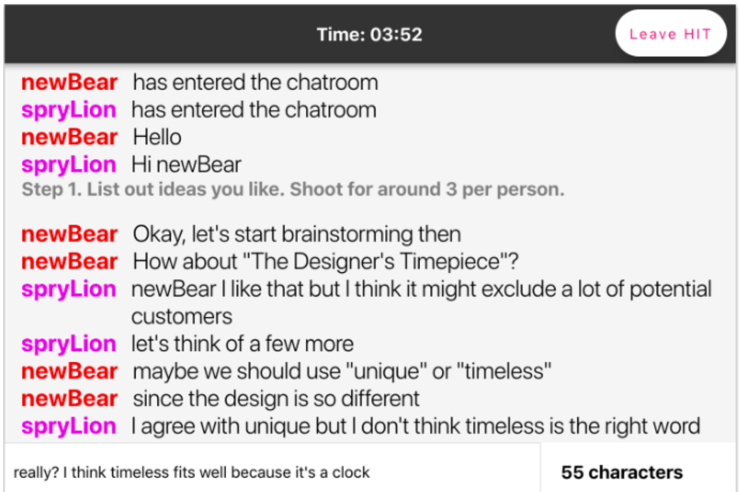

Fig. 2. Participants collaborated in a synchronous chat room environment on one of three different kinds of tasks. This example shows a team doing the ad writing task, collaborating to write a short online advertisement.

Our research introduces a novel technique that provides insight into this question of the inevitability of team fracture. We draw conceptual inspiration from parallel world studies such as MusicLab [54], which randomized people into multiple worlds where the download counts for songs were independent in each world. Likewise, we sought to create multiple parallel instances of the same team. However, prior methods created these parallel worlds by randomizing different people into each world, which is infeasible for us—different people would mean that the team is different. So, we introduce a new method: two-way masking of participants' pseudonym identities. Our intent is to give the same exact group of people multiple chances to construct a team interaction pattern and a cohesive viability or fracture from scratch and to see how consistent their interactions and outcomes are. If successful, we hope that this method could be applied to other team-based social processes such as status and norm development.

3 METHOD

Our research aim is to determine the extent to which a similar fracture outcome emerges when collaborators interact without awareness that they have worked together before. So, we seek a masked membership design that allows us to hide collaborators' interaction histories.

To temporarily hide interaction histories, we introduce a novel deception experiment that leverages pseudonyms so that participants believe that they are working with new collaborators when they are in fact working with prior collaborators. This mechanism relies on repeated group interactions, during which all users appear with pseudonyms at all times.

By default, teams are allocated based on social golfer matching [22], such that each collaboration consists of members who have never interacted before. With this mechanism, each participant interacts with every other participant at most once, unless they are working together for a second time as part of the experiment. In those cases, their interaction will be either *masked*, hiding identities through dynamic pseudonym replacement so that it appears the repeat interactions are with new participants, or *unmasked*, maintaining the same pseudonyms over time so that it is clear when the collaboration is reconvened.

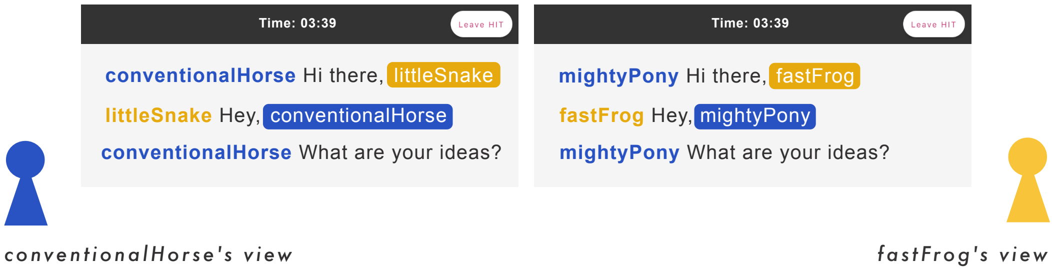

Fig. 3. Chats are converted so all parties see the appropriate unique pseudonyms, and spelling mistakes or typos are automatically corrected, to avoid revealing the manipulation.

To mask interactions, we developed a chat collaboration platform that facilitates multi-person discussions (Figure 2). Our system replaces members' pseudonyms before routing each message out to recipients. It automatically generates pseudonyms in *adjectiveAnimal* format [52], e.g., *conventionalHorse* or *fastFrog*. Participants see their permanent pseudonym, and are not made aware of any new pseudonyms assigned to them when masked. Our system maintains an internal identifier for each participant, as well as the participant's permanent pseudonym and the current public pseudonym assigned to them for the duration of their current collaboration. Because members' private and public pseudonyms may differ, one challenge with our method is that participants may be called on by a pseudonym that they do not perceive as their own. To address this, our system identifies any mentions of their own private pseudonym and replaces it with the public pseudonym on the recipient's client, and likewise replaces any mentions of others' public pseudonyms that they send out with the recipient's private pseudonym. The system uses three strategies to achieve this: (1) an edit distance algorithm to account for misspellings, (2) a starts-with identifier to catch incomplete phrasings of the pseudonym, and (3) an autocomplete interface, similar to those found in tools like Slack, that makes it convenient and faster for users to tab-complete a name than to type it themselves, making errors less likely. Repeated iteration and prototyping established that these algorithmic and interface techniques together prevented any leaked pseudonyms in our dataset.

To prevent participants from noticing this manipulation, the system keeps each participant's own pseudonym static from their own point of view even while it changes from everyone else's point of view (Figure 3). For example, a participant might see themselves with their usual *conventionalHorse* pseudonym and people mentioning them in the chat as *conventionalHorse*. However, to another member of the group, they would be perceived as *mightyPony*. Similarly, they might believe they are interacting with *littleSnake* when in fact that is the masked name pertaining to *fastFrog*.

The identities of participants are entirely isolated from their private pseudonyms, and their public pseudonyms can be replaced, so that others think they are a new collaborator when they are actually not, allowing us to reset the team to a clean slate. In the masked condition teams reconvene with new public pseudonyms, while in the unmasked condition the public pseudonyms remain the same between reconvened rounds of the experiment.

We included a manipulation check question at the conclusion of the study. When debriefing participants in the masked condition, we told them that two of the groups they worked with were actually the same people. We then asked them to identify which collaborators were actually the same (the first group and the third group, the first group and the fourth group, etc.). If the manipulation were weak or if other signals from the team's interactions bled through, participants would be able to guess correctly. If not, they would guess at roughly a chance rate of accuracy.

3.1 Tasks

Participants were organized into groups of four to perform a series of tasks in a series of rounds. They were given ten minutes in each round to work with their group on the task. At the end of each round, participants were moved onto a new group.

We ran the same study design independently across three different tasks in McGrath's task circumplex [41]: creative tasks (type 2), intellective tasks (type 3), and cognitive conflict tasks (type 5). These three task types span three of the four quadrants in the circumplex: creative tasks represent generating ideas and plans, intellective tasks represent choosing a solution, and cognitive conflict tasks represent negotiating the solution to a conflict of viewpoints. By replicating the same study design across a broad swath of the task circumplex, we seek to understand how general any effects are.

For the creative task, we draw from related work [13, 53] to ask collaborators to write short online text advertisements for product campaigns from Kickstarter. In this setup, participants collectively author a 30 character byline advertisement. In each round, all collaborators worked on advertisements for the same Kickstarter project. At the end of the round, the team collectively decided on a final advertisement and one member submitted it on behalf of the team. Teams wrote ads for different Kickstarter projects in each round, so they would never repeat the exact same task.

In the intellective task, participants were asked to correctly answer a series of questions without looking up the solutions online, instead debating with their group members on what the answers to the questions could be [65]. Prompts included estimating the number of states that border the Gulf of Mexico, the percentage of the U.S. population that goes online at least once a week, and the height in feet of Mount McKinley. At the end of each round, the team collectively decided on their final answers and one member submitted a final list of answers on behalf of the team. Teams answered different questions of the same form in each round, so they would never repeat the exact same task.

The cognitive conflict tasks asked a team to allocate funds between a number of competing programs which were each designed to appeal to particular personal values [63]. In each round participants were provided three program options and were asked to collectively decide on how they would choose to allocate a fund of $500,000. At the end of the round, the team collectively decided on a final allocation and a statement of their reasoning and one member submitted it on behalf of the team. The programs' options were rotated such that each round contained a new set of options and a new combination of values.

Each participant was recruited for one of the tasks and completed four rounds of that task: two rounds of work with the same group, and two rounds of work with new groups. The order of these rounds was balanced across groups, so that, for example, the repeated group might occur in the first and third rounds, or second and fourth, and so on. To avoid leaking the manipulation, we excluded all orderings where the same collaborators worked together for two adjacent rounds. So, all participants were reconvened with the same team a second time, but the rounds in which this happened would vary. Rotating membership between teams required that four people be participating in each task. Each group of four was randomized to condition and then proceeded together through each round, interacting in different combinations to form nonoverlapping teams such that each participant worked with each other participant in exactly one collaboration.

In the *masked* condition, when the collaboration was reconvened, the pseudonyms were masked so that participants would be unaware they were working with the same group. In the *unmasked* condition, the pseudonyms were not masked, so participants would recognize each others' usernames.

3.2 Measures

Fracture, in our conceptualization of it, is related to team viability [6]. So, at the end of each round, each participant filled out a 14-item viability scale drawn from prior work [9]. This scale inquired about participants' affect toward current and future interactions with their group. By our definition of the term, team fracture is associated with lower average team viability scores. As the team viability scale is recent, we also explored the first of the remaining steps required for formal validation.

To complement attitudinal data, we sought a behavioral measure of fracture. After all rounds were completed, participants were told there was a chance they might work with one more group, randomly drawn from the groups that they had worked with already. For each of their prior groups, participants were given the option to privately exclude that group from their set of candidates. We defined a team as fractured when at least half of a group's members privately excluded that group from future rounds. Our motivation for choosing half the team was that this represented a nontrivial number of opinions, not just a single unhappy participant.[1]

To understand how much insight team members had into each others' fracture votes [46], we also asked each team member what they thought others on their team would say to this fracture question.

To understand the relationship between fracture and performance, we measured task performance in the creative and intellective tasks. Cognitive conflict tasks do not offer clear performance measures. For the creative ad writing task, each advertisement was added to a campaign for the product on Google Ads. We turned off any optimization, ensuring that ads were seen at the same rate. After 48 hours, we recorded each ad's click through rate (CTR) as a measure of the group's performance [13, 53]. For the intellective tasks, we measured the percent difference each teams answers were from the correct answer, and recorded the median of these values for each team interaction. These two tasks then had individual measures which could capture performance relative to others doing that task.

3.3 Participants

Participants were recruited from Amazon Mechanical Turk. Worker location was restricted to the United States to satisfy language requirements, and we required that all workers had completed at least 100 tasks. Participants were paid $12 per hour for taking part in the experiment, and were incentivized to complete the entire study by receiving a majority of their payment as a bonus at the end of the experiment. We invited workers to sign up ahead of time, then join a waiting room when the task began [35]. The waiting room prompted participants to stay engaged by asking them to respond to questions from a chatbot at regular intervals. Once a sufficient number of participants was active in the waiting room, the main experiment activity would begin.

To ensure that there were enough teams to satisfy the non-overlapping membership constraint, we recruited panels of sixteen people at once. Each panel would proceed through the task rounds together. Each panel was randomized between masked and unmasked conditions.

4 RESULTS

Our results reflect 468 individuals who participated in all four rounds of the study in 79 masked and 61 unmasked teams of 3—4 members. We filtered out teams from the analysis if they lost at least one member due to a disconnection or disinterest during the study, since this would mean that

[1]To test the robustness of our effects, we analyzed the impact of varying this percentage in our analysis code and conducting otherwise identical analysis, e.g., one person voting to fracture at 25%, or a supermajority at 75% of the group voting to fracture. The main results remained consistent. So, we report the results for $>= 50\%$ in this paper.

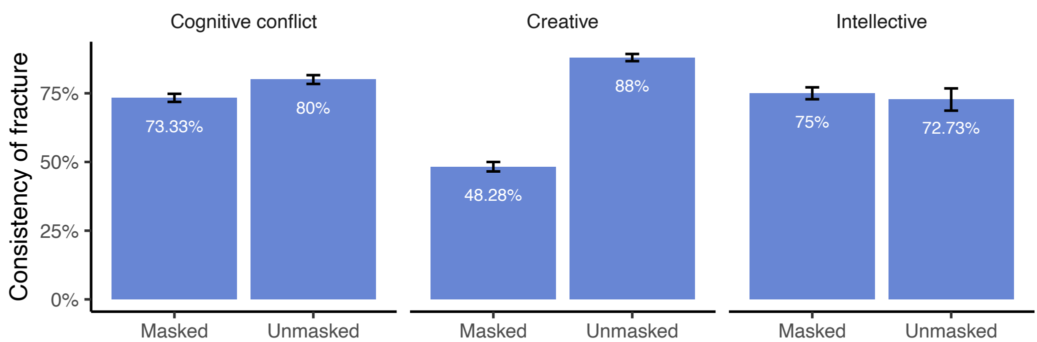

Fig. 4. Unmasked teams exhibited high consistency in their fracture outcomes, 73–88% across tasks. Masked teams were strongly polarized in the consistency of their fracture outcomes: teams in cognitive conflict and intellective tasks were both strongly consistent at 73%, which is essentially the same consistency as unmasked teams; teams in creative tasks were strongly inconsistent at 48%, essentially as consistent as a coin flip.

the team's second convening was missing a member, which would undermine the manipulation. Of remaining teams, 55 performed cognitive conflict tasks, 54 performed creative tasks and 31 performed intellective tasks. The final participant pool was 47.71% female, with an average age of 38.00 years ($SD = 11.54$).

4.1 Manipulation check

Participants worked with four different groups, hence, there are $\binom{4}{2} = 6$ combinations of teams that participants might guess were the repeat teams. A random guessing strategy would be $\frac{1}{6} = 17\%$. Masked participants correctly guessed which teams they had worked with 21.50% of the time, a small amount above chance guessing. We also considered the manipulation on the team level, and found that 65.74% of unmasked teams had correct guesses from at least half their members, while only 15.31% of masked teams guessed correctly.

4.2 How consistent is fracture?

The consistency of fracture in reconvened teams was evaluated by counting the times that teams switched their fracture state, i.e., half or more voted to fracture the first time they interacted but fewer than half voted to fracture the second time, or vice versa. Overall, masked teams changed their fracture outcomes 64.56% of the time, whereas unmasked teams changed fracture outcomes 81.96% of the time. However, broken out by task, the outcomes appear strongly polarized. Teams in cognitive conflict tasks ($masked = 73.33\%, unmasked = 80.00\%$) and intellective tasks ($masked = 75.00\%, unmasked = 72.72\%$) had high consistency for masked teams, at about the same rate as unmasked teams. In contrast, teams in creative tasks ($masked = 48.28\%, unmasked = 88.00\%$) had low consistency, at about a coin toss probability of changing. A logistic regression of each task confirms that masked teams are more inconsistent than unmasked teams in the cognitive conflict task, but not the other two tasks. We show this in Table 1 for the creative tasks, Table 2 for the intellective tasks, and Table 3 for the cognitive conflict tasks.

Table 1. In the creative task, unmasked teams were significantly more consistent in their fracture outcome than masked teams.

	Dependent variable:
	Consistency
Unmasked	2.061^{**}
	(0.719)
Intercept	-0.069
	(0.372)
Observations	54
Log Likelihood	-29.257
Akaike Inf. Crit.	62.514
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 2. In the intellective task, unmasked and masked teams were not significantly different in their fracture consistency.

	Dependent variable:
	Consistency
Unmasked	-0.118
	(0.851)
Intercept	1.099^{*}
	(0.516)
Observations	31
Log Likelihood	-17.692
Akaike Inf. Crit.	39.384
Note:	*p<0.05; **p<0.01; ***p<0.001

This data confirms H1 for the creative task but refutes it for the cognitive conflict and intellective ones, which suggests that during a creative collaboration, a negative outcome with a particular set of collaborators may have been substantially driven by path-dependent behaviors.

Baseline fracture rates provide a useful comparison point. When participants in any condition were collaborating with a team that they would not be seeing again, there was a 24% chance of fracture. So, the consistency we might expect in fracture outcomes between two different teams would be the probability of two fractures plus the probability of two non-fractures, $p^2+(1-p)^2 = 0.52$, roughly even odds. We investigated any possible temporal trends in fracture rates, which would indicate a learning effect or calibration period. While there is some variation by round in these fracture rates, the difference is not significant ($\chi^2(3) = 4.50, p = 0.21$). The emergent trend is that

Table 3. In the cognitive conflict task, much like the intellective task, unmasked and masked teams were not significantly different in their fracture consistency.

	Dependent variable:
	Consistency
Unmasked	0.375
	(0.648)
Intercept	1.012*
	(0.413)
Observations	55
Log Likelihood	−29.908
Akaike Inf. Crit.	63.815
Note:	*p<0.05; **p<0.01; ***p<0.001

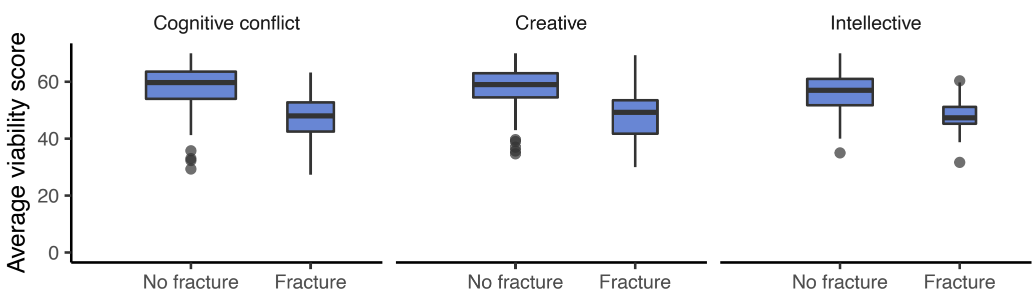

Fig. 5. Collaborations that fractured had lower self-reported viability across all tasks.

the first round had the lowest fracture rate. However, both conditions experienced one of their repeated collaborations in the first round equally, so this is unlikely to affect our conclusions.

To test our behavioral fracture measure's alignment with prior constructs in the literature, we compared fracture outcomes to participants' responses on the team viability scale [9]. The mean viability score was 56 ($\sigma = 9$) out of a possible total of 70. A t-test confirmed that mean viability scores between teams that fractured ($\mu = 48, \sigma = 8$) was lower than teams that did not fracture ($\mu = 57, \sigma = 7$): $t(167.86) = 11.44, p < 0.000$, with a large effect size (Cohen's $d = 1.2$). Figure 5 illustrates this difference.

The 14-item team viability scale used in this paper [9] has not yet been formally validated, so we conducted an exploratory factor analysis (EFA) on the results from 214 teams who completed the scale after their team interaction. Data were screened for multivariate assumptions (normality, linearity, homogeneity, and homoscedasticity). All conditions were met, with slight but not intractable problems with linearity. Following EFA guidelines [45], we first ran Bartlett's test which indicated correlation adequacy ($\chi^2(91) = 5055.522, p < 0.001$) and Kaiser-Meyer-Olkin (KMO) test which indicated sampling adequacy $MSA = 0.97$. A parallel analysis, scree plot and screening of factor values greater than 1.0 and .7 (Old Kaiser and New Kaiser criterion respectively) all suggested one overall factor. Our one-factor model has moderate fit with root mean square error of approximation (RMSEA) at .10 with a 90% CI: $[0.091, 0.119]$, and excellent Tucker Lew Index (TLI): 0.957. Our one-factor model achieved a simple structure with each item loading on one and only one factor, and a raw Cronbach's alpha of 0.99 (indicating high reliability). These exploratory results suggest that team viability could be best measured as a single-dimension. Future research is needed to complete remaining phases of the scale development and validation process [24].

4.3 Do individuals contribute to fracture?

To what extent does membership influence fracture? To answer this, we trained a logistic regression to predict whether the group would vote to fracture the first time they met. We included two main variables: the task condition, and a categorical (dummy) variable identifying each participant in the study. We then measured the AUC—area under the receiver operating curve—as an indication of how well this classifier performed. A baseline model that only had access to condition information had an AUC of .55, indicating relatively poor performance. In comparison, the model that also had access to the individual's identity had an AUC of .76, indicating a more accurate prediction. This result suggests that membership does influence fracture. In particular, specific individuals are either associated with voting to fracture (malcontents) or in causing others to vote to fracture (poor collaborators).

4.4 Is fracture evident from discussion?

We performed an exploratory qualitative analysis of the discussions in the chat logs of teams in the masked condition between their first and second meeting.

We saw few overt signals of fracture, suggesting that private opinions and public behavior differ. However, occasionally, we found that some interactions in chat logs signaled a team's shift. For example, during their first interaction, one team, with a mean viability score of 47.75 and half the members voting to fracture, generated ideas in parallel and with limited interdependence, ultimately just saying which they liked and resulting in teammates overwriting each others' final submissions with no actual consensus:

> *First meeting while masked — fractured*
> **creativeHippo:** I like Time with a Twist too.
> **conventionalHorse:** maybe time telling newly
> **conventionalHorse:** time telling beautifully
> **mightyFrog:** i kinda like just time with a twist also. simple and to the point
> **mightyFrog:** not sure if it needs more
> **mightyFrog:** telling time beautifully is nice
> **mightyFrog:** or time telling if you prefer
> **littleSnake:** LetB: Time with a Twist

In contrast, when this team reconvened later with masked identities, they ended with a mean viability score of 56.75, with no members voting to fracture. Interaction began in much the same way, only this time, a random early event spurred members to pinpoint a common source of inspiration, build upon each others' ideas and provide early feedback:

> *Second (reconvened) meeting while masked — not fractured*
> **mightyFrog:** something about how it makes you smarter? there's smart water so something about smart tea?
> **mightyFrog:** the smart tea for smarties? too corny?
> **creativeHippo:** Thats interesting

conventionalHorse: All you need is tea
conventionalHorse: Smart tea from the sea
mightyFrog: oh thats good too
littleSnake: i like that one

In the cognitive conflict task, one team that was consistent in its non-fracture outcome was also consistent in waiting to hear multiple voices before making any decisions, achieving a mean viability score of 59.5:

First meeting while masked — not fractured
likelyOtter: How are we leaning, with regards to spending the funds?
likelyOtter: I quite like the idea of establishing a community arts program. What do you guys think?
youngFox: I think the veterans support center would be a much better allocation of funds.
spryFish: I would like to give funds to the veterans' center.
newDeer: I also agree with the community arts program. It seems like it would benefit the most amount of people.
spryFish: I have military in my family.
spryFish: I don't have strong feelings about purchasing art. I think the other two options will provide more concrete benefits to people.
likelyOtter: Okay, so it seems like we're divided on two options. Can we pretty much rule out number 2 then?
newDeer: Yeah, i don't think anyone is for #2.
spryFish: We could split the funds evenly between the two we support.
likelyOtter: I think that would represent the feelings of the entire room, so I support that.

In their second interaction, the same norm of waiting and hearing multiple perspectives arose, producing a mean viability score of 67.7:

Second (reconvened) meeting while masked — not fractured
likelyOtter: Libraries are also useful resources to those with low or no income.
newDeer: Yeah, #1 and # 2 seem most important. How about 150k for #2?
spryFish: Sounds good. 400,000 to number 1 and 100,000 to number 2.
newDeer: Yes.
youngFox: I could live with a 350,000 and 150,000 split. Everyone else in agreement?
likelyOtter: I'm good the 350/150 split
likelyOtter: with*
newDeer: How about you spryFish
newDeer: ?
spryFish: Yes
likelyOtter: Great!
youngFox: We're in agreement!
spryFish: I like working with groups with the same values!
likelyOtter: Yeah, me too!
youngFox: It definitely makes life easier.

Another team in the creative task, collaborated effectively when first convening, with a mean viability score of 58.25 and no member voting to fracture. Members took turns generating ideas and playfully encouraged each other in the brainstorming process.

First meeting while masked — not fractured
unconventionalCat: would be fun to play on it with "this tea"
creativeHippo: Thetis tea from the sea
unconventionalCat: you're not a creativeHippo are ya
unconventionalCat: you little unconventional creativeHippo you!
coolWhale: Tis-Tea the stress be gone cure for all
creativeHippo: Tis the time for Tis-Tea was from me
unconventionalCat: using sea is a great play on words
unconventionalCat: esp. cuz seaweed
coolWhale: Tis-Tea the stress be gone cure from the sea.

However, when this same team reconvened, the plainness of one member early on in the interaction stifled the collaborative atmosphere, leading to three of the four members voting to fracture and a mean viability score of 35.

Second (reconvened) meeting while masked — fractured
coolWhale: The Classy stool to be cool
unconventionalCat: meh
creativeHippo: Sit Your Pants (it's a play on words)
coolWhale: Sit you pants with class
unconventionalCat: None B4 Stool N°1
littleSnake: Feel winning, use Stool Number One
unconventionalCat: i don't even know what that means

In addition to understanding fracture from chat logs, we asked participants to privately explain their choices about with which teams they'd be willing to reconvene. Responses for why someone liked a group interaction included aspects such as group contribution, taking the task seriously, and "clashing."

- "I didn't feel like most of the members took the task very seriously."
- "The discussion was not productive."
- "I didn't feel like we clicked well."
- "Not all participated."

On the other hand, people who indicated that they would like to continue working with a given group mentioned aspects such as feedback, creativity, and communication.

- "Because I feel that we had well thought out ideas and discussion."
- "I think they were great to work with! Everyone participated and was helpful. All were kind as well!"
- "They seemed to focus and give good feedback."
- "The team is creative and work hand in hand."

4.5 Behavior under fracture

Although our primary focus is on the viability of the group, one outcome of viable teams is increased performance. Our experiment measured performance signals for the creative ad writing task and the intellective task, and evaluated the magnitude of changes in performance between

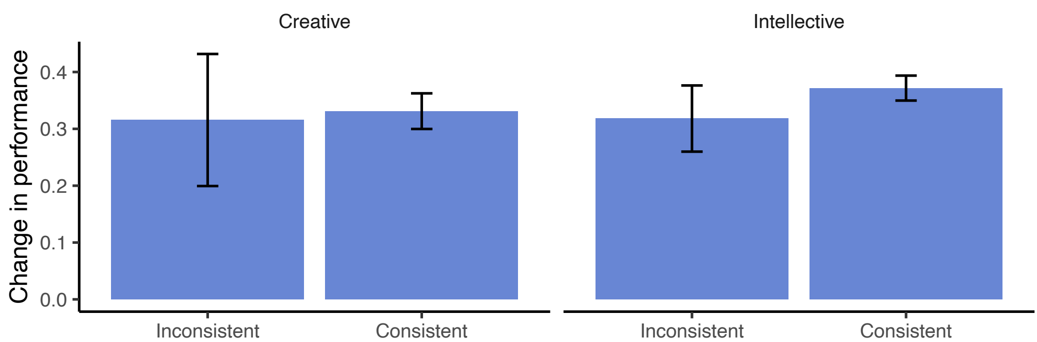

Fig. 6. Consistency in fracture outcomes was unrelated to change in performance outcomes.

first and second interactions of the same teams. Figure 6 shows these changes across tasks, by collaborations that exhibited consistent and inconsistent fracture outcomes between rounds. A logistic regression showed that neither task nor fracture consistency were significantly related to performance (Table 4).

Since we could not evaluate performance in all tasks (e.g. cognitive conflict tasks, due to the nature of this task type), we conducted followup analyses about characteristics present in fractured and non-fractured groups, comparing all the task types in this way. These analyses included a measure of chat turn-taking, length of chat discussions, and sentiment analysis of chat contents. However, they did not yield substantially different results between task types or fracture states.

We analysed turn-taking by counting the number of times new chat messages came from a different team member than the previous chat message. Figure 7 shows how neither participant masking nor fracture outcome affects the number of turns taken in groups. In a model evaluating task type, fracture consistency as predictors of turns taken in a team interaction, in all cases, no factors were significant ($p > 0.5$).

We also evaluated differences in the Gini coefficient when teams did or did not fracture. The Gini coefficient measures how equally distributed the teams' conversations were, with a 0 indicating perfectly equal contributions and a 1 indicating perfectly unequal contributions. Neither message numbers ($t(156) = 1.2410, p = 0.2165$) nor character counts ($t(156) = 0.7078, p = 0.4801$) demonstrated significantly different Gini coefficients. We interpret this result to indicate that fracture is not associated with how dominant participants are in conversation.

We also analysed sentiment using the *sentimentr* package to generate a score between -1 and 1 for each message sent in any given group interaction, with negative and positive scores representing negative and positive messages, respectively [49]. Although sentiment scores varied across different task types, with higher average scores present in cognitive conflict tasks and lower ones expressed in intellective tasks (Figure 8), there was again no significant difference between fracture and non-fracture outcomes ($p > 0.5$).

Taken together, these posthoc analyses suggest that fracture is not a simple function of sentiment, contribution distribution, or performance, and that progress in measuring and theorizing it will be necessary in future work.

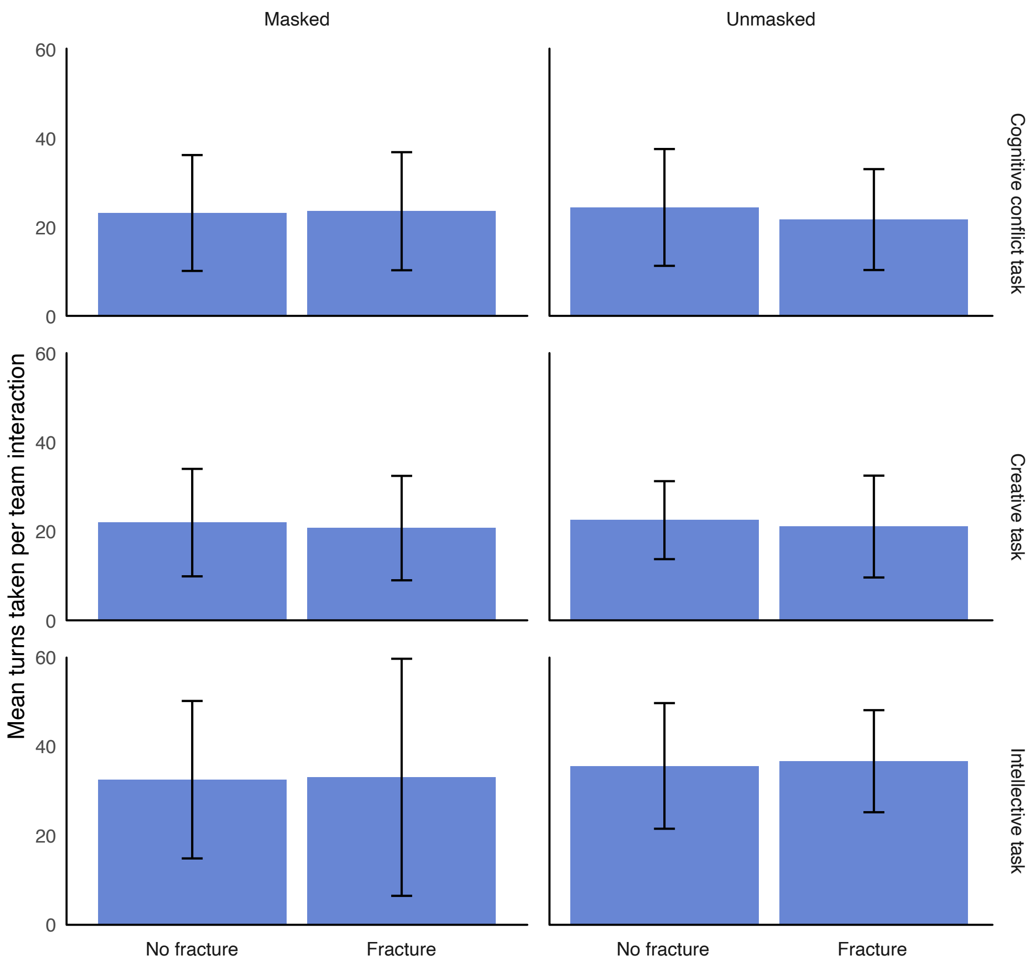

Fig. 7. The number of turns taken between participants in a chat interaction are consistent across different tasks, different rounds, and different fracture outcomes.

5 DISCUSSION

This paper introduces an experimental method for investigating whether history is bound to repeat in fractured collaborations. We find mixed evidence for Hypothesis 1 and Hypothesis 2 — in particular, the effect appears to be moderated by task.

Our results, which indicate that team fracture can be substantially path-dependent in some tasks and substantially robust in other tasks, has bearing on CSCW theory. CSCW has largely focused its dependent variable on team performance; for example, that remote teams suffer lower performance by a factor of two to one [44], that collective intelligence predicts performance [64, 66], that shared awareness can promote more effective information sharing and thus performance [12], and that externally visible conflict can predict performance [30]. Our results report that performance is

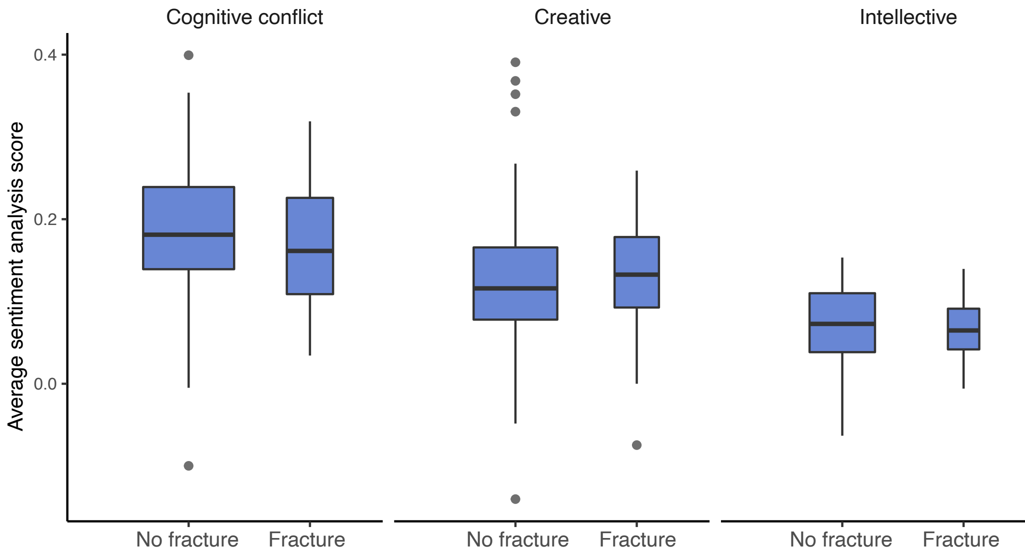

Fig. 8. Sentiment analysis scores were unrelated to fracture outcomes within task types.

statistically orthogonal to an important affective outcome — fracture — that determines whether the team can actually sustain its collaboration. Thus, our work makes clear the need for increased attention on theories of affective and viability-based outcomes in CSCW theory.

The field stands to benefit from theories that extend beyond performance to focus on viability, fracture, and other affective outcomes in team collaboration. Rather than focusing on coordination, or the sheer existence of distance or a mediating channel, these theories will draw focus from other areas such as the psychology of attribution, bias and justice, and computer-mediated communication. For example, CMC has a long history of studying how emotional states can be transmitted. Our results indicate that, at least for the creativity task, *something* unrelated to the task or performance is being communicated via the chat channel which prompts large shifts in the outcome. Identifying this feature and theorizing it will be critical to enabling CSCW to continue to build theories of collaboration that push beyond performance.

We can rule out several theories and explanations given our data. First, there appears to be no difference in baseline fracture rates across tasks, which rules out measurement sensitivity as the issue. Second, fracture is also unrelated to performance in our data, so differences in performance do not appear to be driving the effect. Third, we can rule out some creativity-specific features of the task. The cognitive conflict and intellective tasks are typically deciding a question with an objective right answer, or negotiating between positions. In contrast, the creativity task involved generating new ideas, which suggests that fixation may arise. Fixation [28], when team members prematurely focus their attention and excitement on one option they have generated in a creative process, could cause members to come into conflict over ideas they like. However, we ran pilot experiments with a creative task that did not require teams to decide on a single idea for the team, but instead, asked them to submit their ideas individually, operationally removing the fixation. The results of this pilot mirrored the results shown in the main creative task. Thus, fixation also does not explain the outcome.

Table 4. Absolute performance change is not statistically predictive to consistency of fracture, suggesting that fracture and performance are not tightly related.

	Dependent variable:
	Absolute performance change
Inconsistent	0.071
	(1.208)
Intellective	0.012
	(1.317)
Intellective * consistent	0.166
	(1.493)
Intercept	−0.774
	(1.076)
Observations	49
Log Likelihood	−29.385
Akaike Inf. Crit.	66.770
Note:	*p<0.05; **p<0.01; ***p<0.001

Other explanations arise that cannot be directly tested with our data. For example, there may be effects of the additive vs. integrative nature of the team member contributions. The intellective task and cognitive conflict tasks are additive: the skills and abilities of team members roughly sum. In contrast, design and creativity are integrative: they require combining viewpoints and building on each others' ideas — the experience of being on the team is driven by these integration activities. It is possible that integrative work is more path dependent and sensitive to attributions than additive work.

5.1 Limitations

Our method offers tradeoffs and limitations in its generalizability. First, we focused on teams of four, and it is possible that these dynamics differ with larger groups. Second, we used tasks that are relatively short — while people are effective at making quick judgments of compatibility with team members [37], fracture is very likely to be a time-varying construct. Our method itself likely also has time bounds after which the manipulation becomes apparent, and we have not rigorously tested or characterized those bounds yet. Third, while our method masks old team dynamics, it is not a complete time reversal. In particular, participants are most likely learning from their prior team experiences as they enter each new task. They also likely carry emotional state over from round to round. In other words, if a team fractures viciously, the reconvened and masked team may not know it's the same people, but they may still be gun-shy and try to avoid issues like those that caused problems previously. So, we would never expect to see full consistency in fracture. Fourth, we also want to better explore how norms spread across teams in repeat trials. For example, participants might spread a negative norm as they interact, diffusing an emotional state across teams. Fifth, our definition of fracture focused on attitudes — members' desire to keep working together — but attitudes may not be predictive of a behavioral outcome, such as how well actual teams survive when subsets of the team are inclined to fracture. Moreover, there may be differences based on the proportion of the team that votes to fracture. However, in our analysis of our data stratifying by the percentage of the team voting to fracture, we do not see any strong evidence of impact.

Our experimental approach helps us understand the consistency of team fracture. However, we do not yet have causal evidence of the underlying mechanism that separates creativity tasks from cognitive conflict and intellective tasks. We will continue to work to understand the principles underlying this result.

In this paper, we introduced a conceptualization of team fracture and linked it to team viability. However, we still do not know the full extent of other group and individual level correlates with team fracture. For example, how does individual-level psychological safety [15] relate to team fracture? Or team member familiarity prior to the convening of the team?

5.2 Implications for design

Challenges in remote and distributed teams are a central concern for CSCW [23, 44]. The obvious follow-on question, then, is what role can design play in mitigating fracture? One approach is to try and predict it, then alert the team of the risk and give them the resources to mitigate it. Our results suggest that in some cases, characteristics of the task will suggest how at risk a team might be for a fracture that is path dependent and sensitive to accretions of early actions. Those actions are likely to be visible at some level to the platform, for example through the text chat, and might provide levers for prediction and classification. For the tasks where fracture is consistent, systems might investigate whether fracture could be predicted in advance of the team ever convening, such as through analysis of the individual contributions and behavior of team members before they began collaborating.

With precise knowledge of the kinds of tasks and contexts in which situational attributes are more strongly associated with team fracture, context-specific scaffolds can be designed with a focus on encouraging a pro-social collaboration. For example, intervening to encourage a friendly start to a collaboration where fracture is path dependent could mitigate or reduce it substantially.

On the other hand, in group activities where personal attributes are more strongly associated with team fracture, targeted measures to improve collaborator selection may be taken. For example, pretesting or using low-stakes trial group activities would be likely to surface issues before they occur in a high-stakes situation. In an organizational context, this might involve leveraging practice interactions in which a given team's propensity to fracture could be evaluated without risking a mission critical situation.

More unusual process and system interventions might be feasible as well. Using a technique based on this experiment design, one could experiment with a team intervention that masks teams for their initial interactions and intermixes participants until the team stumbles upon a positive outcome, at which point the team can be unmasked and begin collaborating in earnest. Team fracture may be a random draw on some tasks, but once you have what you want, our experiment suggests that you can keep it. Enabling teams to reduce fracture is valuable academically but also as an opportunity for industrial team contexts.

6 CONCLUSION

In this paper, we introduce the construct of team fracture, defined as a loss of team viability so severe that members no longer want to work together on future teams. We introduce an experimental method for temporarily masking a group's interaction history, allowing us to study the consistency with which the group fractures. We find that the consistency of team fracture is strongly polarized by task type. On creative tasks, it might as well have been a completely different team: teams changed their fracture outcomes at a random chance rate. On cognitive conflict and intellective tasks, the team replayed the same dynamics without realizing it, rarely changing their fracture outcomes. These results suggest that unlike team performance, team fracture cannot be strongly predicted only by stable features of team members.

ACKNOWLEDGMENTS

We are extremely grateful to Alex Huy Nguyen, Pheobe Kimm, Maika Isogawa and Kevin Lin for their contributions to 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, the Office of Naval Research (N00014-16-1-2894), and a National Science Foundation award IIS-1351131.

REFERENCES

[1] Neil R Anderson and Michael A West. 1998. Measuring climate for work group innovation: development and validation of the team climate inventory. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 19, 3 (1998), 235–258.

[2] Caroline Aube and Vincent Rousseau. 2005. Team Goal Commitment and Team Effectiveness: The Role of Task Interdependence and Supportive Behaviors. *Group Dynamics: Theory, Research, and Practice* 9, 3 (2005), 189.

[3] Murray R Barrick, Greg L Stewart, Mitchell J Neubert, and Michael K Mount. 1998. Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology* 83, 3 (1998), 377.

[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] S. T. Bell. 2007. Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology* 92, 3 (2007), 595–615. https://doi.org/10.1037/0021-9010.92.3.595 Bell, Suzanne T. 20th Annual Conference of the Society-for-Industrial-and-Organizational-Psychology Apr, 2005 Los Angeles, CA Soc Ind & Org Psychol.

[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. https://doi.org/10.1177/2041386611405876

[7] Graham Brown, Craig Crossley, and Sandra L Robinson. 2014. Psychological ownership, territorial behavior, and being perceived as a team contributor: The critical role of trust in the work environment. *Personnel Psychology* 67, 2 (2014), 463–485.

[8] 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. https://doi.org/10.1145/2998181.2998213

[9] Jessica Nicole Cooperstein. 2017. Initial Development of a Team Viability Measure. (2017).

[10] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.

[11] Dan R Dalton, Catherine M Daily, Alan E Ellstrand, and Jonathan L Johnson. 1998. Meta-analytic reviews of board composition, leadership structure, and financial performance. *Strategic management journal* 19, 3 (1998), 269–290.

[12] Paul Dourish and Victoria Bellotti. 1992. Awareness and Coordination in Shared Workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work (CSCW '92)*. ACM, New York, NY, USA, 107–114. https://doi.org/10.1145/143457.143468

[13] 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.

[14] 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.

[15] Amy Edmondson. 1999. Psychological safety and learning behavior in work teams. *Administrative science quarterly* 44, 2 (1999), 350–383.

[16] Amy C Edmondson. 2012. *Teaming: How organizations learn, innovate, and compete in the knowledge economy.* John Wiley & Sons.

[17] Frederich E Emery and Eric L Trist. 1969. The causal texture of organizational environments. *Systems thinking* 1 (1969), 245–262.
[18] Shane Flaherty and Simon A Moss. 2007. The impact of personality and team context on the relationship between workplace injustice and counterproductive work behavior. *Journal of Applied Social Psychology* 37, 11 (2007), 2549–2575.
[19] Deborah L. Gladstein. 1984. Groups in Context: A Model of Task Group Effectiveness. *Administrative Science Quarterly* 29, 4 (1984), 499–517. http://www.jstor.org/stable/2392936
[20] 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.
[21] J Richard Hackman. 1980. Work redesign and motivation. *Professional Psychology* 11, 3 (1980), 445.
[22] Warwick Harvey. 2001. Symmetry breaking and the social golfer problem. *Proceedings SymCon-01: Symmetry in Constraints, co-located with CP* (2001), 9–16.
[23] 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. http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=heh&AN=11787478&site=ehost-live&scope=site
[24] Timothy R Hinkin. 1995. A review of scale development practices in the study of organizations. *Journal of management* 21, 5 (1995), 967–988.
[25] M Hoegl and Hans Gemuenden. 2001. Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence. *Organization Science* 12 (07 2001), 435–449.
[26] Helge Hoel, Stale Einarsen, and Cary L Cooper. 2003. Organisational effects of bullying. *Bullying and emotional abuse in the workplace: International perspectives in research and practice* (2003), 145–161.
[27] Daniel R Ilgen, John R Hollenbeck, Michael Johnson, and Dustin Jundt. 2005. Teams in organizations: From input-process-output models to IMOI models. *Annu. Rev. Psychol.* 56 (2005), 517–543.
[28] David G Jansson and Steven M Smith. 1991. Design fixation. *Design studies* 12, 1 (1991), 3–11.
[29] Karen A Jehn and Elizabeth A Mannix. 2001. The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of management journal* 44, 2 (2001), 238–251.
[30] 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.
[31] Atreyi Kankanhalli, Bernard CY Tan, and Kwok-Kee Wei. 2006. Conflict and performance in global virtual teams. *Journal of management information systems* 23, 3 (2006), 237–274.
[32] 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. https://doi.org/10.1145/2441776.2441923
[33] Claus W Langfred. 2007. The downside of self-management: A longitudinal study of the effects tf conflict on trust, autonomy, and task interdependence in self-managing teams. *Academy of management journal* 50, 4 (2007), 885–900.
[34] James R Larson. 2010. *In search of synergy in small group performance*. Psychology Press.
[35] Walter S. Lasecki, Mitchell Gordon, Danai Koutra, Malte F. Jung, Steven P. Dow, and Jeffrey P. Bigham. 2014. Glance: Rapidly Coding Behavioral Video with the Crowd. (2014), 551–562. https://doi.org/10.1145/2642918.2647367
[36] Jeffery A LePine, Ronald F Piccolo, Christine L Jackson, John E Mathieu, and Jessica R Saul. 2008. A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology* 61, 2 (2008), 273–307.
[37] 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. https://doi.org/10.1145/2998181.2998322
[38] Merce Mach, Simon Dolan, and Shay Tzafrir. 2010. The differential effect of team members' trust on team performance: The mediation role of team cohesion. *Journal of Occupational and Organizational Psychology* 83, 3 (2010), 771–794.
[39] Michelle A Marks, Leslie A DeChurch, John E Mathieu, Frederick J Panzer, and Alexander Alonso. 2005. Teamwork in multiteam systems. *Journal of Applied Psychology* 90, 5 (2005), 964.
[40] Michelle A Marks, John E Mathieu, and Stephen J Zaccaro. 2001. A temporally based framework and taxonomy of team processes. *Academy of management review* 26, 3 (2001), 356–376.
[41] Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Vol. 14. Prentice-Hall, Englewood Cliffs, NJ.
[42] Elizabeth Wolfe Morrison, Sara L Wheeler-Smith, and Dishan Kamdar. 2011. Speaking up in groups: a cross-level study of group voice climate and voice. *Journal of Applied Psychology* 96, 1 (2011), 183.
[43] Frank Mueller. 1994. Societal effect, organizational effect and globalization. *Organization Studies* 15, 3 (1994), 407–428.
[44] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.
[45] Kristopher J Preacher and Robert C MacCallum. 2003. Repairing Tom Swift's electric factor analysis machine. *Understanding statistics: Statistical issues in psychology, education, and the social sciences* 2, 1 (2003), 13–43.
[46] Dražen Prelec. 2004. A Bayesian truth serum for subjective data. *science* 306, 5695 (2004), 462–466.

[47] Ray Reagans, Ezra Zuckerman, and Bill McEvily. 2004. How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative science quarterly* 49, 1 (2004), 101–133.

[48] 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.

[49] Tyler W. Rinker. 2016. *sentimentr: Calculate Text Polarity Sentiment.* University at Buffalo/SUNY, Buffalo, New York. http://github.com/trinker/sentimentr version 0.2.3.

[50] Richard Saavedra, P Christopher Earley, and Linn Van Dyne. 1993. Complex interdependence in task-performing groups. *Journal of applied psychology* 78, 1 (1993), 61.

[51] 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.

[52] Niloufar Salehi, Lilly C. Irani, Michael S. Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. 2015. We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1621–1630. https://doi.org/10.1145/2702123.2702508

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

[54] Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 5762 (2006), 854–856. https://doi.org/10.1126/science.1121066

[55] Christine A Sprigg, Paul R Jackson, and Sharon K Parker. 2000. Production teamworking: The importance of interdependence and autonomy for employee strain and satisfaction. *Human Relations* 53, 11 (2000), 1519–1543.

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

[57] Herman HM Tse and Marie T Dasborough. 2008. A study of exchange and emotions in team member relationships. *Group & Organization Management* 33, 2 (2008), 194–215.

[58] Melissa Valentine. 2018. When Equity Seems Unfair: The Role of Justice Enforceability in Temporary Team Coordination. *Academy of Management Journal* 61, 6 (2018), 2081–2105.

[59] 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. (2017), 3523–3537. https://doi.org/10.1145/3025453.3025811

[60] Steven Vallas. 2003. Why Teamwork Fails: Obstacles to Workplace Change in Four Manufacturing Plants. *American Sociological Review* 68 (04 2003). https://doi.org/10.2307/1519767

[61] 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.

[62] Andrew H. Van De Ven, Andre L. Delbecq, and Richard Koenig. 1976. Determinants of Coordination Modes within Organizations. *American Sociological Review* 41, 2 (1976), 322–338. http://www.jstor.org/stable/2094477

[63] Richard T. Watson, Gerardine DeSanctis, and Marshall Scott Poole. 1988. Using a GDSS to Facilitate Group Consensus: Some Intended and Unintended Consequences. *MIS Quarterly* 12, 3 (1988), 463–478. http://www.jstor.org/stable/249214

[64] Anita Williams Woolley, Ishani Aggarwal, and Thomas W Malone. 2015. Collective intelligence and group performance. *Current Directions in Psychological Science* 24, 6 (2015), 420–424.

[65] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330, 6004 (2010), 686–688.

[66] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science* 330, 6004 (2010), 686–688. https://doi.org/10.1126/science.1193147 arXiv:arXiv:1011.1669v3

[67] Norhayati Zakaria, Andrea Amelinckx, and David Wilemon. 2004. Working together apart? Building a knowledge-sharing culture for global virtual teams. *Creativity and innovation management* 13, 1 (2004), 15–29.

Received April 2019; revised June 2019; accepted August 2019