Retrospector: Rapid collaborative reflection to improve collaborative practices

SOYA PARK, MIT, USA
CHINMAY KULKARNI, Emory University, USA

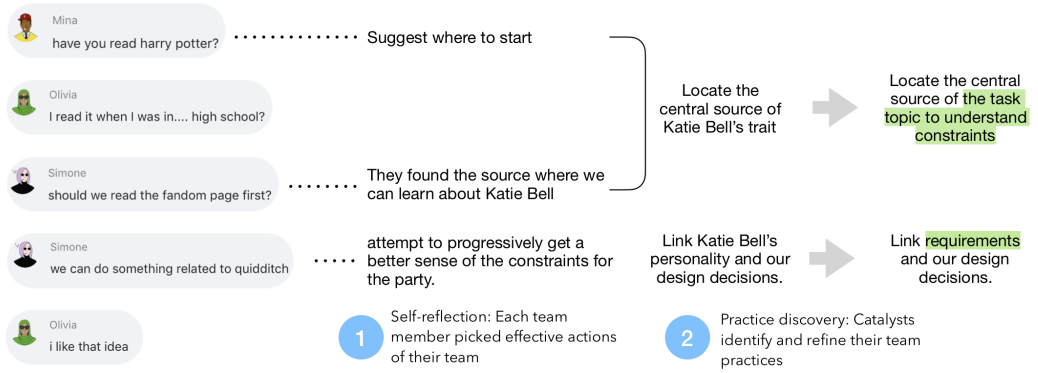

Fig. 1. The Retrospector workflow for helping ad hoc teams discover and reinforce effective team practices: (1) Group members sensemake their work patterns by choosing actions that they find it helpful for their groupwork (2) Catalysts are in charge of specifying and then generalizing team-generated practices. The effective team practices generated through Retrospector improve team performance and viability.

Online platforms for freelancing allow teams performing complex work to be assembled in a matter of minutes and dispersed nearly as quickly. With such short time frames, ad hoc and virtual teams have few opportunities to learn strategies and effective team practices to work with their colleagues. Without such practices, teams are prone to work sub-optimally and lack direction. One key challenge in virtual teams discovering effective team practices is that because the practices ought to involve situated knowledge, it takes time to coalesce, as team members learn about each other over time.

This work introduces Retrospector, that ad hoc teams can use to reflect collaboratively and reinforce effective team practices. Our interface accelerates the discovery of practices in situ and then guides them in reinforcing and applying these practices to future tasks. We conducted a between-subjects experiment (N=75) to assess our design with crowdworkers from the Amazon Mechanical Turk platform. This randomized controlled experiment showed that teams using our system for approximately six minutes of collaborative reflection were able to discover effective practices more successfully and had significantly improved team performance and viability. These results indicate that deliberate support for improving team practices can improve outcomes even through very short interaction. We conclude with design implications and opportunities for future work.

Authors' addresses: Soya Park, MIT, Cambridge, Massachusetts, USA, soya@mit.edu; Chinmay Kulkarni, Emory University, Atlanta, Georgia, USA, chinmay.kulkarni@emory.edu.

2573-0142/2023/10-ART293
https://doi.org/10.1145/3610084

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

Additional Key Words and Phrases: team work, team practice, knowledge capture, knowledge management

ACM Reference Format:
Soya Park and Chinmay Kulkarni. 2023. Retrospector: Rapid collaborative reflection to improve collaborative practices. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 293 (October 2023), 20 pages. https://doi.org/10.1145/3610084

1 INTRODUCTION

Online platforms have dramatically altered the way we work by enabling flexible and distributed forms of computer-mediated work, empowering teams of people to come together quickly and collaborate towards complex goals – in effect forming "virtual teams". With such ad-hoc virtual teams, team membership is volatile, and teams are organized around short-term "projects"; consequently, team members need to adjust their ways of working with each other constantly [56]. Team members also frequently have to work across multiple sites of collaboration, and with different technologies [40]. For example, employees regularly work closely with colleagues they will never meet in person, requiring teams to re-learn how to engage even in basic collaborative activities, such as meetings and socialization [33, 48]. Many teams also do not survive "team fractures" due to interpersonal and emotional conflicts that are particularly frequent in ad hoc teams [28, 52], and online freelancers who participate in such teams report additional social and emotional overhead in navigating team-based work [5].

Unlike traditional teams in an organization, ad hoc teams inherently lack an existing base of effective team practices for working together with their colleagues, a form of organizational knowledge [39]. As a group of people work together, they adjust and tune the way they work and develop team practices that make them effective at their task and in collaborating with each other. Team practices are necessarily a form of situated knowledge and are therefore contingent on team membership and tasks [39].

Effective team practices are hard to discover for any organization, and traditional organizations expend significant effort in discovering those, particularly *best practices*, i.e. practices that have produced outstanding results [38]. Prior work studying such organizations' experience suggests that at least two challenges are heightened for ad hoc and virtual teams. First, effective team practices are not obvious to team members such that when teams work with each other based on the practices, they may not realize it. Therefore, articulating and listing these practices in traditional organizations is often done with the aid of third-party experts (e.g., [4, 38, 39, 43]). Second, because of the situated nature of effective team practices, they require time and accumulated experiences to develop. Due to ad hoc teams' short-lived nature, team membership in virtual teams is more fluid, so practices may not be reinforced over time. As a result, effective team practices, even when discovered, are likely to atrophy over time [14, 46].

Because of their temporary and fluid nature, ad hoc teams cannot rely on experts to accumulate effective team practices, as traditional teams might [38]. Similarly, due to their situated nature, teams cannot simply borrow some other team's effective team practices. In particular, researchers have found that the process of devising effective team practices is a substantial part and highly specific for each group or team [10, 14], and the borrowed practices will likely not be a good "fit" to the team [4, 46].

In this paper, we explore an alternative approach. Instead of using outside experts or borrowing, we seek to accelerate the discovery process of effective team practices such that ad hoc team members can engage in it effectively (our resulting system, Retrospector, only requires six

minutes for practice discovery.) In doing so, we envision a tool that helps teams recognize and reinforce their effective team practices.

RETROSPECTOR offers space and a lens for teams to engage in sensemaking regarding their current practices. RETROSPECTOR is designed around the Shared Mental Model (SMM) framework. SMM is a framework designed to help team members build a common understanding of their team's tasks [7]. The interface asks for users' inputs on examples of SMM practices (planning and sharing task-relevant knowledge) during their team work. By providing such examples, teams can collect a set of instances that captures their teamwork. RETROSPECTOR then helps team members refine the generated practices so they are applicable and actionable by relying on *catalysts*. Catalysts are members of a team whom the team elects to be in charge of applying the practices [14]. RETROSPECTOR interfaces take catalysts through the process, where they can tailor practices to implement goal-directed actions [17, 54]. In effect, RETROSPECTOR provides a structured space for team members to reflect on their practices and identify effective ones, creates a new team role (the catalyst) who reinforces the practices, and identifies members well-suited for this role. Together, these actions create a space to reflect and reinforce effective team practices.

To evaluate and inform future development of RETROSPECTOR, we conducted a randomized, controlled study to assess its effectiveness. We recruited crowdworkers on Amazon Mechanical Turk to conduct group tasks with or without the guidance of RETROSPECTOR (N=75). Recruiting from Mechanical Turk workers yields a conservative sample: not only do participants form ad hoc team, but they are likely never to work together in the future and therefore have minimal incentives to engage in building effective team practices. Even with this conservative sample, however, we found that teams using RETROSPECTOR developed more useful team practices ("I would see which team member has the most knowledge of [relevant task]") than those in the control group, whose practices were often vague truisms ("respect each other's ideas"). Overall, RETROSPECTOR improved team viability (i.e., team members' desire to work with each other again in the future) by 15%. Furthermore, independent raters blind to condition deemed the team outcomes of RETROSPECTOR teams significantly better overall. At the same time, participants reported no significant differences in their satisfaction with the team outcome, or in the effectiveness of their team practices, consistent with prior work showing that the (lack of) effectiveness of team practices is not immediately apparent to team members.

This work makes the following contribution: it introduces a method for ad hoc teams to find effective team practices rapidly and to reinforce them, improving team outcomes. More generally, it suggests a methodology to develop team reflective tools that expose teams' situated knowledge. Although this paper focuses exclusively on effective team practices, similar systems may accelerate other situated organizational knowledge.

2 RELATED WORK

We sought a new design for finding effective team practices. We leveraged structured reflection from team members. In this way, team members can collectively gather their work practice and evaluate what works best for them. Here, we review prior work on interfaces for capturing knowledge through structured reflection. We then explore systems that guide teams to achieve better performance. Lastly, we share work about effective team practices in large organizations.

2.1 Knowledge Capture

RETROSPECTOR is motivated by prior work on knowledge management technologies in teamwork and organizations [1]. In particular, we are interested in knowledge capture. Across many kinds of cognitive work (e.g., programming, UX design), the knowledge and know-how that a worker or team generates may benefit future colleagues. For example, in an organization, scientists who are

looking to analyze their experiment can benefit from inheriting a script from their colleague who conducted a similar experiment [2]. In an open-source repository, newcomers to the code base can benefit from underlying logic and rationales of implementation of functions for why the code is written in a certain manner [41].

Although a substantial amount of knowledge emerges during cognitive tasks, capturing that knowledge is challenging. Specifically, different abstractions of knowledge introduce different challenges for knowledge capture and application. Knowledge abstraction falls along a spectrum, which can be divided into four main levels (Theory - Intermediate - Mediated - Immediate). The most abstract side of the knowledge abstraction spectrum is theory [23]. Although theory holds true and is widely applicable, it is hard to *apply* in practice because it is abstract and unclear how to tailor it to individual situations. For example, although the guidance to "socialize with your team members to strengthen the team" is generally helpful, individual team members might find it difficult to execute it if their team is hybrid or has hundreds of members [22]. On the other hand, the most concrete type of knowledge is so-called immediate knowledge. Immediate knowledge is situational and straightforward to apply to a given situation. However, it is hard to generalize to other situations. For instance, the know-how to run socialization events for in-person teams might not be applicable to hybrid or remote teams. Researchers have explored systems to capture immediate knowledge; one way to capture in situ knowledge is to help users record the knowledge while they are completing the tasks. For example, Krosnick et al. introduce a system that encourages knowledge capture at the moment using an ambient display [30]. Another approach is to help users remember the existence of immediate knowledge by annotating the timestamp during the actual work and reflecting on it later, once the work is completed; Co-notate is a system in which designers can annotate their meeting as it is being recorded and tag notable moments in real time [44], allowing designers to revisit immediate knowledge that emerged during their meeting quickly and reflect on it.

Between theory and immediate knowledge lie mediated and intermediate knowledge. Researchers have proposed various methods to capture intermediate knowledge in the area of design research [23, 32]. Although which format of intermediate knowledge is most useful to practitioners is still in debate, intermediate knowledge offers more concrete guidelines than theory and more general applicability than immediate knowledge [19]. Our work can be viewed as intermediate knowledge capture in the domain of team collaboration. Our interface allows team members to dissect their collaboration, capture effective team practices, and refine to be more situated, rather than merely situational.

2.2 Debugging Teamwork

Previous studies have suggested two main types of interventions at different stages of team development [49] to enhance team performance [31]. The first type of intervention involves *altering* teams, which includes introducing changes to team formation (i.e. who gets to be in a team) [18, 21, 26, 45], team structures (i.e. who gets to be a team leader) [20, 50], or team norms (i.e. how to work together as a team) [40, 59]. Implementing such interventions can lead to improved team performance, but they require a thorough understanding of each team member, the dynamics within teams, and the allocation of resources to experiment with different team configurations. For instance, Zhou et al. [59] propose a system that recommends teams try out various team structures, such as hierarchy and norms, and identify the one that works best for them.

For most teams, trying out various team structures, norms, or memberships might not be feasible [56]. Teams have limited resources and time to experiment with various strategies. Instead, they only can incrementally tweak work practices, if they observe any. In addition, with only

reflection without guidance, teams might remain in the dark regarding how to improve their teamwork without actual implementation or action items [17, 54].

Another type of intervention involves empowering teams to *improve* their current performance. This approach provides guidance to help teams reflect on their teamwork through an elaborated lens, often utilizing visualization techniques or metrics [6, 9, 11, 28]. By engaging in team reflection, team members can proactively address inter-team conflicts and work towards resolving them effectively [28]. Such reflective exercises not only foster sustainable team dynamics but also contribute to the creation of satisfying and productive workplaces [31]. In this study, instead of overhauling and trying out various practices, we explore interfaces that can guide teams to self-reflect with a focus on finding and applying their own effective team practices. Our interface provides lenses which teams can use to understand their team and leads to collaborative practices.

2.3 Effective Team Practices

Effective team practices are powerful [39]. They serve as situated knowledge or guidelines gained from previous experiences. Effective team practices come in various forms and can enlighten various aspects of team performance [14, 24, 25]; for example, effective team practices can turn into checklists, in which each item help teams to ensure they adhere to them [16, 53]. Checklists in a complex and high-risk domain help team members to build a shared context and reduce mistakes [55].

However, finding effective team practices is expensive and time-consuming for at least two reasons. First, effective team practices are often *unrecognized*. During teamwork, various interactions, workflows and activities emerge. For example, once a mere routine of stock brokers to drink coffee together during break, brokers discovered a simple coffee break was an effective team practice because it served as an opportunity to exchange information about the market informally [47].

Although team members are not able to immediately recognize their team's routine and activity as effective team practices [38], prior work has found that designating people to manage practices is effective in identifying and spreading them [12, 14, 27]. Furnari suggested establishing designated individuals called *catalysts*, who are actors sustaining others' interaction, moderating discussion, and reinforcing their group's practices [14]. They found that having catalysts in cross-institutional gatherings helps the group discover new practices. We employ catalysts to identify team practices and make sure they are applied during teamwork. In RETROSPECTOR, catalysts go through an interface for practice discovery that presents records of reflection done by each member individually. After this, during future team interactions, catalysts are asked to remind members about the practices throughout the collaboration, and others are asked to follow their catalysts' guidance.

Second, effective team practices are local and involve contextual knowledge that is specific to teams [39]. Therefore, identifying effective team practices often relies on a teams' accumulated experience and external experts' help. In organization science, best practices are widely explored, as large organizations accumulate knowledge and experience over time [39]. According to management and ethnographic studies done by researchers, organizations can identify and share their effective team practices within their organization [15, 38, 40]. For example, Orlikowski outlined best practices in a global organization through approximately 80 interviews with employees over six months [39]. Although these resources are abundant in big organizations, ad hoc teams, as well as smaller traditional teams do not have enough accumulated knowledge or experience, and they are not experts who can help teams recognize their best practice. In our work, we aimed to build a system that ad hoc teams can use to efficiently discover and reinforce effective team practices themselves.

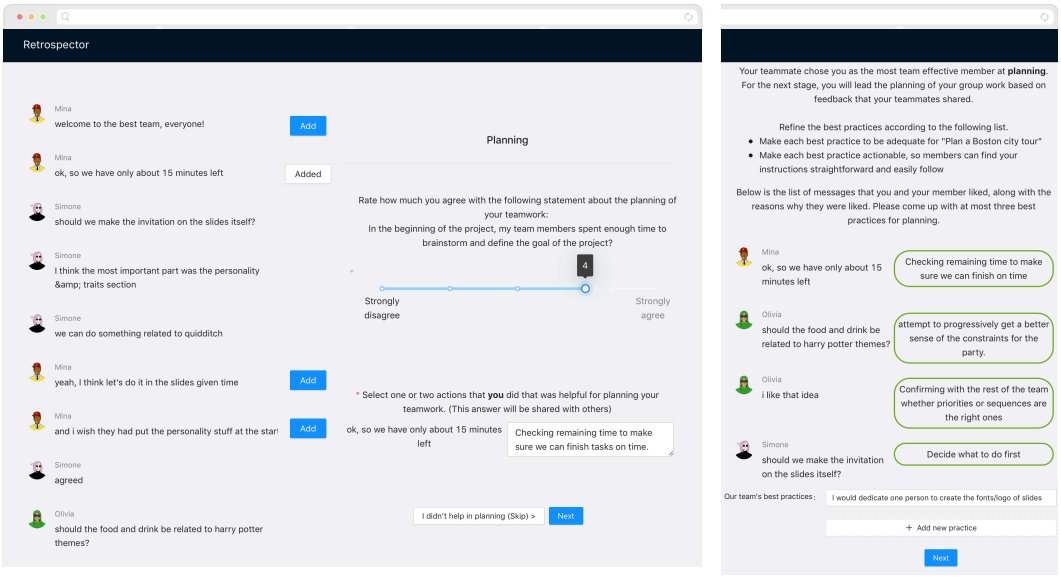

Fig. 2. Retrospector interface: (Left) Self-reflection: each member was asked to revisit their team's collaboration and conversation through an SMM questionnaire (Right) Practice discovery: once all members have submitted the questionnaire, Retrospector selects catalysts. Catalysts are asked to come up with effective team practices for their team based on members' responses from the previous questionnaire.

Table 1. Shared mental models [37] practices. Retrospector guides team members in collecting SMM practices through group reflection. First, each member engages in *self-correction training* by examining their teamwork according to prompts of *planning*, *reflexivity*, and *team-interaction training*. Then, catalysts engage in *leader briefing* based on members' records of *self-correction training*.

Shared Mental Model Practices	Description
Self-correction training	Individual team members reflect on their process and performance.
Planning	Team members set a goal prior to a mission, during a mission or both.
Reflexivity	Team members reflect on how well they fix their mistakes.
Team-interaction training	Team members share task-relevant knowledge.
Leader briefing	Team leaders share the team's goals and priorities.
Cross-training	Individual team members learn their colleagues' responsibilities and roles.

3 SYSTEM DESCRIPTION

This section introduces the interface of Retrospector. To support team members in identifying and applying effective practices for their team, we designed and implemented an interface which allows them to reflect on the team's conversation. Retrospector guides team members through reflection, electing a catalyst, and discovering team practices (Fig. 1).

3.1 Example Scenario

Julia is a real-estate agent who decides to hire a team of workers from a crowdworker platform, because she wants an ad hoc team working on her condo listings which she recently took over. She recruits a designer, a marketing expert, admins and a webmaster who will closely work together to post up-to-date information online and handle customers' inquiries. After a week, the team works well together overall, except for a few instances in which they made mistakes.

Julia decides to use RETROSPECTOR for her team. The team members use RETROSPECTOR to reflect on their work from the past week. Through the reflection, The last time the marketing expert had a list of requests for the designer, she shared a collaborative editing document containing the list with the designer and webmaster. They all found having the shared document helped them be on the same page and track the progress. Another issue they found through reflection dealt with the daily stand-up meetings the admins have to go through the listings. At one point, they needed to ask the webmaster to update the website, but no one told the webmaster about it. As a result, the website was left out-of-date for a few days.

RETROSPECTOR elects a catalyst using these inputs from the team members. The catalyst looks through them and decides on the following effective practices using RETROSPECTOR: from now on, (1) the marketing expert, designer, and webmaster will maintain a shared document that they can add collaborative tasks to. (2) For the daily stand-up meeting, admins take turns being in charge of letting others know about the tasks mentioned in the meeting. Julia finds that since the team uses RETROSPECTOR, there are fewer mistakes and they work more effectively.

3.2 Supporting Individuals' Reflection on Teamwork

We are interested in capturing the team's current way of working so that they can reflect on it. As a result, in the beginning of a team task, RETROSPECTOR does not interfere with users' practices, so they can work as naturally as they would without RETROSPECTOR. After that, the conversation log is analyzed by the group members themselves.

RETROSPECTOR provides a lens where team members can dissect and engage in sensemaking about their work behavior. We adopted the Shared Mental Model (SMM) framework [37] because it provides an abstraction for team members' common understanding of task responsibilities and what the corresponding information needs are. This allows them to more readily anticipate each other's needs and work together more efficiently. RETROSPECTOR lets them reflect on their performance in the following three SMM practices [57]: (1) their planning (whether they spend enough time brainstorming before starting to work on tasks), (2) their reflexivity (whether and how they work on solving their mistakes), and (3) their management of task-relevant knowledge (if team members share information relevant to tasks) individually (Table 1). In more detail, RETROSPECTOR asks the following questions about each SMM practice (Fig. 2. Left):

- Self-rating the team's effectiveness on the SMM practice: rate their performance of the practice on a Likert scale.
- Voting catalysts: select the most effective member for the practice.
- Sharing successful team actions: pick one or two instances of good execution of the SMM practice by the member and how they helped the team.

3.3 Discovering Team Practices through Members' Reflection

Once everyone has finished their self-reflection, RETROSPECTOR appoints *catalysts* [14] for each SMM practice by a majority vote from the previous step. In the case of ties, the system uses the rating as a tie-breaker. Because their team members voted for the catalyst who is the most effective at an SMM practice, they are responsible for carrying out the practice-discovery process. RETROSPECTOR

presents team actions selected by members to catalysts (Fig. 2. Right). Prior work has suggested that people tend to succeed when they set challenging and specific goals as compared with vague goals (e.g., "do your best" goals) [17, 34]. To guide catalysts in coming up with specific practices, Retrospector asks them to find commonality between the selected actions. Then, catalysts refine the actions; in other words, they consider how they would change the actions if their team were to do the same group task again. If they can think of a better way of performing the actions, they are asked to suggest it.

Although the refined actions might be straightforward for the team to re-run in the same task, it might not be generalizable to other tasks. In the next prompt, Retrospector asks catalysts to further refine actions from the previous task and turn them into team practices that are applicable to other tasks. Catalysts are given a checklist as a guideline to reflect and change the actions to team practices. The checklist includes the following items:

- Make each best practice adequate for "[Next group task]".
- Make each best practice actionable so that your instructions are straightforward and easy to follow for members.

At the end, each catalyst generates a maximum of three team practices. Retrospector limits the number of practices so it is manageable for catalysts to keep track of them during later group tasks.

3.4 Reinforcing Team Practices

After catalysts generate the team best practices, Retrospector displays them during subsequent group tasks. The system instructs catalysts to make sure practices are being used in teamwork, whereas the remaining members are also informed to follow the catalysts' instructions.

3.5 Implementation

Retrospector is a Meteor application with Ant Design[1] for interface components and Mongo DB for data storage. The source code of Retrospector is available at https://github.com/[redacted for review].

4 EXPERIMENT

With our system Retrospector, we conducted a controlled study. We sought to explore how team practices using Retrospector impact 1) the performance of teams, and 2) collaboration behavior relative to a control condition in which teams did not use Retrospector.

4.1 Study Design

Task. Two different group tasks were given to participants: the initial task was to (1) make a birthday invitation card for Katie Bell, a fictional character in Harry Potter. After the initial task, and once catalysts had discovered team practices for their group, the groups were asked to (2) plan a day trip to Boston. We chose these two tasks, as they were instantiations of a "creativity" task in McGrath's Task Circumplex Model [36]. Such tasks required teams to work interdependently and engage in discussion and coordination to arrive at a solution. At the same time, the two tasks required different forms of background knowledge and strategies to finish the task. We intentionally designed the two tasks to be different, so that we could inspect whether the team practices generated through our interface could apply to other tasks or only to the initial task and provide task-specific insights.

[1] https://ant.design/

Study protocol. Participants were randomly assigned to either the control or experiment group and that all members in a given team were assigned to the same condition. We used TurkServer [35] to coordinate and match multiple crowdworkers. Once five participants were ready, they were directed to the study website. If too many participants in a team dropped out during the study and it was less than the threshold (three members), RETROSPECTOR finished the session for the team and paid team members for a proportion of the time they participated. The procedures were as follows:

(1) Initial group task (20 mins): The groups were invited to conduct an initial group task together. They were provided with a text-based chat interface and a Google Doc side-by-side.
(2) Self-reflection and practice discovery: Participants then individually engaged in reflection on their collaboration during the initial task. For the RETROSPECTOR groups, individuals filled out the self-reflection questionnaire. Once everyone finished their individual reflection, RETROSPECTOR introduced the two team members who were selected as catalysts to the participants. Catalysts then captured team practices based on members' inputs. For the control groups, individuals did not fill out the reflection questionnaire. The interface randomly assigned two users as catalysts. For both conditions, while catalysts are engaging in team-practices discovery, the remaining participants were asked to pause and revisit the conversation and collaboration experience thus far. We used only *planning* and *team-interaction training* from SMM practices – after iterations of pilots, we found that participants were struggling to find solutions for the remaining SMM practices. Such teams could not discover practices for *reflexivity* as the study task was open-ended, so there were no clear right or wrong answers, and it was short term, so they did not make any notable mistakes.
(3) Last group task (20 mins): The task interface and timer were resumed. Catalysts were asked to lead their teams according to their practices.
(4) Exit survey: At the end of the study, participants were asked to fill out an exit survey focused on understanding their experience and comparing team collaboration between initial and final group tasks.

Participants. Participants were recruited from Amazon Mechanical Turk. We restricted our recruitment to adults located in the United States. They were paid $15 per hour for their time and effort. Seventy-five people (mean age=40) participated. Most participants were employed full time (80%), others had part-time employment (17%), and the rest were unemployed or did not specify. Also, 34% of participants identified themselves as female and the rest as male. The highest educational degree was a bachelor degree (63%), and others had secondary education (33%), or primary education (4%). Participants were randomly assigned to either a control or experiment group.

4.2 Measures

We employed several measures to test the effect of different conditions.

4.2.1 Team interaction. To compare the team interaction of different conditions, the conversation logs during the collaboration tasks were coded and analyzed using the linguistic dictionary Linguistic Inquiry and Word Count (LIWC) [42]. Through multiple iterations of discussion and review of the conversation logs in the research team, we consolidated a codebook that is a modified version of the conversation scheme in [29]. Two coders independently labeled 80 messages based on the codebook, resulting in an inter-rater reliability (IRR) of 76% using Cohen's κ.

In addition, we also collected participants' self-reported reaction through an exit survey. The exit survey included questions about how much they were satisfied with their team, how much

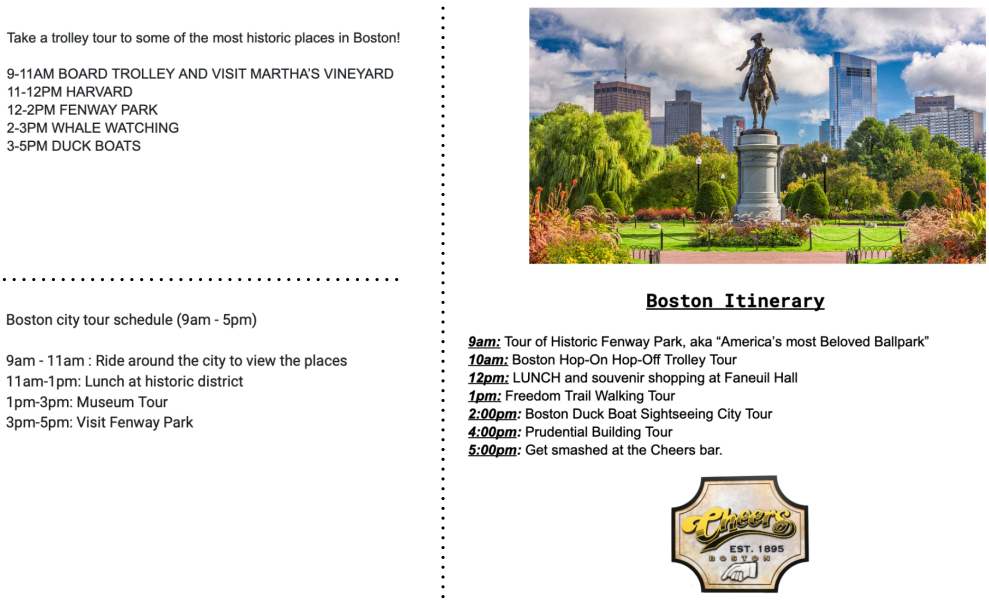

Fig. 3. Examples of itineraries: (Left) Outcomes from two teams with the lowest average score (completeness: 3.5, appropriateness: 3, creativity: 2.5). (Right) Outcome from a team with the highest average score (completeness: 5, appropriateness: 4.5, creativity: 3.5)

conflict persisted in their team, how helpful their team practices were, and their self-perceived team viability [8]. The full questionnaire is attached in the Appendix.

4.2.2 Team effectiveness. To evaluate how different interventions helped teams carry out group tasks, we asked participants to evaluate their final outcome and team. Two raters also independently rated the final outcome of the teams, blind to condition. Raters were asked to rate trip plans generated by each team based on the three criteria:

- Completeness: (Low: 1) The schedule does not fill 9am-5pm, and some events span more than 2 hours. (High: 5) The schedule is from 9am to 5pm, and each event is 2 hours or less. Also, they include some descriptions and visuals.
- Appropriateness: (Low: 1) It is not a realistic trip plan. (High: 5) It is a realistic trip plan.
- Creativity: (Low: 1) It is a list of places you always go in the new city. (High: 5) It includes some unique destinations.

Fig. 3 demonstrates some examples of trip plans and their score. The independent ratings on these trip plans had an inter-rater reliability (IRR) of 70% using Cohen's κ. We averaged the rating of the two raters to determine the final score.

4.2.3 Collaborative practices. Similar to teams' outcomes, external raters provided a score for each team's practice on a set of Likert scales, as below. Scales only had low and high ends marked, with numbers for intervening ratings. Raters were blind to conditions.

- Are the practices insightful? (Low: 1) No, it is too obvious and truism. (High: 5) Yes, it would help me determine how to do the task.
- Are the practices easy to use?: (Low: 1) No, I'm not sure how to actually apply the practices. (High: 5) Yes, it is pretty clear how I would put it into action.

Table 2. Results of self-rating by study participants. They are analyzed using a mixed effects model with random effects for teams fit by maximum likelihood. As a result, there are marginal differences between different treatments (control and experiment groups) in overall quality and appropriateness of outcomes, and significant differences in team viability (p < .01 ***, p < . 05 **, p < .1 *)

Metrics	Overall outcome		Completeness		Appropriateness		Creativity	
Fixed effects	(Intercept)	Treatment	(Intercept)	Treatment	(Intercept)	Treatment	(Intercept)	Treatment
Coefficient	12.7 ***	1.10 *	4.42 ***	.21	4.54 ***	.30 *	3.74 ***	.58
p-value	<.001	.06	<.001	.4	<.001	.09	<.001	.12

Metrics	Satisfaction		Team viability		Task conflict		Team practice	
Fixed effects	(Intercept)	Treatment	(Intercept)	Treatment	(Intercept)	Treatment	(Intercept)	Treatment
Coefficient	4.26 ***	.32	3.78 ***	.77 **	1.86 ***	-.56	8 ***	.21
p-value	<.001	.3	<.001	<.05	<.001	.28	<.001	.66

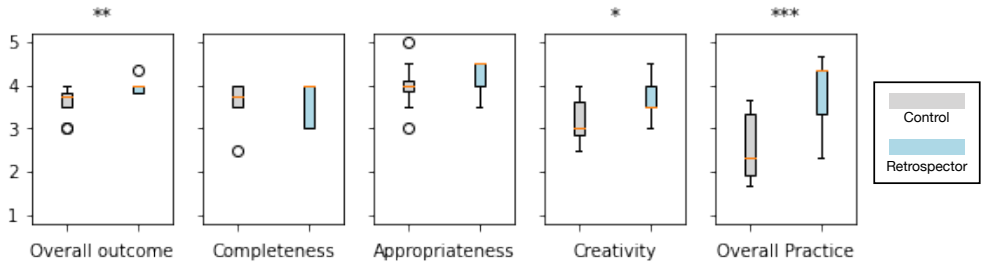

Fig. 4. Results of rating the team's outcomes by external raters. They were analyzed using a paired t-test. As a result, there were marginal differences between different treatments (control and experiment groups) in the outcomes, and significant differences in the overall quality of outcomes (combined score of completeness+appropriateness+creativity) and effectiveness of teams' practices (p < .01 ***, p < . 05 **, p < .1 *)

Table 3 demonstrates some collaborative practices generated by the teams.

4.3 Results

Three participants in a team of the experiment groups dropped out during the Katie Bell task, hence RETROSPECTOR terminates the participation for all participants in the team when the third member dropped out. As a result, eight control and six experiment groups were included in the results. We disregarded one more team from the experiment groups, as the team had a member who spoke a language other than English. We analyzed responses of the exit survey using a mixed effect model (Table 2). As for the ratings of the team's outcome and practices by external raters, we used a two-sided paired t-test (Fig. 4). We also coded the conversation using the codes listed in Table 4.

4.3.1 RETROSPECTOR teams outperformed control teams, but both groups were equally satisfied with their outcomes. The self-survey results indicate that participants from experiment groups highly self-rated their overall outcome than the control group (p=.06, Table 2. top). Among the performance metrics, RETROSPECTOR groups (μ=4.8) thought they were able to deliver more appropriate outcomes than the control groups (μ=4.5). External raters also agreed that the experiment group outperformed the control group (Fig. 4); according to the external raters, RETROSPECTOR groups were particularly more creative than the control groups.

While both participants and external reviewers pointed out the notable performance differences of teams' outcomes, RETROSPECTOR and control groups were similarly satisfied with their final

Table 3. Examples of collaborative practices that the teams came up with in our study

Control group	Retrospector group
- Come to agreements on ideas - Brainstorm - respect each others ideas - Everyone should provide input - Everyone should be communicative - Everyone should try to be as creative as possible	- I would try to find out what kind of things are fun to do there. - I would try to find out what is worth spending most of my time checking out. - I would see which team member has the most knowledge of Boston. - I would dedicate one person to create the fonts/logo.

outcomes. The average satisfaction scores of Retrospector and control groups were 4.5 and 4.3, respectively. There were no statistically significant differences (Table 2. bottom). This indicates that although the control groups realized their outcome was not great, they thought they did their best.

4.3.2 Retrospector teams were able to find effective team practices that they can apply to their teamwork. We compared effective team practices generated from each group. On average, control groups required three minutes for practice discovery, and Retrospector groups required six minutes to engage in self-reflection and practice discovery. Unlike the performance rating, there was a discrepancy between participants' self-rating and external raters' scores on team practices; participants from both the control (μ=4) and Retrospector groups (μ=4.1) rated that their team practices were similarly effective (Table 2. bottom). However, external raters rated that team practices from Retrospector groups were significantly more effective and customized to the task (Fig. 4). Table 3 lists team practices sampled from both groups; team practices from Retrospector groups tended to be appropriately specific to the tasks. Practices from the control groups were too general and unclear to actually execute.

We also observed differences between the two groups in behavior. Specifically, Retrospector groups (μ=4.4) had 15% higher team viability than the control groups (μ=3.8, Table 2. bottom). However, there was no significant difference between groups in self-perceived satisfaction and conflicts.

As for the conversational behavior, there were noticeable differences, whereas *express opinion* was most dominant in both groups (Table 4). The experiment group tended to be more implicit and self-driven rather than demanding work from other members (e.g., *share domain knowledge*, *make a change in the group work*). The control group had more explicit direction (e.g., *manage the discussion*, *ask or demand a change*). When comparing catalysts and the remaining members within teams in Retrospector groups, we found that there were no differences; most of the messages of both catalysts and others were *express opinion* and *share domain knowledge* in that order. However, in the control groups, catalysts were mostly managerial (*manage the discussion*), whereas the others were complying to the demand (*express opinion* and *share domain knowledge*). For example, there was an increase in the use of *you* significantly in the control groups more than Retrospector groups (p=.05). Catalysts of the control groups tended to rely on other team members (e.g., *can **you** elaborate please? Can **you** suggest an idea? What is **your** idea?*). We also found that when participants used the control version of the interface, there were also more usages of words in the *discrepancy* category ($p < .01$). Retrospector groups used more *conjugation*, e.g., and, but, whereas

Table 4. Result of conversation-log analysis (because of rounding, categories may not sum to 100%)

Codes	Description	Examples	Control	Experiment
Express opinion	State their subjective preference in general, or regarding a specific instance.	*"A boat tour could be fun"*	32%	36%
Share domain knowledge	Post background knowledge (i.e., Boston) related to the task.	*"well those stadiums and teams I think are landmarks of city"*	23%	30%
Manage the discussion	Take initiatives or ask for inputs or knowledge to shape the discussion. Time management also goes here.	*"We have 12 mins left"*	23%	13%
Ask for opinions	Ask a specific group member or the group in general for their preferences.	*"What do you think Stuart?"*	8%	8%
Make a change in the group work	Indicate that they will edit the itinerary.	*"I'm doing it"*	-	7%
Ask or demand a change	Ask or command another group member(s) to make changes to their group invitation.	*"please set time.."*	4%	-
Misc	The remaining messages	*"Haha"*	11%	8%

(p=.04), and their catalysts offered more direction to their members. Finally, RETROSPECTOR groups tended to use more *focus future* (p=.09), whereas the control groups used more *focus present* (p=.02). This suggests that RETROSPECTOR users tended to plan ahead, whereas control users came up with actions on the fly.

5 DISCUSSION

We presented a team-practice identifying tool, RETROSPECTOR. Our design was grounded in previous literature regarding practice diffusion in organizations and interfaces for knowledge capture. Here, we share design implications of our work and the alternative designs of discovery, application, and diffusion of group practices.

5.1 Implications for Organizational Design and the Role of Catalysts

RETROSPECTOR facilitates team reflection and leads to better team performance. When teams reflect through the guidance of RETROSPECTOR, our results suggest they are able to produce better situated practices, good outcomes, and a healthy team culture. While the effect of RETROSPECTOR is apparent externally, it is less obvious internally; team members themselves do not perceive severe conflicts or harmony in their team, and are satisfied with their poorly performed work or not satisfied with their well-performed work.

The design of Retrospector also influences the role of catalysts. We adapted catalysts from the field of multidisciplinary groups [14]. Much like previous work [14], catalysts' required effort is less demanding (*sustain* others' interactions and *assist* the construction of shared meanings) than traditional leaders' roles (*assign* and *manage* tasks and teams), but their contribution is significant. With Retrospector, catalysts act similarly to what is suggested by prior work; they collaborate and work as much as others, but in addition, they also remind other members of their collaborative practices from time to time. However, in the control version, catalysts take a managerial role that is comparable to traditional leaders. They demand work and delegate tasks to their team members. These role differences may be the result of the differing quality of collaborative practices: whereas practices from Retrospector groups were more well-defined, for groups without Retrospector, practices were vague; hence, teams relied on their catalysts to come up with strategies and plans on the fly.

Our research also suggests implications for employers of ad hoc teams. In particular, even when teams are only working on short-term projects, our work suggests employers should dedicate some part of their team's time to engaging in reflection. Doing so evidently improves the product and helps teams work better together, even if reflective activities are reasonably short. Our work also suggests that having explicit catalysts in teams and organizations can be beneficial. Catalysts entail a unique role and facilitate reflection and discovery of collaborative practices. We must caution that such role assignments may not be straightforward – in our study, they were chosen democratically by the team, rather than assigned by a manager. Future work could further investigate how catalysts should be chosen within the constraints of real-world organizations.

At the same time, organizations could benefit from having catalysts, as they are closer to the desired forms of leaders. Recent work has proposed three types of leaders in organizations [3]; low-level leaders are needed to capture and realize new and half-baked ideas from employees. Such leaders are more appropriate in the flattened hierarchy and have flexible commitment, where people without a formal position should be able to lead. Middle-level leaders are responsible for helping employees get the resources they need. High-level leaders are concerned about the structures of organizations. Having these three layers of leadership helps organizations grow and allows them to maintain a dynamic mindset. Among these three types, given the flattened hierarchy, the role, and the nature of commitment (i.e., stepping-up and stepping-down leadership), catalysts are similar to the low-level leaders. Catalysts may thus allow organizations to ensure organizational growth by reflecting and discovering new ideas and practices. However, because catalysts play a fluid role in organizations, their contributions may also be overlooked. Our work suggests how these contributions are valued (and compensated!) may impact organizational effectiveness in the long term.

5.2 Sociotechnical Design Considerations

In the following section, we discuss our design choices of Retrospector as a sociotechnical system, and how it influenced different stages of collaborative-practice generation. We present trade-offs in this design, which are the result of design considerations that apply broadly beyond Retrospector.

Practice discovery through distributed sensemaking. Through our study, we found that catalysts were able to come up with more focused and customized team practices based on the inputs from team members. In this sense, Retrospector acts as a sociotechnical system for distributed sensemaking around collaborative practices.

We hypothesize that the success of Retrospector for sensemaking is because it allowed individual members to (1) collaboratively filter their team interaction into a manageable amount of instances for catalysts to go over; and (2) share with catalysts their individual reactions to their

teams, so that catalysts have a better sense of what interaction or activity is preferred and the reasons why they were liked. This is consistent with prior work in distributed sensemaking [13]; through shared representation among team members, other members' sensemaking process could contribute to catalysts, understanding which in turn leads to effective team practices that capture teams' characteristics. RETROSPECTOR guides users to share the sensemaking process in accordance with SMM practices, which leads to distributed sensemaking.

In the current design of RETROSPECTOR, the interface asks members to pick chat messages, which is the lowest level of interaction, in the self-reflection phase. Limiting the interface to such a low-level instance might not capture general patterns of team interaction. As future work, we plan to employ higher levels of signals in team interactions. Prior work has explored different visualizations and abstractions to team sensemaking [11]. With collaborative tagging of group chat, teams can remain on the same page and decrease communication efforts [29, 58].

Another consideration of practice discovery is the starting ground. Currently, catalysts are only given inputs from other members and have to build practices from the scratch. Only guidance and specification are the categories of SMM practices. Instead, our system can provide templates, that catalysts can tweak to customize team practices.

Practice application. Our study indicates that even though external reviewers found team practices from RETROSPECTOR groups to be more effective, there is no significant difference in the self-assessed rating of team practices by participants. In other words, participants of RETROSPECTOR groups did not appreciate their own (more effective) practices, and participants of the control groups were equally satisfied with their ineffective ones.

One possible explanation is that it is possible that team members are not aware of how they benefit from underlying planning and direction provided by catalysts. On the one hand, this implies that it is not necessary for all team members to be able to actually list or feel the existence of their team practices. Instead, it may be sufficient for some leaders (namely catalysts) to manage and update team practices over time, and make sure that their team members are working according to the practices. It also suggests that the contributions of catalysts are likely invisible, and systems like RETROSPECTOR could be designed to make their work more visible, such that it is fairly compensated.

The current RETROSPECTOR system requires catalysts to carry out multiple tasks (i.e. practice discovery and application) on their own. It is also possible to imagine sociotechnical systems where catalysts share the burden of practice discovery with their team members, hence they can solely focus on application of these practices. For example, an alternative design could let practice discovery be done through team discussion so that catalysts can moderate the discussion [14]. In addition, a more elaborate catalyst-election algorithm can help teams to find more reliable and suitable catalysts for the team. The algorithm can factor in not only self-reflection by individual users but also the conversational behavior of each member.

Practice diffusion. While practice diffusion was not included in the scope of our work, our findings suggest future directions for transferring effective practices across teams. Our result implies that embedded interventions for teams to reflect on their teamwork lead to customized practices that represent the characters and dynamics of teams. Prior literature has also warned about the challenges and consequences of using someone else's practices [51]. Taken together, instead of focusing on diffusing practices, teams can diffuse theirs as a medium of communicating team cultures. As practices are highly customized for each team, one could imagine sharing theirs to explain their team dynamics to people outside of the team. Such information sharing could be useful for newcomers who are looking to join the team and judge if the team would be a good fit for them.

5.3 Limitations

Our experiment has several limitations. First, we only tested our interface for a pair of tasks. Both tasks are somewhat similar as they are both within the creativity category of McGrath's Circumplex Model [36]. We attempted to mitigate the limitation by designing two different tasks in a few parameters. The two tasks ask for different background knowledge (Harry Potter vs. A city), different outcome (Invitation vs. Itinerary), and different requirements (Food, beverage vs. Time). Additional studies with different sets of tasks and more design probes may paint a richer picture. Second, there were more control groups in our experiment than groups using Retrospector. Most participants who dropped out did so before the reflection phase, and so the dropout rate was not informative regarding the design of our tool. However, the imbalance reduced the power of our experiment. Finally, while our system helps ad hoc teams develop more effective team practices, these are not necessarily the "best" practices for the team. As O'dell and Grayson write, it is difficult to prove which practices are "best" for a team, and such proof can require both time and data [38]. This is because the effectiveness of practices is contingent on team membership and practices are collaborative behaviors learned over time. Furthermore, it is always possible (though unlikely over time) that better practices exist that the team has not experimented with. Given the short-term nature of many ad hoc teams, it may be impossible to discover the "best" practices before the team disbands, but Retrospector may still help teams find a starting place for effective practices.

In addition, the current study does not investigate whether teams can iteratively enhance practices over time. Because our decision was both pragmatic (due to challenges running long-lived tasks on Mechanical Turk) and deliberate (because ad hoc teams may not work together long enough to iterate), future work could study the benefits and limitations of iterative team reflection.

6 CONCLUSION

In this work, we introduced a novel system, Retrospector, that allows teams to engage in team reflection and discover their effective team practices. Retrospector offers a space and lens for teams to collaboratively filter and make sense of their teamwork. We employed the SMM framework and asked users to pick one or two of the best instances of planning and knowledge sharing. Retrospector then used this input to elect leaders of the team ("catalysts") and presented selected instances from their team members. Through a set of interactions guided by Retrospector, catalysts discovered and refined effective team practices. We assessed our system by conducting a lab study with online crowdworkers. Our results suggest that teams who reflected and engaged in practice discovery with Retrospector outperformed other groups and discovered effective team practices, and Retrospector teams improved team viability by around 15%. This work provides a foundation for designing systems that leverage collaborative reflection to foster the development of more effective team practices within virtual teams. More broadly, we hope that, going forward, this work signals a greater focus on leveraging a team's situated knowledge and practices as greater priorities for the design of technological supports for virtual groups.

ACKNOWLEDGMENTS

We thank Pranav Khadpe, April Wang, Theia Henderson and Farnaz Jahanbaksh for their assistance in analyzing our data. We also appreciate feedback from Julia Cambre, David Karger and Yasmine Kotturi on our early draft.

REFERENCES

[1] Mark S Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. 2013. Sharing knowledge and expertise: The CSCW view of knowledge management. *Computer Supported Cooperative Work (CSCW)* 22, 4 (2013), 531–573.

[2] Mark S Ackerman and Thomas W Malone. 1990. Answer Garden: A tool for growing organizational memory. *ACM SIGOIS Bulletin* 11, 2-3 (1990), 31–39.

[3] Deborah Ancona, Elaine Backman, and Kate Isaacs. 2019. Nimble leadership. *Harvard Business Review* 97, 4 (2019), 74–83.

[4] Michael C Beers. 1996. The strategy that wouldn't travel. *Harvard Business Review* 74, 6 (1996), 18–27.

[5] Allie Blaising, Yasmine Kotturi, Chinmay Kulkarni, and Laura Dabbish. 2021. Making it work, or not: A longitudinal study of career trajectories among online freelancers. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–29.

[6] Hancheng Cao, Vivian Yang, Victor Chen, Yu Jin Lee, Lydia Stone, N'godjigui Junior Diarrassouba, Mark E. Whiting, and Michael S. Bernstein. 2021. My Team Will Go On: Differentiating High and Low Viability Teams through Team Interaction. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 230 (jan 2021), 27 pages. https://doi.org/10.1145/3432929

[7] Sharolyn Converse, JA Cannon-Bowers, and E Salas. 1993. Shared mental models in expert team decision making. *Individual and group decision making: Current issues* 221 (1993), 221–46.

[8] Jessica Nicole Cooperstein. 2017. Initial development of a team viability measure. (2017).

[9] Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. 107–114.

[10] Amy C Edmondson, Richard M Bohmer, and Gary P Pisano. 2001. Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative science quarterly* 46, 4 (2001), 685–716.

[11] Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)* 7, 1 (2000), 59–83.

[12] Martha S Feldman. 2000. Organizational routines as a source of continuous change. *Organization science* 11, 6 (2000), 611–629.

[13] Kristie Fisher, Scott Counts, and Aniket Kittur. 2012. Distributed Sensemaking: Improving Sensemaking by Leveraging the Efforts of Previous Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 247–256. https://doi.org/10.1145/2207676.2207711

[14] Santi Furnari. 2014. Interstitial spaces: Microinteraction settings and the genesis of new practices between institutional fields. *Academy of management review* 39, 4 (2014), 439–462.

[15] William Gaver, Peter Gall Krogh, Andy Boucher, and David Chatting. 2022. Emergence as a Feature of Practice-Based Design Research. In *Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 517–526. https://doi.org/10.1145/3532106.3533524

[16] Atul Gawande. 2010. *Checklist manifesto, the (HB)*. Penguin Books India.

[17] Peter M Gollwitzer. 1999. Implementation intentions: Strong effects of simple plans. *American psychologist* 54, 7 (1999), 493.

[18] Diego Gómez-Zará, Matthew Paras, Marlon Twyman, Jacqueline N. Lane, Leslie A. DeChurch, and Noshir S. Contractor. 2019. Who Would You Like to Work With?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300889

[19] Colin M. Gray and Yubo Kou. 2017. UX Practitioners' Engagement with Intermediate-Level Knowledge. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17 Companion)*. Association for Computing Machinery, New York, NY, USA, 13–17. https://doi.org/10.1145/3064857.3079110

[20] Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. 2016. The Knowledge Accelerator: Big picture thinking in small pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2258–2270.

[21] Emily M Hastings, Farnaz Jahanbakhsh, Karrie Karahalios, Darko Marinov, and Brian P Bailey. 2018. Structure or nurture? the effects of team-building activities and team composition on team outcomes. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–21.

[22] 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.

[23] Kristina Höök and Jonas Löwgren. 2012. Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research. *ACM Trans. Comput.-Hum. Interact.* 19, 3, Article 23 (oct 2012), 18 pages. https://doi.org/10.1145/2362364.2362371

[24] Mark A Huselid. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of management journal* 38, 3 (1995), 635–672.

[25] Daniel R Ilgen, John R Hollenbeck, Michael Johnson, and Dustin Jundt. 2005. Teams in organizations: From input-process-output models to IMOI models. *Annual Revision Psychologist* 56 (2005), 517–543.

[26] Farnaz Jahanbakhsh, Wai-Tat Fu, Karrie Karahalios, Darko Marinov, and Brian Bailey. 2017. You want me to work with who? Stakeholder perceptions of automated team formation in project-based courses. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3201–3212.

[27] Aimée A Kane, Linda Argote, and John M Levine. 2005. Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organizational behavior and human decision processes* 96, 1 (2005), 56–71.

[28] Pranav Khadpe, Chinmay Kulkarni, and Geoff Kaufman. 2021. Empathosphere: Promoting Constructive Communication in Ad-hoc Virtual Teams through Perspective-taking Spaces. *arXiv preprint arXiv:2111.13782* (2021).

[29] Tae Soo Kim, Nitesh Goyal, Jeongyeon Kim, Juho Kim, and Sungsoo Ray Hong. 2021. Supporting Collaborative Sequencing of Small Groups through Visual Awareness. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.

[30] Rebecca Krosnick, Fraser Anderson, Justin Matejka, Steve Oney, Walter S. Lasecki, Tovi Grossman, and George Fitzmaurice. 2021. Think-Aloud Computing: Supporting Rich and Low-Effort Knowledge Capture. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 199, 13 pages. https://doi.org/10.1145/3411764.3445066

[31] Christina N Lacerenza, Shannon L Marlow, Scott I Tannenbaum, and Eduardo Salas. 2018. Team development interventions: Evidence-based approaches for improving teamwork. *American psychologist* 73, 4 (2018), 517.

[32] Jonas Löwgren. 2013. Annotated Portfolios and Other Forms of Intermediate-Level Knowledge. *Interactions* 20, 1 (jan 2013), 30–34. https://doi.org/10.1145/2405716.2405725

[33] Kelly Mack, Maitraye Das, Dhruv Jain, Danielle Bragg, John Tang, Andrew Begel, Erin Beneteau, Josh Urban Davis, Abraham Glasser, Joon Sung Park, et al. 2021. Mixed Abilities and Varied Experiences: a group autoethnography of a virtual summer internship. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–13.

[34] Ravindranath Madhavan and Rajiv Grover. 1998. From embedded knowledge to embodied knowledge: New product development as knowledge management. *Journal of marketing* 62, 4 (1998), 1–12.

[35] Andrew Mao, Yiling Chen, Krzysztof Z Gajos, David C Parkes, Ariel D Procaccia, and Haoqi Zhang. 2012. Turkserver: Enabling synchronous and longitudinal online experiments. In *Workshops at the twenty-sixth AAAI conference on artificial intelligence*.

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

[37] Susan Mohammed, Lori Ferzandi, and Katherine Hamilton. 2010. Metaphor no more: A 15-year review of the team mental model construct. *Journal of management* 36, 4 (2010), 876–910.

[38] Carla O'dell and C Jackson Grayson. 1998. *If only we knew what we know: The transfer of internal knowledge and best practice*. Simon and Schuster.

[39] Wanda J Orlikowski. 2002. Knowing in practice: Enacting a collective capability in distributed organizing. *Organization science* 13, 3 (2002), 249–273.

[40] Soya Park, April Yi Wang, Ban Kawas, Q Vera Liao, David Piorkowski, and Marina Danilevsky. 2021. Facilitating knowledge sharing from domain experts to data scientists for building nlp models. In *26th International Conference on Intelligent User Interfaces*. 585–596.

[41] Soya Park, Amy X. Zhang, and David R. Karger. 2018. Post-Literate Programming: Linking Discussion and Code in Software Development Teams. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings* (Berlin, Germany) *(UIST '18 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 51–53. https://doi.org/10.1145/3266037.3266098

[42] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.

[43] David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael Muller, and Felix Portnoy. 2021. How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 131 (apr 2021), 25 pages. https://doi.org/10.1145/3449205

[44] Søren Rasmussen, Jeanette Falk Olesen, and Kim Halskov. 2019. Co-Notate: Exploring Real-Time Annotations to Capture Situational Design Knowledge. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 161–172. https://doi.org/10.1145/3322276.3322310

[45] Niloufar Salehi and Michael S Bernstein. 2018. Hive: Collective design through network rotation. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–26.

[46] David Strang and Sarah A Soule. 1998. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual review of sociology* (1998), 265–290.

[47] Edward Peter Stringham. 2002. The emergence of the London stock exchange as a self-policing club. *Journal of Private Enterprise* 17, 2 (2002), 1–19.

[48] John Tang. 2021. Understanding the telework experience of people with disabilities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–27.
[49] Bruce W Tuckman. 1965. Developmental sequence in small groups. *Psychological bulletin* 63, 6 (1965), 384.
[50] 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. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3523–3537.
[51] Rachel E Watson-Jones and Cristine H Legare. 2016. The social functions of group rituals. *Current Directions in Psychological Science* 25, 1 (2016), 42–46.
[52] 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.
[53] WHO. 2008. *WHO Surgical Safety Checklist.* Retrieved July 7, 2022 from https://www.who.int/teams/integrated-health-services/patient-safety/research/safe-surgery/tool-and-resources
[54] David Sloan Wilson, Elinor Ostrom, and Michael E Cox. 2013. Generalizing the core design principles for the efficacy of groups. *Journal of Economic Behavior & Organization* 90 (2013), S21–S32.
[55] Leslie Wu, Jesse Cirimele, Kristen Leach, Stuart Card, Larry Chu, T. Kyle Harrison, and Scott R. Klemmer. 2014. Supporting Crisis Response with Dynamic Procedure Aids. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS '14)*. Association for Computing Machinery, New York, NY, USA, 315–324. https://doi.org/10.1145/2598510.2598565
[56] Longqi Yang, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, et al. 2022. The effects of remote work on collaboration among information workers. *Nature human behaviour* 6, 1 (2022), 43–54.
[57] Xiaodan Yu and Stacie Petter. 2014. Understanding agile software development practices using shared mental models theory. *Information and software technology* 56, 8 (2014), 911–921.
[58] Amy X Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.
[59] 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. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

A SELF-REFLECTION QUESTIONNAIRE

A.1 Planning

- Rate how much you agree with the following statement: In the beginning of the project, my team members spent enough time to brainstorm and define the goal of the project.
- Select one or two messages that you did that was helpful for planning your teamwork by choosing corresponding chat message(s). Or, skip if you don't think if you didn't help in planning.
- Who was the most effective team member at planning?
- Select one or two messages that [the selected member] did that was helpful for planning your teamwork by choosing corresponding chat message(s).

A.2 Sharing Relevant Knowledge

- Rate how much you agree with the following statement: Team members efficiently share enough information relevant to tasks
- Select one or two messages that you did that was effective at sharing relevant knowledge by choosing corresponding chat message(s). Or, skip if you don't think if you didn't help in sharing relevant knowledge.
- Who was the most effective team member at sharing relevant knowledge?
- Select one or two messages that [the selected member] did that was effective at sharing relevant knowledge by choosing corresponding chat message(s).

B EXIT-SURVEY QUESTIONNAIRE

All of the questions have five options – strongly disagree, disagree, neutral, agree and strongly agree.

- (If leader) During Boston city planning, as a leader, I was able to come up with instructions that team members could easily follow.
- (If leader) During Boston city planning, as a leader, I was able to come up with useful instructions to improve teamwork.
- (If not leader) During Boston city planning, instructions from team leaders were easy to follow.
- (If not leader) During Boston city planning, instructions from team leaders were useful.
- I am satisfied with my team's final outcome of Boston city planning.
- Our team's Boston city planning is complete.
- Our team's Boston city planning is appropriate.
- Our team's Boston city planning is creative.
- Most of the members of this team would welcome the opportunity to work as a group again in the future.
- As a team, this work group shows signs of falling apart.
- The members of this team could work for a long time together.
- There was a lot of conflict of ideas in our group.

Received July 2022; revised January 2023; accepted March 2023