Working Harder but not Smarter: Experimental Results on the Effects of Collective Intelligence Awareness

Chase McDonald
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Thuy Ngoc Nguyen
University of Dayton
Dayton, Ohio, USA

Carlos Botelho
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Christopher Dishop
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Anita Williams Woolley
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
a2w@cmu.edu

Cleotilde Gonzalez
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
coty@cmu.edu

ABSTRACT

A growing number of studies have demonstrated that collective intelligence is an important factor in team success. Recent studies have also demonstrated the utility of real-time collaborative process metrics for measuring and predicting collective intelligence as teams work together, opening the possibility of their use to guide interventions to improve team performance. We report the results of an experiment in which teams collaborated on a search and rescue task online while some were randomly assigned to see real-time displays of their team's collective effort – the collaborative process metrics our pilot studies showed most strongly predicted task performance. However, we find that providing this information does not reliably improve team performance, in some cases leading teams to alter behavior to maximize effort to the detriment of other processes and performance. We discuss the implications of our results for the design of digital nudges as interventions into collaborative processes and collective intelligence.

CCS CONCEPTS

• **Applied computing → Psychology**; • **Human-centered computing → Empirical studies in collaborative and social computing**.

KEYWORDS

collective intelligence, human teams, collaboration, teamwork

1 INTRODUCTION

Collective intelligence is the ability of a group to perform a wide range of tasks or achieve a wide range of goals in different environments that vary in complexity [11, 12]. Extant work has demonstrated that measures of a team's collective intelligence are predictive of their future performance [2, 7], as well as real-time measures of collaborative processes captured as groups are working to predict collective intelligence [4, 9, 13]. This raises the question of whether real-time collaborative process metrics can be used as an intervention to enhance collective intelligence as teams collaborate.

There is a reasonable basis to expect that interventions that raise awareness of collaborative processes in teams should lead to better performance. Widely-accepted frameworks for team coaching [6] supported by empirical evidence [10**?**] demonstrate that providing real-time feedback to teams as they collaborate leads to significantly better performance. This has been supported by related studies incorporating technology-based interventions including real-time feedback displays focused on relative speaking time and task engagement [1, 3] should be beneficial. However, there is other evidence demonstrating that attempts to make such real-time technology-based interventions can be more mixed [4]. In the experiment we report here, we examined whether making team members aware of the quality of their collaboration by displaying real-time collaborative process metrics as they worked could improve collective intelligence and collaboration. We specifically focused on the effect of collaborators' awareness of collective effort, using a measure developed and validated in recent research [9] and which our pilot tests demonstrated as a strong predictor of performance on our task. We randomly assigned some collaborators to see real-time displays of their collective effort as they worked. As we report, we found that the intervention did not reliably improve performance, and in some cases, led collaborators to overly focus on increasing their collective effort at the expense of their overall performance.

2 METHODS

Team Minimap

We conducted our experiment using a multi-agent implementation of a Search and Rescue task in the Minimap environment developed by Nguyen and Gonzalez [8], herein referred to as the Team Minimap task. In this setting, a team of four human players navigated a grid-like environment to search for and rescue victims of three types: minor, serious, and critical. The team members were assigned one of two distinct roles: medics and engineers. Medics were able to rescue minor, serious, and critical victims; however, an engineer also had to be adjacent to a critical victim for it to be rescued. Engineers, on the other hand, could only rescue minor victims on their own but had the ability to open doors and clear the rubble that surrounded the serious victims. The team's goal was to maximize the points earned by collaborating to rescue as many victims as possible. The layout of the environment and victim locations is held constant and contains X minor, X serious, and X critical victims.

2.1 Measures of Collective Intelligence

Building on a team effectiveness framework introduced by Hackman [5], Riedl et al. [9] demonstrated three real-time collaborative process measures are significant predictors of collective intelligence, capturing teams' use of member knowledge and skill, the quality of

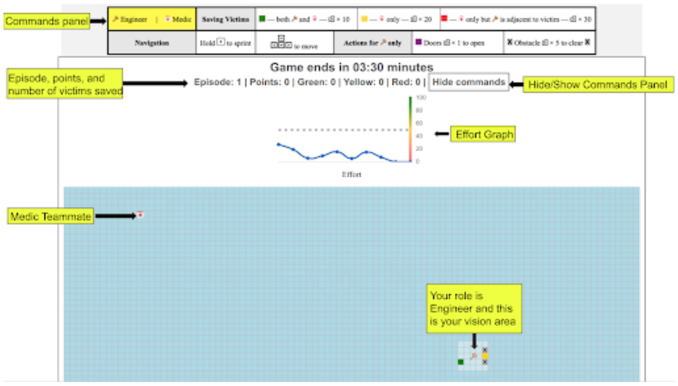


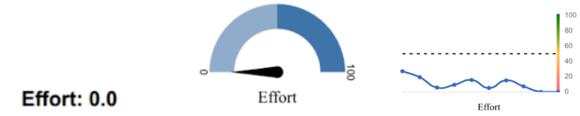
Figure 1: Task setting for the "Line" display condition, as shown in our task instructions to participants. Participants only have a limited view of their immediate surroundings in the environment, but can also see the positions of their teammates.

their task strategy, and the sufficiency of collective effort. Building on that research, we have implemented these metrics in other task contexts, including an online Search and Rescue game [13], and we refer to them collectively as Team Effectiveness Diagnostic (TED) metrics.

In this experiment, we focused on the TED metric capturing collective effort. In developing this TED metric, we adapted its operationalization to a search and rescue task environment to capture the behaviors in this context that signify effort. In this setting, we formalize effort as a linear combination of the number of victims saved, doors opened, rubble cleared, and cells moved by each team member. In our calculations, we weight some activities more than others; each victim saved is weighted 20 times higher than other inputs, and the number of cells moved is scaled for each player based on the maximum number of cells moved over the measurement period. The calculation of effort is made in real-time, captured in each three-second interval.

2.2 Visual Displays

Our experiment included four different conditions, including a control condition plus three different variations on a display of the real-time TED Effort metric. In our control condition (No TED), we provide no information about collective effort to participants. In the three display conditions, depicted in Figure 2, we show the TED Effort metric, updated in three-second intervals, depicted in either a numeric, line graph, or gauge format. In both the gauge and numeric display formats, only the current value of TED Effort is shown, whereas the line graph shows the trend based on the most recent 10 values. We elected to incorporate conditions with different display formats in order to distinguish the effect of the information alone (i.e., the Numeric condition), versus a display including a normative reference point (i.e., the Gauge condition), or a display demonstrating relative progress (i.e., the Line condition).



(a) Numeric **(b) Gauge** **(c) Line**

Figure 2: The three different TED display conditions each have a distinct representation of the TED Effort.

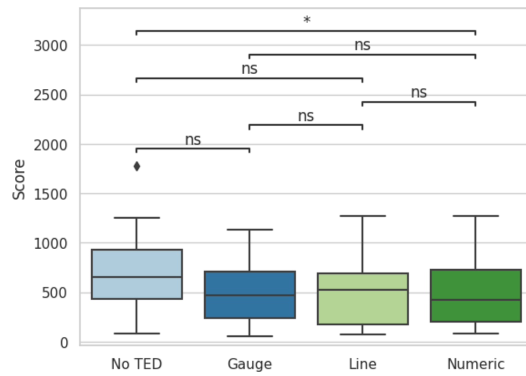


Figure 3: Score, as expressed by the total number of points earned, by condition.

3 RESULTS & DISCUSSION

We recruited participants via the Amazon Mechanical Turk platform, resulting in complete data from 436 participants. These participants were randomly assigned to groups of four in each condition. This resulted in 32 groups in the No TED condition, 25 groups in the Gauge condition, 27 groups in the Line condition, and 25 groups in the Numeric condition. All teams completed two 5-minute episodes of the task and were compensated with a base payment in addition to a bonus proportional to the number of points earned by their team. Participants in the display conditions were not informed how the effort was calculated, solely that it measured the collective effort of their team.

The average score over both episodes, as measured by the number of points earned, is shown for each condition in Figure 3. We observe no significant difference between any of the TED display conditions, and only a marginally-significant difference ($p < 0.05$) between the Numeric display condition and the No TED control condition; no other conditions were significant. This result tells us that providing the supplementary TED information did not directly improve team performance, and in the case of the Numeric condition resulted in a marginal decrease.

To further explore our findings, we examined our data for evidence of whether or not the participants were aware of the feedback displays by evaluating if they altered their behavior in any systematic way. We observed that the participants in the TED Effort display conditions all registered significantly higher values of effort than participants in the No TED control condition, shown in Figure 4. We interpret this pattern as indicating that participants noted the

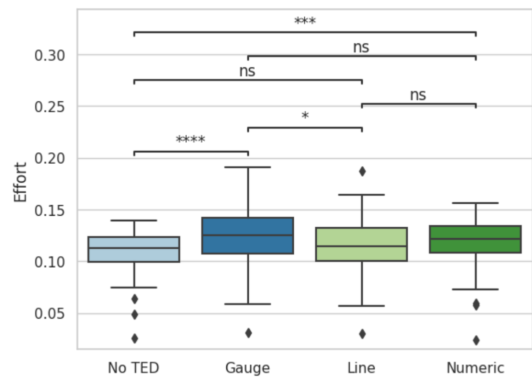


Figure 4: Average team effort per 3-second interval by condition. Effort is scaled via min-max scaling to be in the $[0, 1]$ **range.**

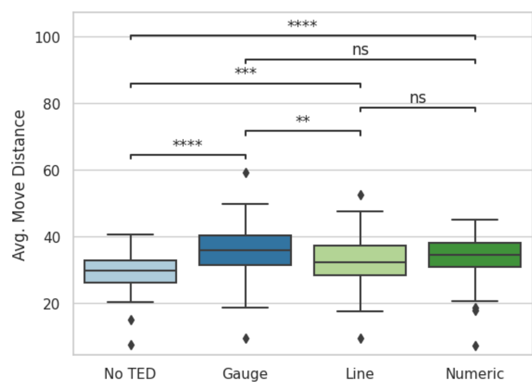


Figure 5: Average number of cells covered by each team per three-second interval.

information, and responded, but unfortunately not in a way that improved performance. We further broke down the TED Effort metric into its components to better understand how participants changed their behavior: the only significant difference in behavior between the NO TED control condition and the others was in the distance moved, which is shown in Figure 5. Although participants were not explicitly told which behaviors resulted in changes in the TED Effort values, it appears that participants learned the association and moved more to increase the displayed value.

In summary, we are building on extant research on collective intelligence to explore interventions to improve collaboration. Our results demonstrate that real-time displays of TED metrics can shape collaborative behavior, but also come with a cautionary note: participants may adopt behaviors that are ultimately not conducive to the maximization of any focal metrics. This is not entirely surprising, given the history of social science research demonstrating how readily behavior can be shaped by focusing on particular performance metrics. However, our pilot tests, some including displays of all three TED metrics (including use of knowledge, skill, and efficiency of task strategy, in addition to effort) suggested that the amount of information was too complex for participants to process while working on the task, and thus had little impact on their behavior. Thus the question of how to increase collective intelligence via feedback to collaborators remains open. The successful interventions demonstrated in extant research have involved tasks that were more heavily influenced by a single type of collaborative process (such as using primarily information-sharing; e.g., [1]) or the intervention stimulated a more integrative, team-level consideration of collaborative process (such as the successful nudges in [4]) versus individual-level adjustments of individual behaviors which can result in the over-correction we observed here. Therefore, it is likely that technologically-administered interventions to improve collective collaborative processes need to facilitate adjustments to collective cognition in order to drive change in collective intelligence.

REFERENCES

[1] Joan Morris DiMicco and Walter Bender. 2007. Group reactions to visual feedback tools. In *International Conference on Persuasive Technology*. Springer, 132–143.

[2] David Engel, Anita Williams Woolley, Ishani Aggarwal, Christopher F Chabris, Masamichi Takahashi, Keiichi Nemoto, Carolin Kaiser, Young Ji Kim, and Thomas W Malone. 2015. Collective intelligence in computer-mediated collaboration emerges in different contexts and cultures. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3769–3778.

[3] Ella Glikson, Anita Williams Woolley, Pranav Gupta, and Young Ji Kim. 2019. Visualized automatic feedback in virtual teams. *Frontiers in Psychology* (2019).

[4] Pranav Gupta, Young Ji Kim, Ella Glikson, and Anita Williams Woolley. in press. Digitally nudging team processes to enhance collective intelligence. *MIS Quarterly* (in press).

[5] J. R Hackman. 1987. *The design of work teams.* Prentice-Hall, Englewood Cliffs, NJ.

[6] J. R. Hackman and Ruth Wageman. 2005. A theory of team coaching. *Academy of Management Review* 30, 2 (2005), 269–287. http://amr.aom.org/content/30/2/269.short 00520.

[7] Young Ji Kim, David Engel, Anita Williams Woolley, Jeffrey Yu-Ting Lin, Naomi McArthur, and Thomas W Malone. 2017. What makes a strong team? Using collective intelligence to predict team performance in League of Legends. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 2316–2329.

[8] Thuy Ngoc Nguyen and Cleotilde Gonzalez. 2022. Minimap: an interactive dynamic decision making game for search and rescue missions. *Available at SSRN 4123762* (2022).

[9] Christoph Riedl, Young Ji Kim, Pranav Gupta, Thomas W Malone, and Anita Williams Woolley. 2021. Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences* 118, 21 (2021), e2005737118.

[10] Anita Williams Woolley. 1998. Effects of intervention type and timing on group task performance. *Journal of Applied Behavioral Science* 34 (1998), 30–46. Woolley 1998.

[11] 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.

[12] Anita Williams Woolley and Pranav Gupta. in press. Understanding Collective Intelligence: Investigating the Role of Collective Memory, Attention, and Reasoning Processes. *Perspectives on Psychological Science* (in press).

[13] Michelle Zhao, Fade R. Eadeh, Thuy-Ngoc Nguyen, Pranav Gupta, Henny Admoni, Cleotilde Gonzalez, and Anita Williams Woolley. 2023. Teaching agents to understand teamwork: Evaluating and predicting collective intelligence as a latent variable via Hidden Markov Models. *Computers in Human Behavior* 139 (2023), 107524.