The Entropy of Communication Turn Taking during a Collaborative Problem-Solving Task

Hamish Stening (hamish.stening@mq.edu.au)
Department of Psychology, Macquarie University
Sydney, NSW, Australia

Jayna Lee (jayna.lee@students.mq.edu.au)
Department of Psychology, Macquarie University
Sydney, NSW, Australia

Jan Groeschel (jan.groeschel@students.mq.edu.au)
Department of Psychology, Macquarie University
Sydney, NSW, Australia

Professor Michael J Richardson (michael.j.richardson@mq.edu.au)
Department of Psychology, Macquarie University
Sydney, NSW, Australia

Associate Professor Rachel W. Kallen (rachel.kallen@mq.edu.au)
Department of Psychology, Macquarie University
Sydney, NSW, Australia

Abstract

Collaboration and teaming are critical for solving complex problems. However, little is known about how group dynamics affect teaming behaviours and, ultimately, problem-solving effectiveness. The present study aimed to validate a novel measure of the dynamics of team communication – here termed turn-taking entropy – and to investigate what aspects of those dynamics affect collaborative-problem-solving performance. Thirty-two teams of 4 were asked to complete a simulated crisis-response task in which they had to rank 15 items in order of their importance to their team's survival (first individually and then as a team). Group responses were better than the aggregated individual responses of team members (suggesting teaming benefits), and were better when team members had task-relevant skills and knowledge. However, response quality was not significantly related to task completion time. Additionally, the proposed entropy measure appeared to capture group communication dynamics, and appeared to differentiate stable an unstable patterns of communication. Implications and directions for future research are discussed.

Keywords: teaming; problem solving; turn taking; entropy; Lost at Sea.

Introduction

By working collaboratively, agents can often work more efficiently or achieve better outcomes than if they had worked independently (for example, Barron, 2000; Hill, 1982; Johnson & Johnson, 2009; O'Donnell, 2006; Saleh et al., 2005). It is this very feature of collaboration that makes teaming such an effective strategy for solving a wide variety of complex tasks (Graesser et al., 2018). Indeed, teams are used to maximise the quality of medical care (Baker et al., 2006; Ervin et al., 2018; McKee & Healy, 2002), to solve complex design and engineering problems (Galbraith, 2009; Reiter-Palmon & Leone, 2019), and to meet the complex demands of military operations (Goodwin et al., 2018; Shuffler et al., 2012). Given the increasing number and complexity of problem-solving challenges found in modern workplaces (Gatewood et al., 2015; Hoffman et al., 2020; World Economic Forum, 2020) and the increased dependence on teams to solve these problems (Chowdhury & Murzi, 2020; Hall et al., 2018), understanding the dynamics, mechanisms, and necessary precedents of effective teaming is more important than ever (Graesser et al., 2018; RamosVillagrasa et al., 2018; Wiltshire et al., 2018).

Some researchers have suggested taking a Dynamical Systems Theory or complex-adaptive-systems approach to investigating teams (for example, Gorman et al., 2017). However, dynamical modelling of any system requires defining system variables that best capture the behaviour/dynamics/team characteristic of interest. In the past, some researchers have used the ratio of the latency between regular communication functions (Gorman et al., 2010), the Shannon entropy (Shannon, 1948) of communication functions (Wiltshire et al., 2018), the frequency and variety of turn-taking patterns (Hoogeboom and Wilderom, 2020), and dynamic complexity (Schiepek et al., 2010; Wiltshire et al., 2021), just to name a few examples.

Many complex adaptive systems, including problem-solving teams, over time exhibit qualitatively distinct stable phases in their behaviours (for example, Barnosky et al., 2012; Folke et al., 2004; Hughes et al,. 2013; Scheffer, 2020; Wang et al., 2009; Zhou et al., 2011). While transitions between these phases can be gradual (for example, Heerklotz & Tsamaloukas, 2006; Shimizu et al., 2017; Song et. al, 2008), often the transitions are sudden, nonlinear, and unpredictable (May et al., 2008; McSharry et al., 2003; Scheffer et al., 2009; Venegas et al., 2005). Wiltshire et al. (2018) theorised that teams' stable communication patterns emerge as a direct result of the constraints placed on the teams' components. During stable phases, there is large constraint on the system, and a large basin of attraction towards a particular system state (or pattern of team-member interactions). In such cases, the system will recover quickly from minor perturbations or disturbances (such as changing task contexts or new ideas being suggested by team members; Scheffer et al., 2009) which can be seen in high autocorrelation or low variability (Dakos et al., 2012; Scheffer et al., 2009), low dynamic complexity (Schiepek &

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

Strunk, 2010; Wiltshire et al., 2021), and low entropy (Likens et al., 2014; Wiltshire et al., 2018) in measures of the team's system state. Conversely, approaching or during the transitions between stable phases, there is less constraint on the system and a smaller (or flatter) basin of attraction, and the system will recover more slowly from perturbations (a phenomenon known as "critical slowing down"; Scheffer et al., 2009), if indeed it does recover at all. As such, approaching or during a phase transition, CPS teams may exhibit critical instability (Wiltshire et al., 2021) which may be seen in lower autocorrelation, greater variability, greater dynamic complexity, and greater entropy in the teams' dynamics.

One aim of the present study was to validate assumptions regarding collaborative problem solving (specifically, that the processes of teaming provide non-linear performance benefits above solely the aggregation of knowledge and skills). Another aim was to validate a new measure for capturing the dynamics of team communication, here termed "turn-taking entropy", that represents the unpredictability of the speaking order of team members, and that can be calculated without the need for transcription and semantic coding of communication (as was the case in Wiltshire et al., 2018) or EEG recording equipment (as was the case in Likens et al., 2014). Additionally, the present study investigated how aspects of those dynamics affect teaming behaviours and, ultimately, problem-solving performance. To do this, teams of participants were asked to complete the same task as in Wiltshire et al. (2021)'s study, the Lost at Sea task (a widely used crisis-response simulation task in which participants are asked to imagine that they are stranded on a lifeboat in the middle of the ocean, and to rank 15 items in order of their importance to the team's survival; Neminoff & Pasmore, 1975).

It was hypothesized that teams' rankings lists would be closer to the "correct" rankings than the lists aggregated from their members' individual lists (Hypothesis 1). However, it was also expected that team members' individual-task ability would still positively contribute to group-task performance (Hypothesis 2). Furthermore, it was hypothesized that the longer that teams worked together and the more that members exhibited teaming behaviours, the greater the teaming benefits and the better they would perform on the Lost at Sea task (Hypothesis 3).

It was hypothesized that turn-taking entropy would vary across teams and across communication, and would exhibit clear peaks and troughs (Hypothesis 4). It was also hypothesized that higher entropy would be associated with longer task completion time and increased performance (Hypothesis 5). It was hypothesized that the more that teams reorganised (both in terms of the magnitude of the reorganisation and in terms of the number of reorganisations), the longer they would take to complete the task, but the better the quality of their final list (Hypothesis 6), and finally, it was hypothesized that the faster that teams reorganised, the less time that they would take to complete the task (Hypothesis 7).

Method

Participants

Data were collected for 32 four-person teams (128 participants) as part of a larger study on collaborative problem solving. Participants were first-year psychology students from Macquarie University who signed up for the study as individuals using the university's SONA system in exchange for course credit. The majority of participants (116) had never interacted with any of their teammates prior to the study, but 8 teams contained at least two team members that were at least familiar acquaintances.

Ninety-nine participants identified as female, 28 identified as male, and 1 identified as non-binary or third gender. 15 teams comprised of only females and 1 team comprised of only males. 60 participants identified as White/Caucasian, 33 identified as Asian, 2 identified as African, 1 identified as Aboriginal Australian, 1 identified as Pacific Islander, and 31 identified as Other. Ages ranged from 17 to 49 ($M = 19.8$, $SD = 4.7$).

Materials

Lost at Sea Task. The Lost at Sea task is a crisis-response simulation in which participants are asked to imagine that they have been stranded on a lifeboat in the South Pacific after their yacht and most of its contents have been destroyed in a fire. The participants are then tasked with ranking a list 15 items in order of importance to their survival (Nemiroff & Pasmore, 1975). In our version of the task, participants read the task instructions and completed their rankings in a virtual presentation of the task (designed using the cross-platform Unity game engine, version 2020.3.17f1 LTS; Unity Technologies, San Francisco, California) displayed on touchscreen tablet computers (Microsoft Surface Pro 9). On their tablets, participants were able to access/hide task instructions, drag items into ranking positions, see how much longer they had to rank the items, and, in the group-ranking portion of the task, see their team's rankings list.

Task performance was operationalized as ranking accuracy (or more specifically, as the Spearman's rank correlation, ρ, between an individual's/team's ranking list and the "correct" rankings list that was developed by the US Merchant Marines; Nemiroff & Pasmore, 1975). Task performance was measured both for final submitted rankings, but also for each update/change to the ranking list. Because the magnitude of changes in ρ was inversely proportional to the number of items in a team's\individual's list at the time of the update (ρ would increase more for a "correct" addition of a third item to a two-item list than for a "correct" addition of a fifteenth item to a fourteen-item list), changes to ranking accuracy were coded as negative (-1), neutral (0; when ρ did not change), or positive (1).

Finally, to aggregate team members' individual rankings, the ranking positions of each item for each member where summed and then ranked in order of this sum. ρ was calculated for the aggregated rankings, and "teaming benefits" for each team were operationalised as the difference between the team's group ρ and the ρ for its aggregated rankings.

Recording Equipment. Audio was recorded using four Shure MX53 over-ear omni-directional headset microphones and four Audio-Technica System 10 PRO wireless transmitters/receivers feeding into a Zoom LiveTrak L-12 audio recorder. Video was recorded using four Razer Kiyo Ring Light Equipped Broadcasting Cameras (RZ19-02320100-R3M1) feeding individually via USB 3.0 into a custom HP computer (Intel Corei9-9900K CPU, NVIDIA GeForce RTX 2080Ti GPU) running the OBS Studio (27.2.3) multi-channel recording software.

Procedure

Participants were randomly allocated to teams by signing up to research time slots. The study design was correlational, and no experimental manipulations were made. On arrival, participants were fitted with microphones, wireless transmitters, and name tags displaying a unique colour (Blue, Red, Green, or Grey, printed in large, coloured, Calibri size 80 font) such that participants were able to refer to each other by either colour or their real names. They sat at a round table (in front of their tablet and such that the team members were facing each other), gave consent, completed a demographic information survey also hosted in Qualtrics, and began the Lost at Sea task.

Participants were given a short verbal description of the task before being instructed to read the full task information presented in the virtual Lost at Sea environment. Participants were given 10 minutes to complete their individual rankings, starting from the moment that they hid the task instructions and accessed the ranking screen. When all team members had finished their individual rankings, they were given 20 minutes to complete the group rankings, starting from the moment the experimenter remotely opened the group-ranking instructions simultaneously on all team members' screens.

During the tasks, the experimenter observed the participants from an adjoining room via the video camera feed. Participants were not allowed to talk to each other before, between, or after the ranking tasks, except when completing the group ranking.

Data Processing

Speaker coding. Using Adobe After Effects (18.4.1), animated 495ms-duration audio waveforms for the microphone audio for each team member were generated and laid over the OBS Studio videos of the four camera feeds for each team. This was done to make it clear which team members were speaking and which team members were silent during periods of cross-talk (i.e., when one member's microphone picked up the sound of another member talking). Four research assistants were trained on how to label the outputted videos using Prodigy (v1.11.8; Python 3.10.0) annotation software, and one assistant per video recorded speaker, start time, and end time for each speech utterance (i.e., speaking turn). Interjections such as "um", "oh", and "mmhm", as well as any vocalized laughter, were coded as utterances/turns, whereas sighs, heavy breaths, and non-vocalized laughs were not. All videos and labels were then reviewed by the primary investigator for errors.

To determine who was the primary speaker (i.e., whose turn it was) during periods of cross talk, each list of speaking utterances was first converted into four separate time series (one for each team member) of whether that team member was (1) or was not (0) speaking (at each 30ms window of the task) using custom Python (3.10.0) code. Then, more Python code was used to merge each set of four time series back into a list of speaker changes that occurred during the task. A change of speaker was determined to have occurred every time a new speaker started a new utterance (even if they overlapped the previous speaker). If, however, the overlapping speaker finished their utterance before the previous speaker finished theirs, a speaker change back to the original speaker was deemed to have occurred at the end of the new speaker's utterance. Additionally, a change of speaker to "No Speaker" was recorded if there was 3s or more of no team members talking.

Finally, each list of speaker changes was converted into a list of speaking turns taking the five values "Blue", "Red", "Green", "Grey", and "No Speaker". Additionally, each list was subdivided into lists of turns during each 60s sliding window (using a step size of 1s) across the task.

Turn-Taking Network Graphs. Treating speaking turns as system states, each list of turns was converted into a Markov chain and transition matrix. Then, each chain was visualized as a 5-node bidirectional network graph using the Matplotlib (3.6.3; Hunter, 2007) and NetworkX (3.0; Hagberg et al., 2008) Python packages (see Figure 1 for examples). In the graphs, the nodes represented who was speaking (treating "no speaker" as a speaking option) during a turn (i.e., the state of the system), and the proportional area of the nodes represented the proportional frequency of each speaker/silence in the list of turns that the graph was constructed from. The arrows between nodes, on the other hand, represented the one-step transition probability of who would be the next speaker in the next turn, with the thickness of the arrow's lines corresponding to the magnitude of this probability. Unlike traditional Markov chains, however, the speaker of a turn could not be followed by themself given how turns were defined. That is, states in the chain could not be recurrent.

Turn-Taking Entropy. Shannon entropies (Shannon, 1948) were calculated for the lists of turns (both for the turn taking across the entire group task, and for the turn taking within each possible 60s-wide sliding window using a slide step of 1s; i.e., for each matrix) using the equation:

$$\mathrm{H}(\boldsymbol{X}) = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{n}} \mathbf{P}(\boldsymbol{i})\, \mathbf{H}\left(\boldsymbol{X}_{\boldsymbol{j}|\boldsymbol{i}}\right) \tag{1}$$

where H(X) is the entropy of the next speaker j given the current speaker i is known, i and j have possible outcomes "Blue", "Red", "Green", "Grey", and "no speaker", P(i) is the probability that i was the current speaker (i.e., the diagonals in the turn-taking matrix), and $H(X_{j|i})$ is the entropy of who will follow i with equation:

$$H(X_{j|i}) = \sum_{j=1}^{n} P(j|i) \log_2 P(j|i) \qquad (2)$$

where $P(j|i)$ is the probability that speaker j would follow speaker i (i.e., the off-diagonals in the turn-taking matrix). The minimum theoretical value for $H(X)$ for a four-person team – which would occur when there was only one speaker or just silence during the observation window – is 0, and the maximum theoretical value – which would occur when all speakers and silence had equal numbers of utterances – is 2.32.

Higher values in turn-taking entropy correspond to greater symmetry in the group's turn taking, as well as lower predictability of who will be the next speaker. Hence, higher entropy values are indicative of less constraint on the turn-taking dynamics. Conversely, lower entropy values represent less symmetry, less uncertainty, and more constraint.

Peaks in Entropy. The sliding-window entropies for each team were combined into single time series. Using Python code, these series were smoothed using a Gaussian kernel with a standard deviation of 10 and a window size of 50. Then, local maxima and minima were identified in the smoothed time series using find peaks function from the SciPy (1.10.1) signal package (Virtanen et al., 2020). These peaks indicate critical instabilities in the team's communication dynamics, where the team's behavioural patterns and structural organisation are more susceptible to perturbations. The periods between peaks correspond with stable phases where the behavioural patterns and structural organisation are more resistant to perturbations (see Figure 1).

Results

Group-Task Performance

Individuals' ranking-accuracy scores (ρ; n = 128) were approximately normally distributed ($\gamma = 0.10$, $\kappa = 2.74$) with a mean of .167 (SD = .204) and a range of -.307 to .714. Groups' ρ scores (n = 32) were also approximately normally distributed ($\gamma = 0.38$, $\kappa = 2.81$) with a mean of .248 (SD = .178) and a range of -.067 to .682. Teaming-benefit scores (i.e., the differences between aggregated-list ρ and group ρ), on the other hand, were positively skewed ($\gamma = 0.92$) and platykurtic ($\kappa = 5.91$), and ranged from -.28 to .65 with a mean of .06 (SD = .17) and median of .07.

Dissimilarly, individuals' task-completion times (n = 128) were left skewed ($\gamma = -0.39$) and leptokurtic ($\kappa = 2.25$), with a mean of 226.4s (SD = 51.1) and a range of 102 to 300s. Groups' task completion times (n = 32), on the other hand, had a symmetric ($\gamma = 0.03$), u-shaped distribution ($\kappa = 1.64$), with a mean of 693.8s (SD = 305.9) and a range of 219 to 1195s. This suggests that no team found the 20min group-task time limit too short and that the observation period was long enough to observe the natural group dynamics.

Hypothesis 1 (average teaming benefits would be greater than 0) was tested using nonparametric bootstrap estimation of the mean. Supporting the hypothesis, the test revealed that the mean was significantly greater than 0 ($Z = 2.27$, $p = .023$). This suggests that collaboration and discussion improved task performance above solely sharing and averaging team members' initial individual conclusions.

Hypothesis 2 (average individual task ability would predict group performance) was tested using an OLS regression of group ρ on individual ρ (averaged across team members). In support of the hypothesis, higher individual ρ was significantly associated, on average, with higher group ρ (b = 0.84, SE = 0.26, $t(30) = 3.28$, $p = .003$), suggesting that the better team members performed on average in their individual tasks, the better they performed in the group task.

Hypothesis 3, on the other hand, was tested with two separate regressions. First, to determine whether working longer at the task would improve performance, group ρ was regressed on group-task completion time. Second, to determine whether teaming behaviours (operationalised as updating the rankings as updates were indicative of challenging others' perspectives and/or revisiting ideas) led to better task performance, group ρ was regressed on the number of updates to the list during the task. The regressions revealed that time taken to complete the task was not significantly related to task performance ($b < 0.01$, $SE < 0.01$, $t(30) = 0.74$, $p = .465$) but number of updates was ($b = 0.006$, $SE = 0.003$, $t(30) = 2.14$, $p = .040$). The more times a team updated their list, the more accurate, on average, their list was. Therefore, Hypothesis 3 was partially supported by the data.

Turn-Taking Entropy

Entropy across the entire group task (n = 32) was left skewed ($\gamma = -1.26$) and platykurtic ($\kappa = 5.66$) with a mean of 1.54 (SD = 0.11) and a range of 1.18 to 1.71 (50.9 to 73.6% of maximum possible entropy for a four-person group). Sliding-window entropy (n = 22,595), on the other hand, was also left skewed ($\gamma = -1.66$) and platykurtic ($\kappa = 6.78$). It had a mean of 1.24 (SD = 0.28) and a range of 0 to 1.72 (0 to 74.1% of maximum possible entropy). Clearly, in support of Hypothesis 4, turn-taking entropy varied across teams and across communication.

The number of peaks (local maxima) in the smoothed entropy time series appeared to be randomly distributed, however when controlling for the time taken by the group to complete the task, the distribution became approximately normally distributed ($\gamma = -0.40$, $\kappa = 2.69$). Whereas the absolute number of peaks ranged from 4 to 16 (M = 9.7, SD = 4.1), the peaks per minute ranged from 0.55 to 1.10 (M = 0.86, SD = 0.14). Also in support of Hypothesis 4, sliding-window entropy clearly exhibited peaks and troughs. Visual inspection of the network graphs at the times of peak and troughs revealed that teams' communication patterns differed between periods of stable dynamics and critical instabilities, and often between different periods of stable dynamics (for an example, see Figure 1).

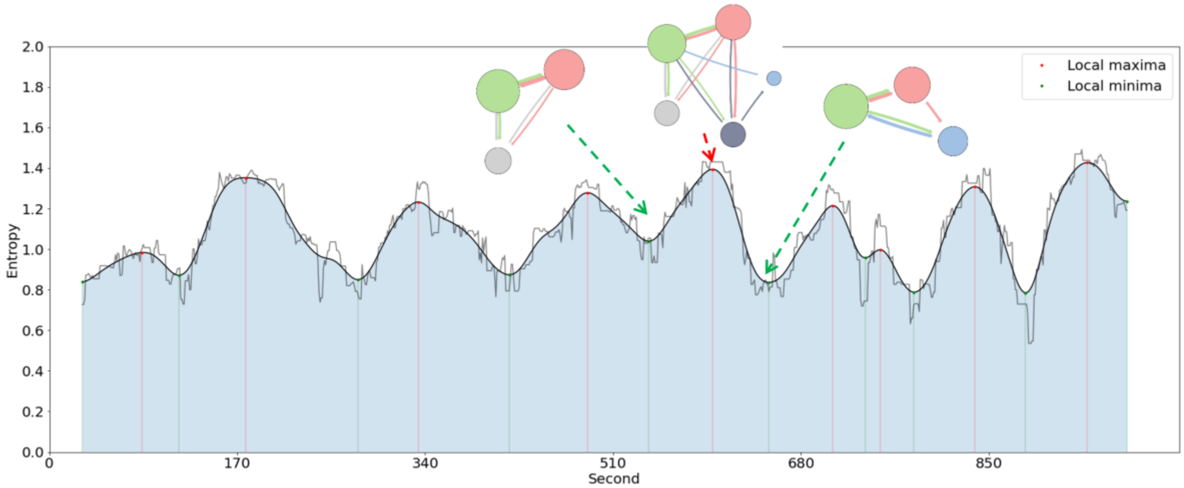

Figure 1: The entropy time series for one team during the group task. Network graphs of the turn-taking dynamics at and either side of a critical instability have been superimposed over those points.

Hypothesis 5 (higher entropy would be associated with longer task completion time and greater ranking accuracy) was tested with three regressions. First, group-taskcompletion time was regressed on whole-task entropy. Contrary to what was hypothesized, whole-task entropy did not significantly predict task time (b = 156.13, SE = 527.43, t(30) = 0.30, p = .769). Second, group ρ was regressed on whole-task entropy. Again contrary to what was hypothesized, whole-task entropy did not significantly predict group ρ (b = 0.16, SE = 0.31, t(30) = 0.53, p = .600). Finally, to assess whether sliding-window entropy at the time of an update to a team's ranking list affected whether that update improved (Δρ > 0, n = 391) or worsened the list (Δρ < 0, n = 345), a multilevel logistic model (predicting improvement/no improvement with a fixed effect of entropy across the window 30s before and 30s after the update and a random effect of team) was fit. Because logistic distributions model only binary outcomes, the n = 492 neutral updates (Δρ = 0) were ignored when fitting the model. Contrary to what was hypothesized, entropy did not predict the quality of updates to the list (b = 0.36, SE = 0.35, Z = 1.03, p = .303). When neutral updates were not ignored in the model and instead coded as improvements, entropy remained non-significant (b = 0.30, SE = 0.23, Z = 1.30, p = .193).

Although no predictions were made regarding the relationship between entropy and teams' likelihood of making changes to their rankings, this relationship was also tested. Each of the n = 22,629 sliding-window entropies calculated across the 32 time series were coded 1 if the team made an update to their list during that second, and 0 if they did not. The multilevel logistic model predicting update/no update with a fixed effect of entropy and a random effect of team revealed that the higher a team's entropy, the more likely they were to update the list (b = 0.40, SE = 0.14, Z = 2.79, p = .005).

Hypothesis 6 (the more that teams reorganised their communication dynamics, the longer they would take to complete the task but the better the quality of their final list) was tested with several regressions. More varied entropy scores suggests more varied communication dynamics, and hence more reorganisation (in absolute terms) in the dynamics across the task, controlling for task length. Hence, the average (with respects to time) magnitude of reorganisation was operationalised as the standard deviation (SD) of the sliding-window entropies for that task. These n = 32 SDs were right skewed (γ = 1.09) and platykurtic (κ = 4.23), had a mean of 0.167 (SD = 0.093), and ranged from 0.049 to 0.461. To test whether they predicted task completion time and group ρ, two regressions predicting each of time and ρ, respectively, were fit. Entropy SD was significantly, positively associated with completion time in the first regression (b = 1535.3, SE = 531.8, t(30) = 2.89, p = .007) but was non-significant in the second (b = -0.04, SE = 0.35, t(30) = -0.13, p = .899). This provides partial support for Hypothesis 6: The more varied a team's turn-taking dynamics, the longer they took to complete the task. However, this variance had no effect on the accuracy of the final rankings list.

The number of communication-dynamics reorganisations was defined as the number of reorganisations (i.e., critical instability to stable phase events, or just critical instabilities), and operationalised as the number of peaks in the entropy time series, given peaks in entropy were assumed to represent critical instabilities. To test whether the number of reorganisations predicted completion time and group ρ, two more regressions were fit. Again in partial support of Hypothesis 6, the number of peaks was significantly positively related to completion time (b = 68.44, SE = 5.61, t(30) = 12.19, p < .001) but not significantly related to group ρ (b = .01, SE = 0.1, t(30) = 0.83, p = .415). That is, the more times a team reorganised their dynamics, the longer they took to complete the task. However, the number of reorganisations did not affect the accuracy of the final rankings list.

Finally, for Hypothesis 7 (the faster that teams reorganised, the less time that they would take to complete the task), reorganisation speed was operationalised, following the same logic as above, as the number of peaks in the entropy time series, but this time divided by the task length (in minutes). Regressing task completion time on entropy peaks per minute revealed that peaks per minute was significantly, negatively associated with completion time (b = -952.44, SE = 367.38, t(30) = -2.59, p = .015), thereby supporting Hypothesis 7. The more frequent the reorganisations in turn-taking dynamics, the longer the task took.

Discussion

The purpose of this study was to validate assumptions regarding collaborative-problem-solving performance, to validate a novel measure for capturing the dynamics of team communication, and to see what effects those dynamics have on problem solving. In support of Hypothesis 1, groups performed better through collaboration compared to solely aggregating individuals' answers. This suggests that teaming (sharing information, challenging perspective, triggering novel ideas, etc.) is an important process for optimal problem solving. Additionally, in support of Hypothesis 2, the better team members were at the individual task, the better their teams performed in the group task. That is, team collaborative-problem-solving performance was still largely dependent on the task-relevant abilities and knowledges of team members.

However, there was mixed support for Hypothesis 3, and there were limits to the benefits of teaming. For example, spending more time on the task was not associated with better performance. It may be that it does not take effective teams much time to achieve all – or at least close to all – the available benefits of teaming. On the other hand, the number of updates to the rankings list did positively correlate with performance. This suggests that some of the dynamics of effective teaming may be indirectly captured in how rankings editors interacted with the list. For example, if the team were willing to challenge the list (potentially leading to new insights and collaborative learning), this may be seen in items already ranked being moved again (i.e., more changes to the list).

In accordance with Hypothesis 4, the proposed turn-taking entropy measure varied across teams and across communication, and there were clear peaks and troughs in the entropy time series (between which the network graphs were clearly visually different). This suggests that teams' turn-taking dynamics evolved over time, and that the entropy measure was able to, at least in part, capture that evolution. However, contrary to Hypothesis 5, higher entropy – that is, more symmetric and less restricted turn taking – was not associated with better group performance, nor was it associated with longer task duration. It may be that symmetric turn taking and equal sharing of speaking time (which would result in higher entropy scores) was not a good strategy for teams, especially when task-relevant knowledge was not shared equally between team members. In such cases, it may have been more optimal for one or two team members to dominate the conversation.

One interesting finding is that teams were more likely to make changes to their rankings during periods of higher entropy. This suggests that teams were less likely to make changes to the list during periods of stable communication. This is not wholly surprising. For example, a lack of consensus in the group regarding the utility, limitations, and correct ranking of an item would be likely to result in prolonged, low-entropy, stable communication dynamics characterised by one team member explaining their ideas (or the principal "owner" of an idea repeatedly trading utterances with the main detractor). Conversely, entropy scores would be higher in the subsequent situation when all team members would take turns to give verbal endorsement of an idea/decision just before it was implemented into the rankings list. This brief uptick in entropy could be conceived as a critical instability that one would expect to quickly settle on a more stable dynamic, presumably when the team had to discuss its next problem. It is not clear, however, whether these instabilities can be triggered by external interference, and whether naturalistic and forced instabilities resolve in different ways.

In partial support of Hypothesis 6, more peaks in the entropy time series and higher variance in the entropy values (indicating more critical instability-resolution events and greater variance in communication dynamics, respectively) were both associated with longer task-completion times. This suggests that the more a team's dynamics have to change (both with respects to the absolute difference in entropy peaks and troughs and with respects to absolute changes in entropy) to find the stable dynamics, task strategies, or insights required to get the rankings to a sufficient level of quality (for the team to be satisfied enough to submit the rankings and end the task), the longer the task will take.

Concordantly, in support of Hypothesis 7, the faster that teams found their next stable pattern of communication (i.e., the greater the peaks per minute), the faster they could find optimal dynamics, find optimal solutions, or resolve disagreements, etc., and the faster they could complete the task. However, contrary to what was predicted, having more frequent critical instabilities was not associated with better task performance, despite such instabilities leading to more updates to the rankings. Hence, having group dynamics become unstable and change may not be sufficient for finding optimal task solutions/strategies, particularly if the dynamics ultimately reached are suboptimal. Future research should investigate what constitutes optimal turn-taking dynamics and determine what factors make teams more likely to find those dynamics.

In conclusion, this study has reaffirmed the importance of collaboration in effective problem solving and demonstrated the utility of turn-taking entropy as a measure of team communication dynamics. Future research should validate this measure against other measures of group dynamics, particular with respects to the identification of critical instabilities, and further examine how critical instabilities can be triggered and manipulated to guide their resolution.

Acknowledgments

This work was supported by the Australian Department of Defence, Defence Science and Technology Group (partnership grant MyIP8655) and Human Performance Research Network (HPRNet, partnership grant ID9024). MJR was supported by the Australian Research Council Future Fellowship (FT180100447).

References

Baker, D. P., Day, R., & Salas, E. (2006). Teamwork as an essential component of high-reliability organizations. *Health services research*, *41*(4p2), 1576-1598. https://doi.org/10.1111/j.1475-6773.2006.00566.x

Barnosky, A. D., Hadly, E. A., Bascompte, J., Berlow, E. L., Brown, J. H., Fortelius, M., ... & Smith, A. B. (2012). Approaching a state shift in Earth's biosphere. *Nature*, *486*(7401), 52-58. https://doi.org/10.1038/nature11018

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The journal of the learning sciences*, *9*(4), 403-436. https://doi.org/10.1207/S15327809JLS0904_2

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510-516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chowdhury, T. M., & Murzi, H. (2020, June). The Evolution of Teamwork in the Engineering Workplace from the First Industrial Revolution to Industry 4.0: A Literature Review. In *2020 ASEE Virtual Annual Conference Content Access*. https://doi.org/10.18260/1-2--35318

Dakos, V., Van Nes, E. H., d'Odorico, P., & Scheffer, M. (2012). Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*, *93*(2), 264-271. https://doi.org/10.1890/11-0889.1

Ervin, J. N., Kahn, J. M., Cohen, T. R., & Weingart, L. R. (2018). Teamwork in the intensive care unit. *American Psychologist*, *73*(4), 468–477. https://doi.org/10.1037/amp0000247

Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., & Holling, C. S. (2004). Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review of Ecology, Evolution, and Systematics*, *35*, 557-581. https://doi.org/10.1146/annurev.ecolsys.35.021103.105711

Galbraith, J. R. (2009). Multidimensional, multinational organizations of the future. In F. Hesselbein & M. Goldsmith (Eds.), *The organization of the future 2*, (pp.174-187). Jossey-Bass.

Gatewood, R., Feild, H. S., & Barrick, M. (2015). *Human resource selection*. Cengage Learning.

Goodwin, G. F., Blacksmith, N., & Coats, M. R. (2018). The science of teams in the military: Contributions from over 60 years of research. *American Psychologist*, *73*(4), 322. https://doi.org/10.1037/amp0000259

Gorman, J. C., Dunbar, T. A., Grimm, D., & Gipson, C. L. (2017). Understanding and modelling teams as dynamical systems. *Frontiers in psychology*, *8*, 1053. https://doi.org/10.3389/fpsyg.2017.01053

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, *19*(2), 59-92. https://doi.org/10.1177/1529100618808244

Hagberg, A.A., Schult, D.A., & Swart, P.J. (2008). Exploring network structure, dynamics, and function using NetworkX.In G. Varoquaux, T. Vaught, & J Millman (Eds.), *Proceedings of the 7th Python in Science Conference (SciPy2008)* (pp. 11–15). Available online at: http://bit.ly/3nXMmx1

Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., & Fiore, S. M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, *73*(4), 532. https://doi.org/10.1037/amp0000319

Heerklotz, H., & Tsamaloukas, A. (2006). Gradual change or phase transition: characterizing fluid lipid-cholesterol membranes on the basis of thermal volume changes. *Biophysical journal*, *91*(2), 600-607. https://doi.org/10.1529/biophysj.106.082669

Hill, G. W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological bulletin*, *91*(3), 517. https://doi.org/10.1037/0033-2909.91.3.517

Hoffman, B., Shoss, M., & Wegman, L. (2020). The Changing Nature of Work and Workers: An Introduction. In B. Hoffman, M. Shoss, & L. Wegman (Eds.), *The Cambridge Handbook of the Changing Nature of Work* (pp. 3-19). Cambridge University Press. https://doi.org/10.1017/9781108278034.001

Hoogeboom, M. A., & Wilderom, C. P. (2020). A complex adaptive systems approach to real-life team interaction patterns, task context, information sharing, and effectiveness. *Group & Organization Management*, *45*(1), 3-42. https://doi.org/10.1177/1059601119854927

Hughes, T. P., Carpenter, S., Rockström, J., Scheffer, M., & Walker, B. (2013). Multiscale regime shifts and planetary boundaries. *Trends in ecology & evolution*, *28*(7), 389-395. https://doi.org/10.1016/j.tree.2013.05.019

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90-95. https://doi.org/10.5281/zenodo.592536

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher*, *38*(5), 365-379. https://doi.org/10.3102/0013189X09339057

Likens, A. D., Amazeen, P. G., Stevens, R., Galloway, T., & Gorman, J. C. (2014). Neural signatures of team coordination are revealed by multifractal analysis. *Social neuroscience*, *9*(3), 219-234. https://doi.org/10.1080/17470919.2014.882861

May, R. M., Levin, S. A., & Sugihara, G. (2008). Ecology for bankers. *Nature*, *451*(7181), 893-894. https://doi.org/10.1038/451893a

McKee, M., & Healy, J. (Eds.). (2002). *Hospitals in a changing Europe* (Vol. 3). Buckingham: Open University Press.

McSharry, P. E., Smith, L. A., & Tarassenko, L. (2003). Prediction of epileptic seizures: are nonlinear methods relevant? *Nature medicine*, *9*(3), 241-242. https://doi.org/10.1038/nm0303-241

Nemiroff, P.M. & Pasmore W. A. (1975). Lost at sea: A consensus-seeking task. In W. Pfeiffer & J. Jones (Eds.), *1975 Annual Handbook for Group Facilitators* (pp. 28–34). University Associates.

O'Donnell, A. M. (2006). The Role of Peers and Group Learning. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 781–802). Lawrence Erlbaum Associates Publishers.

Ramos-Villagrasa, P. J., Marques-Quinteiro, P., Navarro, J., & Rico, R. (2018). Teams as complex adaptive systems: Reviewing 17 years of research. *Small Group Research*, *49*(2), 135-176. https://doi.org/10.1177%2F1046496417713849

Reiter-Palmon, R., & Leone, S. (2019). Facilitating creativity in interdisciplinary design teams using cognitive processes: A review. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *233*(2), 385-394. https://doi.org/10.1177%2F0954406217753236

Saleh, M., Lazonder, A. W., & De Jong, T. (2005). Effects of within-class ability grouping on social interaction, achievement, and motivation. *Instructional Science*, *33*(2), 105-119. https://doi.org/10.1007/s11251-004-6405-z

Scheffer, M. (2020). *Critical transitions in nature and society*. Princeton University Press.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., ... & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53-59. https://doi.org/10.1038/nature08227

Schiepek, G., & Strunk, G. (2010). The identification of critical fluctuations and phase transitions in short term and coarse-grained time series—a method for the real-time monitoring of human change processes. *Biological cybernetics*, *102*(3), 197-207. https://doi.org/10.1007/s00422-009-0362-1

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shimizu, Y., Matsui, J., Unoura, K., & Nabika, H. (2017). Liesegang mechanism with a gradual phase transition. *The Journal of Physical Chemistry B*, *121*(11), 2495-2501. https://doi.org/10.1021/acs.jpcb.7b01275

Shuffler, M. L., Pavlas, D., & Salas, E. (2012). Teams in the military: A review and emerging challenges. In J. H. Laurence & M. D. Matthews (Eds.), *The Oxford Handbook of Military Psychology* (pp. 282-310). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399325.013.0106

Song, J. K., Fukuda, A., & Vij, J. K. (2008). Gradual phase transition between the smectic-C* and smectic-C A* phases and the thresholdless antiferroelectricity. *Physical Review E*, *78*(4), 041702. https://doi.org/10.1103/physreve.78.041702

Venegas, J. G., Winkler, T., Musch, G., Melo, M. F. V., Layfield, D., Tgavalekos, N., ... & Harris, R. S. (2005). Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature*, *434*(7034), 777-782. https://doi.org/10.1038/nature03490

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272. https://doi.org/10.1038/s41592-020-0772-5

Wang, W., Chen, Y., & Huang, J. (2009). Heterogeneous preferences, decision-making capacity, and phase transitions in a complex adaptive system. *Proceedings of the National Academy of Sciences*, *106*(21), 8423-8428. https://doi.org/10.1073/pnas.0811782106

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-solving phase transitions during team collaboration. *Cognitive science*, *42*(1), 129-167. https://doi.org/10.1111/cogs.12482

Wiltshire, T. J., Hudson, D., Lijdsman, P., Wever, S., & Atzmueller, M. (2021). Social analytics of team interaction using dynamic complexity heat maps and network visualizations. In *2021 IEEE 2nd International Conference on Human-Machine Systems* (pp. 1-6). IEEE. https://doi.org/10.1109/ICHMS53169.2021.9582454

World Economic Forum. (2020). *The Future of Jobs Report 2020*. World Economic Forum. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

Zhao, L., Yang, G., Wang, W., Chen, Y., Huang, J. P., Ohashi, H., & Stanley, H. E. (2011). Herd behavior in a complex adaptive system. *Proceedings of the National Academy of Sciences*, *108*(37), 15058-15063. https://doi.org/10.1073/pnas.1105239108