Tackling the Zombie Apocalypse: sensemaking in simulated disaster management

Tom Duffy  
University of Birmingham  
School of Engineering  
Birmingham  
UK  
tomduffy1984@gmail.com

Chris Baber  
University of Birmingham  
School of Engineering  
Birmingham  
UK  
c.baber@bham.ac.uk

In this paper, we report the design of a simulated disaster management exercise that is used to explore the manner in which teams of emergency responders make sense of unfamiliar, dynamic situations. The paper develops a notion of sensemaking that combines semantic (i.e., extracting meaning from cues in the environment) and pragmatic (i.e., recognising opportunities for courses of action). This notion of sensemaking is explored in a study in which teams of experienced emergency responders, organised into different command structures, deal with the spread of a threat on a University campus. The results indicate how the command structures affect the ability of the teams to employ semantic or pragmatic sensemaking. Implications for supporting emergency response are considered.


1. INTRODUCTION

Disasters and emergencies are separated by scale: a disaster is an event that causes large scale damage and / or loss of life; an emergency may involve damage and loss of life but on a manageable scale. Emergency responders tend to be highly trained individuals working in teams, the purpose of training being to maintain a community of practice (Wenger et al., 2002). In contrast, disaster management personnel are usually working in exceptional circumstances and teams of people are newly formed in response to the disaster. These ad hoc teams are often only in existence for the duration of that particular disaster. Consequently, these groups can work (during initial phase of the disaster) as networks of exploration (Baber et al, 2008) or adhocracies (Mendonca et al. 2007). These networks form around problems, have low formalisation of behaviours, few shared Standard operating procedures and tend to have high capacity for problem solving and creating innovations. Such networks can start as loosely defined groups with little consensus and differing ideas to what the problem domain is. Over time they find a common understanding of the problem, a shared language for talking about it and an agreed way of responding. As such, there is a need “to practice generic skills related to information sharing and collaborative sensemaking” (McMaster & Baber, 2011). The question, therefore, is what term ‘sensemaking’ could mean, beyond a literal definition, and how common understanding could be trained and supported. In particular, there is a need to support the ways in which emergency response teams can use sensemaking to develop Situation Awareness.

Computer Supported Cooperative Work has long recognised the importance of common understanding in shared activities (Monk, 2003; Olson and Olson, 2000), but it has paid less attention to Situation Awareness. Gergle et al. (2013) showed that displayed information supports the development of common ground and Situation Awareness in distinct but complimentary ways. The benefit of providing different forms of information, for grounding, depends on the nature of the activity and environment (Carroll et al., 2003). For example, when using shared video, commanders had a tendency to place inappropriately high trust in the video, at the expense of consulting other data sources (McGuirl et al., 2008). This suggests that some forms of information can interfere with the ability to make sense of a situation (by unduly focusing attention on some aspects rather than others).
De Jaegher and Di Paolo (2007) proposed that sensemaking is concerned with the coordination of action. From this perspective, it becomes important to know the response that could be made. De Jaegher and Di Paolo (2007) use an example of two people ballroom dancing to illustrate this point. The ‘sense’ that each person makes in the dance involves their ability to remember the steps, interpret the music and respond to their partner in order to make the correct move. The changing situation creates constraints on the possible actions that can be performed, and the correctness of the action depends on the ability to recognise which constraint is most important at that stage in the process. As Weick (1995) notes, in his analysis of the Mann-Gulch forest fire, “Social construction of reality is next to impossible amidst the chaos of a fire, unless social construction takes place inside one person's head, where the role system is reconstituted and run.” In his account, knowing which action to perform and whose lead to follow in performing actions, was paramount in this incident. What both examples imply is that making ‘sense’ involves knowing what to do in response to a given situation.

This notion of sensemaking can be contrasted with one that focuses on interpreting a situation in terms of the ‘meaning’ that can be extracted from it. In this approach, making ‘sense’ is a cognitive process of information collection and assimilation. For example, the data-frame approach (Klein et al., 2005a,b) is about linking and appreciating aspects of a situation in order to comprehend it. An example might be of an intelligence analyst who is receiving data from various sources and trying to link that data together in such a way that it forms a plausible hypothesis. The criterion for what data are allowed to contribute to this hypothesis can be considered a frame. Data that does not fit the frame is dismissed as noise or it might prompt the analyst to reconsider the criteria for what can be considered as evidential data. In this scenario, we would expect an experienced analyst to see patterns in the data that a more junior analyst might not.

We view sensemaking as operating on two planes. The first is a process that is concerned with defining the constraints under which action and communication takes place. This will be called pragmatic sensemaking. The second is a metacognitive process, concerned with managing the meaning of the situation. This will be called semantic sensemaking. Both forms of sensemaking are essential to emergency response.

2. SENSEMAKING AND COMMAND STRUCTURE

While one might assume that disaster management involves a central commander marshalling personnel, the reality is more likely to be a loose collection of agencies, each with its own command structure, sharing information on a piecemeal basis. The sharing of information could be through face-to-face meetings or through radio communications. While there is interest in the development of Common Operating Pictures on which a shared view of the situation and response could be presented, this is not common practice at present (particularly when multiple agencies are involved in the response). Thus, a pressing question for emergency response is how information can be shared. One approach to this question is to utilise an information-management structure, analogous to the command structure. For a centralised structure, all information would flow to and from a hub, from which the commander could issue orders to manage the response. However, this structure is not the only possibility. Traditionally, such structures follow the classification of Leavitt (1951) shown in figure 1.

![Figure 1: Types of Command Structure (after Leavitt, 1951)](image)

More recently, Alberts and Hayes (2003) contrasted a fully connected circle, in which all nodes connect to all other nodes (which they term ‘edge’) with a wheel (which they term ‘Command and Control’, C2). They suggest that the edge network ought to be more agile, flexible, adaptive and better able to cope with unexpected events. From this one might anticipate an edge network to be most appropriate for disaster management. In an experimental comparison of command structures, Stanton et al. (2015) show that the edge network actually performs less well; the Y network performs best, with the wheel next and the edge last. Perhaps this should not be too surprising as the Y could be considered a form of small-world (scale-free) network. In previous work (using large-scale models of communication networks for military command and control), we showed that the
command structures tend to have similar performance in small networks dealing with predictable events, but that as network size or event predictability changes, then the scale-free network shows best performance. Indeed, the network models would spontaneously split into set of Y networks at certain network sizes (Baber et al., 2012).

We propose that the wheel (C2) structure ought to be most appropriate for pragmatic sensemaking; each member of the structure need only know how to connect to the central node. Under such a structure, one would also expect each node to take on a specialised role, with demarcation between roles being coordinated by the central node. This means that semantic sensemaking is mainly a function of the central node. As the volume of messages and decisions increases, the central node could become overwhelmed and become a bottleneck for decision making. In contrast, the fully-connected circle (edge) structure would divide semantic sensemaking between nodes, which could reduce the bottleneck. However, this could lead to high overheads in pragmatic sensemaking (because it is not clear which node is responsible for which role, leaving scope for duplication or omission of roles). This means that having some means of indicating which node is fulfilling which role would be beneficial. To this end, we implement a communications broker which suggests not only which nodes might need to connect with each other (in order to share information) but also what role each node is performing. In this way, the edge network could have its pragmatic sensemaking enhanced.

From this discussion, we develop a set of working hypotheses. We would expect the C2 network to be affected by the quantity of information, such that high volume communications could lead to bottlenecks in this network. In contrast, volume of communications should not affect the edge network, because it would be able to adjust its pattern communications. We would expect the communications broker to support communication of relevant information between nodes in the edge network, and minimise risk of duplication of effort. Finally, in both network structures, we would expect participants to balance taskwork (in terms of addressing the problem) with teamwork (in terms of managing communication).

3. SIMULATING DISASTER MANAGEMENT

In Emergency response multiagency training, aims to encourage personnel to cooperate on large-scale situations that simulate, as realistically as possible, the threats that they might face in real-life. In such situations, Emergency Response training is primarily directed towards ensuring appropriate procedural compliance and communications between agencies. For Disaster Management, the aim might be to challenge the manner in which unusual, unfamiliar or unexpected events are dealt with. To this end, the use of a Zombie Apocalypse is common in disaster management training (e.g., the Zombie Preparedness Campaign1). This is not because anyone is anticipating such an event but rather that it provides a number of features which can be used in emergency response training. The threat, while respondents might have some familiarity from television and films, can be unpredictable. This means that the mechanics of the threat, in terms of its spread, effect, defence etc., can be managed within the simulation and presented in ways which challenge the trainees so that they do not fall back on familiar procedures or training but respond to the novel aspects of the situation. The motivation of the threat ‘actors’ can be presented without any underlying terror dimension; this means that the exercise can address a threat from the ‘other’ without needing to engage in debate about the political or religious motivation of the threat. Consequently, an important aspect of Disaster Management is the need to comprehend the situation in sufficient detail to decide on the appropriate response. The focus of this paper is on this initial situation assessment.

### 3.1 The Experimental Environment

The participant’s view of the experimental platform is shown in figure 1. The platform was hosted on a webserver which allowed participants to access it from any personal computing device with an internet connection and a web browser. An interactive map was developed using an application program interface (API) provided by Google (Google Maps API v3) and jQuery JavaScript was used to make customised draggable objects. The four objects that could be dropped on to the map were high, medium and low threat-level icons and an ‘unknown’ pin. This latter pin was used to mark an area of risk where the threat level had not been identified.

A chat system, was integrated into the HTML page using jQuery and PHP scripts. Messages sent between users were stored in a MYSQL database, which was used to inform a Communications Broker. The Communications Broker continuously monitored the chat, within a team, for key terms. The key terms were used as search terms to query a table of locations on the region of the map (which was based on a University campus).

---

The Communications Broker took locations retrieved from this query and added them as an indexed location to the appropriate record in the chat table in the database. On the participant view of the platform, the broker maintained a list of locations and the participants who were talking about them. This can be seen in figure 2 under the heading ‘talking about:’ to the right of the map. There are two reasons why locations are used as the topics in the Communications Broker. Firstly, it was imagined that participants would refer to locations in the scenario by using familiar or shortened versions of the building name. This helps the broker establish a common lexicon for referring to locations that is common to all participants. Secondly, it is envisaged that as participants are looking to resolve information on their maps, they can match the location of unknown threats on their map to locations and associated participants presented by the Communications Broker.

The chat system can be seen on the right of figure 2 under the title ‘Communications Links’. When a participant clicks a button it opens a chat window with the participant whose ID was on the button. The text chat system was restricted to 140 characters per message. This served to stop participants broadcasting large amounts of information to their group by copying and pasting large descriptions of current situational awareness. It also mimics a real response scenario where bandwidth can be restricted to basic text messaging. The social media site Twitter uses the same character limit in its posts.

4. THE STUDY

4.1 Participants and Procedure

The trial took place at Florida State University, Tallahassee, and involved 25 delegates from a ‘Disaster Systems’ module. This is a core module for the Emergency Management and Homeland Security certificate. Delegates on this program were experienced emergency response personnel. Consequently, it could be assumed that they would have good appreciation of risk management, situation awareness, communication protocols and the importance of map-based planning of response. As such, it is reasonable to assume that the
participants in the study would approach the task in a manner similar to their operational experience (even if the scenario is very unfamiliar to them, they could still be expected to approach the gathering and sharing of information in a way that corresponded to their usual practice). Having said this, discussion with participants prior to the experimental trials suggests that the command structures with which they had experience tended to be more similar (but not identical) to the centralised condition in the experiment (although several participants had, on occasion, used something closer to an edge organisation with no obvious central command).

Participants accessed the system by logging on with an email address (provided on registration to the study). Following log on, the administrator allocated participants, on presentation, to one of three groups. Each group operated under a different command structure:

- **Group 1: Centralised Command and Control Network (C2)** in which one participant acted as the Commander and managed the response to the other participants. All communication would be via the Commander.

- **Group 2: Edge Network (E)**, in which all participants were able to communicate directly with each other and there was no designated Commander.

- **Group 3: Edge Network with Communications Broker (CB)**, in which communications could be between all participants (as in Group 2) but a communications broker was used to suggest links.

Once the group had been assigned, the participants were presented with information concerning the nature of the threats that they were facing and were allowed to ask questions of clarification. Prior to the trial starting, threat information was shared between participants. That is, there were 13 Zombie threats in each scenario. The threats were dealt out to each member until there were no threats remaining. This means that a group of eight, five members could have two Zombie threats each and three members would have one threat each. Each participant was also given the location of all threats; those threats which had not been dealt to the specific participant would appear as ‘unknown’ on the map. In this way, no single group would have an undue advantage at the start of the experiment.

Each trial has three time periods; t1, t2, and t3. At the beginning of time period’s t2 and t3, some Zombies change location and / or threat level. This created a very challenging environment for the participants. During t1, the participants know the locations of all the threats but, with the exception of the threat(s) that were dealt to them, do not know the threat level. Therefore at t1, the task of the group is to resolve information about the unknown markers. The intention was to encourage semantic sensemaking amongst participants in the groups. As the experiment moves into t2 and t3, work became more challenging as any information about threats that was previously resolved may now have become obsolete. The intention was to encourage participants to not only consider the semantics of the situation (in terms of defining threat level) but also the pragmatics (in terms of how the actions of their fellow group members).

### 4.2 Measures of Performance

At the end of each time period, the experiment was paused and each participant was asked two questions:

- Which building (on the campus map) do you think is the MOST safe to be in currently?
- Which building do you think is the LEAST safe?

The participants selected their answer from a list of all buildings within the campus area. This selection was assumed to be based on the participants’ situation awareness of the scenario at that point in time. In performing the task, participants were trying to identify areas of risk and changes from the initial scenario. This information is distributed amongst a given group but may not be realised in individual situational awareness. The risk presented in the scenario can be thought of as a signal. Unresolved and old information in the participant’s situational awareness can be considered the noise that is diluting this signal. As such we can represent and assess these signals as a signal detection problem.

The receiver in this instance is the participant in a group working collectively to resolve the risk signal. The group as a collective is acting as a receiver responding to stimuli or targets in an environment. The group can be assessed on its sensitivity to detecting these stimuli. Signal detection classifies responses made by the participant into 4 categories; hit, miss, false alarm and correct rejection. A hit occurs where a stimulus is present and reported as such. Stimuli that are present but not reported are a miss. If a stimulus is not present but the participant reports detection, this is marked as a false alarm. A correct rejection occurs where a stimulus is not present but is reported as such. Sensitivity represents how well a group can detect that a target is present and is a function of hit rate and false alarm rate. The hit rate is the number of correctly detected stimuli divided by the total
number of responses. The total number of responses is equal to the number of people in each group.

The criteria for hits, misses, correct rejections and false alarms is summarised in table 1:
- Hit = correct identification of area of risk
- Miss = failure to identify area of risk
- Correct Rejection = correct identification of area of safety
- False Alarm = identifying area of safety as an area of risk.

<table>
<thead>
<tr>
<th>Risk Area</th>
<th>Safe Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skull (risk)</td>
<td>Hit (H)</td>
</tr>
<tr>
<td>Shield (safe)</td>
<td>Miss (M)</td>
</tr>
</tbody>
</table>

Table 1: Signal Detection Matrix

Sensitivity is a function of hit rate and false alarm rate and is represented by the symbol d’ (d-prime). A group that performs well in the task should have a high hit rate and low false alarm rate. d’ is the difference between the z-transform of the probability of a hit and the z-transform of the probability of a false alarm, and is given by equation 1:

\[ d' = Z(p(H)) - Z(p(FA)) \] eqn. 1

Another indicator of performance is the Total Proportion Correct (TPC). A correct response is a hit or a correction rejection. Combining these scores and dividing by the total number of responses gives us a TPC score for the group:

\[ TPC = \frac{H + CR}{No. responses} \] eqn. 2

The answers given by the participants were layered onto a map (figure 3). This map was not shown to participants but was used by the experimenters to analyse results. Participants only had access to the chat log and the map shown in figure 2. A skull and crossbones icon was used to represent a location that a participant had reported as high risk. A green shield was used to represent a location that a participant had reported as safe area. Locations with zombie threats were represented as a heat-map layer shown on the map below as red, yellow and green circles (figure 3). The level of zombie threat at a location dictates the size of the radius of the circle. Using this visual representation, results were taken of where participants correctly identified dangerous and safe locations.

4.3 Results

Results were analysed for periods t1 and t3 to see how the groups had performed after the initial period and then again after they had completed 3 rounds of the experiment. At the beginning of t1 every participant should have had a good awareness of where the threats in the scenario were located, as this information had been dealt to them at the beginning of the experiment. They might not be aware of the severity of this threat since this information was distributed amongst the groups. It was anticipated that t1 would also act as a period for the participants to become familiar with the experimental platform. Activity in this experiment was measured as the number of updates that participants made to their maps during the trial and the number of communications sent was measured to see to what extent participants were communicating with each other.

4.3.1 After t1

Table 2 shows the number of participants (N); the number of hits (H'), misses (M), false alarms (F') and correct rejections (C) for each group. d’ and TPC are calculated using the formulae presented above.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>H</th>
<th>M</th>
<th>F</th>
<th>C</th>
<th>d’</th>
<th>TPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>.62</td>
<td>.61</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>.00</td>
<td>.5</td>
</tr>
<tr>
<td>CB</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>.75</td>
<td>.64</td>
</tr>
</tbody>
</table>

Table 2: t1 performance results

Table 2 shows that the group using a communication broker (CB) has the highest sensitivity when detecting risk in the scenario. The group using a central commander (C2) has a slightly lower sensitivity but its TPC score is only 3% lower than that of CB. This is a result of strong score in correct rejections coinciding with a high miss-rate. The Edge Group (E) has neutral sensitivity in that its false alarm rate is equal to its hit rate and a TPC score of 0.5%.
4.3.2 After t3
Results after the third period (Table 4) show CB sensitivity had declined to be the worst of the three groups, which is a reflection of high miss rate. TPC had fallen also mostly because of a rise in false-alarm rate. E shows an improved sensitivity and TPC and is the best performing group.

<table>
<thead>
<tr>
<th>N</th>
<th>Comms</th>
<th>Activity</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>9</td>
<td>3.11</td>
<td>17.8</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>3.5</td>
<td>13.9</td>
</tr>
<tr>
<td>CB</td>
<td>7</td>
<td>7.29</td>
<td>10.6</td>
</tr>
</tbody>
</table>

Table 3: t1 Network Statistics
Average communications rates for CB are more than double that of the other groups (Table 3). Density for CB is the highest, showing it is utilizing 50% of the possible communication links in its network, but the average activity per participant in CB is the lowest of the groups. C2 is approaching its highest possible density score (of 0.222) as it.

4.3.3 Communication Activity
Table 6 shows the C2 group commander (12) dictating, what he or she considers the safest building in the scenario area currently.

<table>
<thead>
<tr>
<th>Time</th>
<th>From</th>
<th>To</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:53:13</td>
<td>12</td>
<td>15</td>
<td>tully good</td>
</tr>
<tr>
<td>12:53:20</td>
<td>12</td>
<td>29</td>
<td>tully good</td>
</tr>
<tr>
<td>12:53:56</td>
<td>12</td>
<td>35</td>
<td>tullys good</td>
</tr>
<tr>
<td>12:54:07</td>
<td>29</td>
<td>12</td>
<td>Tully's good.</td>
</tr>
<tr>
<td>12:55:46</td>
<td>12</td>
<td>46</td>
<td>go to tully</td>
</tr>
<tr>
<td>12:55:58</td>
<td>12</td>
<td>35</td>
<td>none are at tully</td>
</tr>
</tbody>
</table>

Table 4: t3 performance results
Table 5 shows the network statistics for each group following the third period. C2 has exhausted all the possible communication links available to it. E and CB have similar density scores showing they have utilised around ¾ of the available connections (77% and 71% of network capacity respectively). Average communications per participant in CB is more than double that of C2 and E. As with results after t1, CB has the highest communication rate but also has the lowest activity rate. In contrast, the group with the lowest communication rate (C2) has the highest activity rate.

<table>
<thead>
<tr>
<th>N</th>
<th>Comms</th>
<th>Activity</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>9</td>
<td>8.44</td>
<td>34.4</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>9.88</td>
<td>25.4</td>
</tr>
<tr>
<td>CB</td>
<td>7</td>
<td>24.86</td>
<td>19.4</td>
</tr>
</tbody>
</table>

Table 5: t3 Network Statistics

\[
\text{Density}_{\text{max}} = \frac{\text{Possible C2 links}}{N(N-1)} = \frac{8 \times 2}{9 \times 8} = 0.222
\]

\footnote{The highest possible density score is calculated by dividing the total number of possible communication links in the C2 network by the number of possible communications links the group would have as a fully connected directed network.}

Table 6: Sample Communications from Group 1 t3
Table 7 shows participants from CB broadcasting the same question to multiple group members. Participants in this group appear to be asking questions in an effort to collect information more proactively.

<table>
<thead>
<tr>
<th>Time</th>
<th>From</th>
<th>To</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:56:46</td>
<td>26</td>
<td>34</td>
<td>Who has a Zombie at Dittmer lab</td>
</tr>
<tr>
<td>12:56:47</td>
<td>26</td>
<td>6</td>
<td>Who has a Zombie at Dittmer lab</td>
</tr>
<tr>
<td>12:56:49</td>
<td>26</td>
<td>16</td>
<td>Who has a Zombie at Dittmer lab</td>
</tr>
<tr>
<td>12:56:51</td>
<td>26</td>
<td>17</td>
<td>Who has a Zombie at Dittmer lab</td>
</tr>
<tr>
<td>12:56:53</td>
<td>26</td>
<td>76</td>
<td>Who has a Zombie at Dittmer lab</td>
</tr>
<tr>
<td>12:56:56</td>
<td>26</td>
<td>45</td>
<td>Who has a Zombie at Dittmer lab</td>
</tr>
</tbody>
</table>

Table 7: Communications Sample CB at t3
Table 8 shows a sample of messages from E. The members, in t3, appear to be sending threat information arbitrarily to other members in the group, but show less of the mass broadcasting behaviour of CB.

<table>
<thead>
<tr>
<th>Time</th>
<th>From</th>
<th>To</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:46:15</td>
<td>11</td>
<td>9</td>
<td>Medium zombie threat mendenhall A</td>
</tr>
<tr>
<td>12:46:29</td>
<td>47</td>
<td>37</td>
<td>High level at leach</td>
</tr>
<tr>
<td>12:46:31</td>
<td>9</td>
<td>37</td>
<td>High alert at landis</td>
</tr>
<tr>
<td>12:46:36</td>
<td>11</td>
<td>31</td>
<td>Medium zombie threat mendenhall A</td>
</tr>
<tr>
<td>12:46:36</td>
<td>40</td>
<td>37</td>
<td>Medium zombie threat montgomery</td>
</tr>
<tr>
<td>12:46:44</td>
<td>40</td>
<td>47</td>
<td>Medium zombie threat montgomery</td>
</tr>
<tr>
<td>12:46:46</td>
<td>47</td>
<td>37</td>
<td>Lowlevel at Katherine Montgomery hall</td>
</tr>
<tr>
<td>12:46:46</td>
<td>40</td>
<td>11</td>
<td>Medium zombie threat montgomery</td>
</tr>
<tr>
<td>12:46:48</td>
<td>40</td>
<td>9</td>
<td>Medium zombie threat montgomery</td>
</tr>
</tbody>
</table>

Table 8: Communications Sample E at t3
4.4 Conclusions

Activity rate and communications rate are indicators of how the participants in the experiment are focusing their efforts. Activity rate can be considered a measure of task work, i.e., the amount of effort that a participant is expending on updating and resolving their map information. Communications rate, comparatively, is a reflection of teamwork. Due to its structure, participants in C2 can only communicate with the commander in the network. This inhibits communications and thus teamwork in the group. The central node in this network is responsible for passing on information critical for other participants’ situational awareness. Figure 4 shows the state of C2’s network following the third time period. From the inward and outward edges from User ID 12 (acting commander) it can be seen that this node has become overwhelmed by inward communications and comparatively there is little outward communication back to the other group members.

The bottle-necking of information is a probable cause for the poor performance of C2 by the end of time period t3. C2, however, maintains the highest activity level throughout the trial. It could be the case, that as communications were stagnating, group members had more time to explore and play with the map while they waited for information from the commanding node in the network. One participant attempted to overcome the restrictions of the organisational structure by taking a screenshot of the situation map, uploading it to an image hosting website and then sharing the link to this image with the group commander (Figure 5). The commander did not pass on this method of spreading situation awareness to the other participants and the rest of the network did not adopt the behaviour.

Towards the end of t3, the commander in group C2 adopted an approach by which he / she stopped passing on situation information to other participants and instead broadcast what was believed to be the area of lowest risk. This approach could explain why C2 obtained a strong performance score.

CB had the highest levels of communications of all groups. The communications broker seemed to encourage communication within the group. This coincided with the lowest activity rate. In this group, there seemed to be an emphasis on teamwork and information sharing at the cost of carrying out taskwork. The cost of this is that the group has a lower sensitivity to risk. The communications broker appears to have had a detrimental effect on performance by encouraging too many links between participants. In evaluating the communications broker, it can be seen how this detrimental effect may have occurred. The communications broker listed the names of locations and the participants who had talked about them. Where a participant is searching for information about a certain location and sees the list of participants talking about that location, it could be that the searcher makes multiple enquiries with the listed participants associated with the location. The multiple enquiry approach seen here might be more expected of E where no communications broker is available. However, E maintains a good balance of activity and communications, which results in a high sensitivity to the risk targets in the experiment. This is achieved by using a slightly different strategy. E participants, on the whole, send small snippets of factual information rather than ask questions about locations.

5 DISCUSSION

The study has explored how different team structures have an impact on pragmatic and semantic sensemaking. If the task was solely concerned with defining threat level and location, then performance ought to be measured in terms of semantic sensemaking. However, such a measure could fail to explain why there are differences between the groups and how these differences change over time.

We draw a distinction between sensemaking and Situation Awareness; assuming that sensemaking is a precursor to Situation Awareness. In other
words, before it is possible to develop and maintain awareness of a situation, it is necessary to define the situation in the first place. As one of the reviewers of this paper notes, this means that the study only focuses on the first level (perception) of a commonly cited model of Situation Awareness (Endsley, 1995). Our intention was to challenge teams of responders in terms of defining the threat in a dynamic environment. In a real emergency, the response phase would work from this initial perception (or sense) of the situation and seek to apply appropriate Standard Operating Procedures, e.g., in terms of managing evacuation of areas, providing resources to support activity, setting up casualty stations or exclusion zones etc. The scenario in this study did not explore any of these operational issues.

In broad terms, the three conditions in the experiment provoke three different behaviours and strategies from the corresponding groups taking part. C2’s structure restricts communications and isolates group members. The result is they attempt to find alternative ways to communicate or tinker with their maps. E operates by sending about the situation without a great deal of searching and question asking occurring. The communications broker in CB encourages participants to look for information rather than just share the information they have on their map.

One conclusion from this exploration is that the factors which influence pragmatic sensemaking in this study have the biggest impact on performance. The manner in which communication is managed either lead to participants filling time with unproductive task activity (C2) or to participants attending more to the communications than to the task (CB). This suggests an interaction between the manner in which a team is organised and the technology which is used to support this organisation. Further work could address the potential role of the communications broker on a C2 structure. In terms of semantic sensemaking, the dynamics of this task was sufficient to create confusion in the groups, leading to reduction in performance (in terms of d’ and TPC) from t1 to t3 for groups C2 and CB. However, group E showed a marked increase in both measures across t1 and t3. This suggests that, as the attention of groups C2 and CB shift from taskwork to teamwork (i.e., from dealing with information about the task to dealing with communications) so there is a related reduction in performance. Group E showed an improvement in performance and a more efficient approach to communication between the t1 and t3.

One explanation for the differences in that C2 reached saturation in its communication links and this had the effect of inducing a bottleneck in performance. This is one of the problems that was cited for the use of edge rather than C2 networks (Alberts and Hayes, 2003). However, the suggestion that the edge network produced superior performance contradicts the findings of Stanton et al. (2015). In addition to differences between the two studies in terms of the tasks that were performed, we also note that the Edge network sought to organise the way in which it communicated information. This reflects the expectation that an edge network could adapt to changing situation demands.

The communication broker condition resulted in the best performance at t1, with participants sharing information and developing a good understanding of the problem. However, this group produced the worst performance at t3. At both t1 and t3, the group using the communications broker had low activity scores and a highly connected network. This suggests that the provision of this support encouraged the participants to attend more to communicating than to acting of the task.

In terms of lessons learned for the human-computer interaction, the study has raised some questions concerning the ways in which teams of people share information to either reach consensus (common ground) on a problem or to define appropriate communication strategies. Presenting participants with a map with incomplete information was, we feel, a valid approach to an experimental study. In an operational environment, the challenge is to provide an overview of the situation that can show not only the geography (and spread) of the situation but also the intended actions of other agencies and the intended consequences of these actions. In other words, there is a need to not only show what is happening now, but also to indicate what actions are planned (and what these actions are intended to achieve). For the Zombie study, for example, we could have provided opportunity for selecting escape routes or for blocking the paths of Zombies. One would hope that the escape routes defined by one participant did not conflict with the blocked paths set up by another, but this would provide interesting scope to explore the challenges of deconfliction in emergency response. It also raises some interesting challenges as to how one presents the state of the world in the future (after actions have been performed) while also presenting the current state of the world, or how to present information which is uncertain or contested at the same time as information that is agreed.

The use of a communication broker was almost too successful in encouraging communication, in that it had the potential to distract participants from the risk evaluation task (encouraging them to spend more effort on communicating with each other). Clearly further work would be required to establish a means of balancing taskwork and teamwork in this scenario. However, the metrics used in this study could provide a useful starting point for
monitoring performance of teams in emergency response in order to support an adaptive communications broker / decision support system.

3. REFERENCES


Olson, G. M., and Olson, J. S., 2000, Distance matters, Human-Computer Interaction, 15, 139-178.

