Collaborative Sensemaking

By

Thomas Joseph Duffy

A thesis submitted to

The University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

The School of Electronic, Electrical and Systems Engineering.

The University of Birmingham

March 2015

Abstract

The research presented in this thesis seeks to separate the notion of collaborative sensemaking into two different modes; a semantic mode, which describes the cognitive and meta-cognitive processes of sensemaking and a pragmatic mode which recognises the constraints under which collective sensemaking takes place. Using quantitative data obtained from three novel experiments designed by the Author and one external study, the thesis seeks to find ways to measure collaborative sensemaking. Two organisational structures are compared and contrasted for abilities to support sensemaking processes and experimental results support previous research that decentralised edge networks perform better then hierarchical networks at sensemaking tasks. The concept of a Communications Broker is presented as a potential technology for aiding pragmatic collaborative sensemaking and two prototypes are built. Experimental data suggests that the Communications Broker does alter the behaviour of the participant networks performing sensemaking tasks and accordingly is proposed as an area of interest for future research.

For my parents;

Dr Thomas & Judith Duffy

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Professor Chris Baber from whom I have received continuous inspiration, encouragement and patience throughout my PhD journey. His knowledge and expertise have been invaluable and I am thankful I was able to work along side him for so many years.

I would also like to extend my appreciation to my colleagues Drs James Knight, Richard McMaster and Robert Houghton for their general mentoring and insights throughout my time at the University of Birmingham.

A special thanks also goes to Professor David Merrick and Dr Audrey Heffron-Casserleigh, and to the rest of the team at the Center for Disaster Risk Policy at Florida State University. The experience gained from my employment with them in emergency management proved invaluable to this thesis and helped to situate my academic work in real-world scenarios.

Finally, I would like to recognise my family for being an unshakeable source of support throughout this whole process.

Work from this thesis was published in the following papers:

Duffy, T., Baber, C., & Stanton, N. (2013) Measuring Collaborative Sensemaking. Proceedings of the 10th International ISCRAM Conference – Baden-Baden, Germany, May 2013. T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and L. Yang, eds

Table of Contents

Table of Figures

List of Tables

List of Equations

1 Introduction

This chapter introduces the approach to sensemaking that will be developed in this thesis, i.e., that it is often a collaborative process that operates in two different modes. The Mann Gulch disaster and 9/11 Word Trade Center attacks are used as examples from the field of disaster management to emphasise the nature of sensemaking in real-world scenarios and how factors such as environment, organisational structure and technology can alter the performance of sensemakers. These examples are used as inspiration for the research questions presented in this thesis.

1.1 Sensemaking

Sensemaking is the everyday process by which we give meaning to our environments. Sensemaking happens when people are confronted with unknowns, novel situations, confusing events or gaps in knowledge. One of the first firefighters to respond to the World Trade Center attacks on September 11th 2001, described the situation as “beyond our consciousness” (9/11 Commission Report, 2004). The firefighter’s words help illustrate the challenge of sensemaking; rather than merely making sense of a situation, sensemaking is enacted when we are forced to go beyond what we know. In other words, sensemaking involves a set of processes by which features of a situation need to be selected and combined into a meaning, or sense, which goes beyond the features themselves. This implies that sensemaking is partly a cognitive process (involving the selection and combination of features), partly a metacognitive process (involving the management of selecting, combining and interpreting features) and partly, as will be explored in this thesis, a collaborative process (involving people interacting with each other and with the technologies they use to manage the selection, combination and interpretation of features). Sensemaking is a worthy topic for study because it is challenging to observe and difficult to assess. This is partly because sensemaking, on one level, is an internal cognitive process that can only be indirectly measured by its output. It could be implied that if sensemaking is the name of the process, then ‘sense’ must be the output. But how is ‘sense’ represented or measured? Sense might be an agreement on what something might mean or an agreement on how to respond to an unfamiliar situation. This suggests that sensemaking is often a collaborative process since ‘agreement’ happens between two or more people. Note that agreement on a course of action or response is not necessarily predicated upon a shared understanding of what something might mean. As an example, imagine this scenario that is borrowed and adapted from Walker et al (2013) of drivers on a motorway (freeway). The drivers are unable to communicate outside the constraints of their individual system, i.e., the braking and turning indicators on the vehicle, yet they are still collaborating on the global goal of keeping the traffic moving. Each driver is engaged in the process of making his/her own sense of a dynamic situation yet is not able to share this interpretation with the other road users. However, the actions of all drivers imply that they are working towards an 'agreed' set of behaviours in their management of their activity on the road. In essence they are coordinating pragmatically towards short-term goals (i.e., avoiding collisions) and long-term goals (i.e., reaching their destination in a timely manner) without reaching agreement on their current understanding of the road conditions.

In this thesis, it is suggested that sensemaking can be contrasted between two modes; a *semantic* mode, drawing on the cognitive and metacognitive processes mentioned previously, where meaning is given to stimuli in an environment, and a *pragmatic* mode where an agreed meaning might not have been established but a common course of action is being enacted. We can imagine this pragmatic mode of sensemaking working in scenarios such as disaster management where high tempo events dictate decision making in scenarios where there is not time to share all information or to reach an agreement on meaning, but where there might be sufficient *common ground* (Clark & Brennan, 1991) to guide a response. A comprehensive review of sensemaking is given in chapter 2. The following two sections present the challenging nature of collaborative sensemaking in the realm of disaster management.

1.2 Mann Gulch Disaster

On August 5th 1949, 15 airborne firefighters, known as 'smokejumpers', parachuted into the steep Mann Gulch valley near the Missouri River in the Helena National Forest, Montana. Two hours after being dropped, 10 of the smokejumpers were dead and a further 2 fatally injured after changing wind conditions caused a 'blow up', which saw the fire cover approximately 3000 acres in 10 minutes and engulf the fleeing firefighters. Including a local fire marshal, James Harrison, who was already on the scene, 13 men died that day. Of the three smokejumpers that survived, two survived by escaping through a crevice in the ridge of the valley. More controversially, the foreman and oldest of the smokejumpers Wagman Dodge, avoided death by laying down in the embers of an escape fire that he had created but, in which, he had failed to convince the other members of his crew to join him. In the confusion of the unstable environment created by a rapidly moving fire, Dodge's young team of smokejumpers (aged between 17-28 years) had failed to understand their foreman's intentions and continued to flee the fire but were rapidly overrun by it.

This tragedy of the Mann Gulch was comprehensively researched and reported in the book '*Young Men and Fire*' by Norman Maclean(1992). Karl Weick(1993) revisits the sequence of events that Maclean(1992) recounts and addresses questions about why organisations unravel and how they might be made more resilient. For Weick(1993) a key issue was the loss of meaning amongst the firefighters in not recognising the escape fire as a possible salvation. Weick states "the basic idea of sensemaking is that reality is an ongoing accomplishment that emerges from efforts to create order and make retrospective sense of what occurs". What is particularly relevant to the discussion here is that the semantic sensemaking led Dodge himself to rapidly assess the situation and define an unusual course of action, but the pragmatic sensemaking of his colleagues as they watched him set a fire and then lie in the embers could not interpret his behaviour.

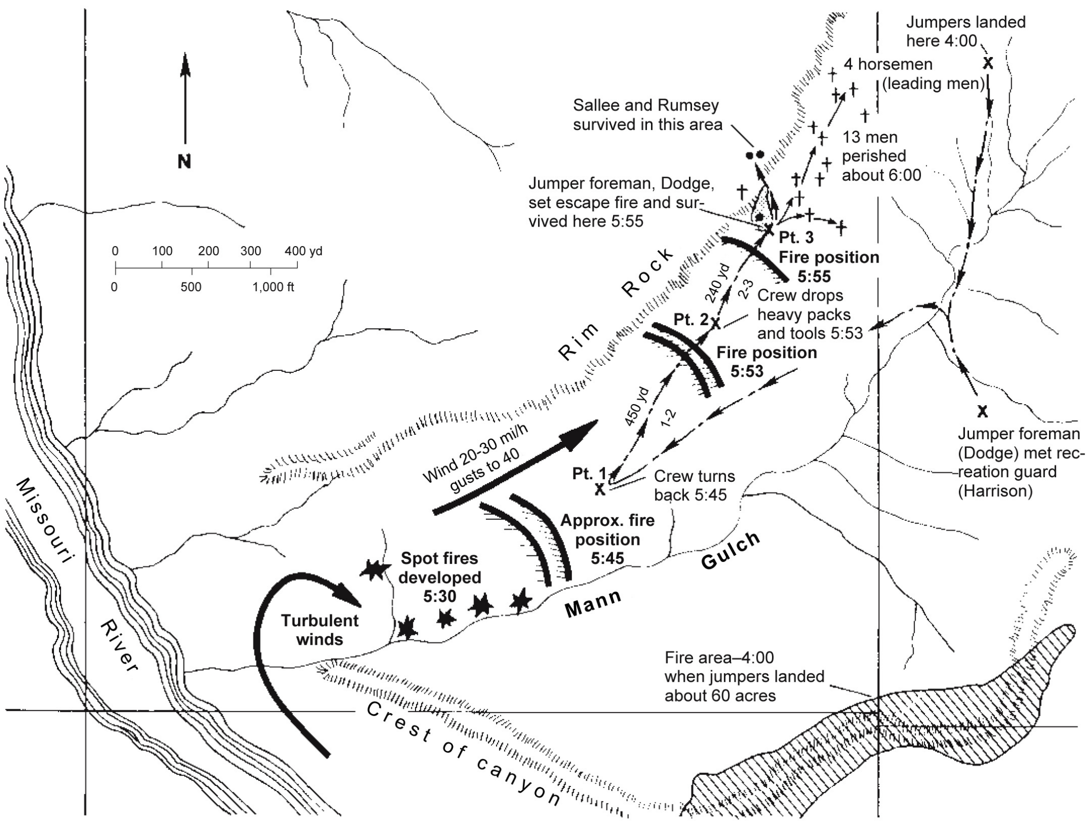

FIGURE 1 - MAP OF MANN GULCH ILLUSTRATING THE MOVEMENT OF CREW (ROTHERMEL, 1993)

At Mann Gulch the smokejumpers arrived expecting to face what was known as a 10am fire i.e. a fire they would have under control by 10am the following day. The context under which they went about their work was reinforced by factors such as Dodge (the foreman) stopping to eat supper with Harrison while hiking West towards the Missouri river on the North side of the Mann Gulch valley. In addition to this, while en route, one of their colleagues, David Navon, stopped to photograph the intense flames present on the South side of the canyon (Rothermel 1993). These are not the actions of people who feel they are in imminent danger. Thus, the context that the smokejumpers were facing anything other than 10am fire was not challenged until it was too late. Dodge and Harrison overtook the rest of the smokejumpers on the hike towards the Missouri river where the crew believed that they could fight the fire from the relative safety of being upwind and having the river on one side of them. Dodge and Harrison were the first to see that the fire had managed to cross the Mann Gulch valley and had blocked their route to the Missouri river. Changing wind conditions meant that the fire was being pushed up the North side of the valley towards the firefighters. At this point the two men turned and start retreating from the fire ordering the rest of the smokejumpers to do likewise. At this moment the remainder of the smokejumper crew are plunged into confusion. The context under which they were working has been mystified and they are asked to follow a retreat order from the foreman that they are to blindly follow. The label of the '10am fire' no longer applied and they had lost sense of their situation and their ability to make decisions based on an environment, they thought they understood, disappeared. The fire quickly headed towards the firefighters and created a situation in which it was impossible to share any individual sense of the events that were unfolding. Such was the fear of the other crew members, when Dodge lit an escape fire and ordered everyone to lie down in its embers the crew could not make sense of what he was trying to do and carried on fleeing.

In Weick's (1993) analysis of the events at Mann Gulch he poses the idea that the disaster was caused by an "interrelated collapse of sensemaking and structure". This raises the question; what is the relationship between organisational structure and the ability of that organisation to do sensemaking? Maclean (1992) asks a similar question early on in his book; "what the structure of a small outfit should be when its business is to meet sudden danger and prevent disaster?" (cited in Weick 1993). Looking more closely at these questions, there is a distinction to be made. It is possible for an organisation to face danger without having to do a great deal of sensemaking. It could be that a well-drilled army company frequently faces danger but has experience and roles defined to such an extent that its response becomes routine. The context they are working in, although dangerous, still makes sense to the organisation. The smoke jumpers were organised in a hierarchical structure. Dodge was the foreman and there was a second-in-command called William Hellman. Dodge was the oldest and most experienced of the crew and had the responsibility of establishing the strategy by which the fire would be fought. Hellman was closer in age to the other crew members than Dodge and was better at implementing orders rather than constructing them (Weick 1993). Hellman was good at coordinating the activity of the organisation rather than taking cues from the environment and creating new operational strategy. This means that the collective sense that the group shared was not arrived upon through collaborative effort. Dodge was using his own sense of the situation and passing down commands to the rest of the crew. This could have meant that the rest of the crew, reliant on Dodge's experience, did not engage in sensemaking at an individual level but waited for the 'sense' of the situation to be presented to them by their foreman. When the environment inhibits the communication between the firefighters, the organisation is not resilient to this change due to its reliance on the hierarchical command structure. This isolates the crewmembers, causes panic and they cease to operate as an organisation. This breakdown in the organisation causes them to lose "access to the novel ideas of other people" in the group and is, perhaps, why they did not acknowledge the escape fire as a credible way out of the danger (Weick 1993). Ironically, their experience that 'sense' would be presented by their foreman broke down when his actions stopped making sense to them. Why did the other firefighters not see the innovation of the escape fire as a solution? Weick (1993) asks the question 'how could more people either see this escape fire as a solution or develop their own solution?' One conjecture presented in this thesis is that the structure of the smokejumpers was not suited to the sharing of innovative ideas. The above example demonstrates that there is a relationship between collaborative sensemaking and organisational structure. From the viewpoint of this thesis, it is suggested that Dodge was performing his own semantic sensemaking and was using his experience and knowledge to create a solution. It was the group's failure in pragmatic sensemaking which produced a break-down in collaboration to the extent that the activity of the foreman made no sense to the others. This idea is explored in this thesis and forms part of Research Question 1 (section 1.4.1).

1.3 World Trade Center Emergency Operation Center

On the morning of September 11th 2001 members of the al-Qaeda terrorist group hijacked four commercial passenger planes with the intent of using them to carry out suicide attacks in the US. Two planes were flown into the North and South towers of the World Trade Center (WTC), one plane was crashed into The Pentagon and the fourth plane came down near Shanksville, Pennsylvania after passengers attempted to subdue the hijackers. In total there were 2,977 victims of the attacks.

At the time of the attacks, the New York City Office of Emergency Management (NYOEC) had its headquarters and Emergency Operations Center (EOC) located on the 27th floor of 7 WTC (Simon & Teperman, 2001). The role of an EOC is to coordinate the strategic response to large-scale emergencies and disasters. In the United States, its exact composition can vary from location to location but commonly, when activated, an EOC acts as a place where representatives of public (city, state and federal), private, non-profit and volunteer organisations can be located in an effort to monitor and coordinate all aspects of the response. Members of the EOC do not usually have direct control over individual response assets but instead are responsible for collecting, analysing and sharing data to inform strategic decisions that govern an inter-agency response. This means that the EOC is performing semantic sensemaking in that it is triaging information to form an interpretation of the situation. Additionally, it is assessing where this sense of the situation needs to be sent and what courses of action should follow, which represents a pragmatic sensemaking effort. Coordinating the communications between various agencies is one of the primary roles in the EOC and as such EOCs tend to be awash with various communications capabilities. The EOC at the NYOEC was one of the most sophisticated centres of its type in the world (Kendra and Wachtendorf, 2003):

> *"The site was equipped with computer messaging systems for communication among staff, a phone system with provision for microwave back-up, separate systems for fire department, police department and EMS communications, coastguard-operated video monitoring of New York's waterways and traffic monitoring of the city's streets... [Also] video conferencing and ARCVIEW and MAPINFO geographic information system (GIS) packages"*
>
> *- (Kendra & Wachtendorf 2003)*

The first plane hit The North Tower (1 WTC) at 8.46am and the NYOEC started directing resources to the affected area. However, when the North Tower collapsed at 10:29am the building containing the EOC sustained so much damage that it had to be evacuated and the building collapsed later that day. The evacuation of the EOC was so sudden that almost none of the supporting technology or equipment was taken. In the initial period following the evacuation, the Mayor, emergency managers and some agency representatives had to move between make-shift command posts as the scale of the disaster forced them back from WTC. Two parallel attempts were made to create temporary EOCs in a school and a library with limited success and it was not until approximately 60 hours after the initial attack that a semi-permanent EOC was established at a location on Pier 92 of the Hudson River. In addition to these problems, the radios used by the three emergency services – New York Police Department (NYPD), New York City Fire Department (FDNY) and Emergency Medical Services (EMS) - were incompatible. There was also a long-standing rivalry between NYPD and FDNY which contributed to them operating in a fairly autonomous mode. Other communication problems with radio repeaters and the inability of the emergency responders' radios to broadcast through buildings were also reported. The oral histories of the first responders on 9/11 provide insights into the nature of information that day:

1. Coordination and communication were serious problems;
2. Securing information early in the emergency was difficult;
3. Commanders lacked solid information to direct efforts;
4. Information was contradictory and difficult to interpret;
5. The collapse of the buildings was difficult to conceptualize;
6. Improvisation was common;
7. False information compounded confusion;

- (Dearstyne, 2007)

The points listed above raise questions about the nature of collaborative sensemaking. Firstly, were these problems caused or compounded by the evacuation of the EOC? Secondly, if the technology in the EOC had been available, would the response have been better enacted?

The complexities and challenges of inter-agency communication during disaster management are well documented (Bharosa, Lee & Janssen, 2010; Wolbers & Boersma 2013; Dearstyne, 2007). In essence, during a disaster the EOC is a socio-technical organisation engaged in sensemaking; personnel are collecting information from field units, the public and other sources; attributing meaning to events and sharing what they interpret with other agency representatives. The sense of the situation is then fed back to field commanders who can delegate and coordinate activity on the ground. It can be seen that there is a reliance on the hierarchical command structure from EOC to field commander for the sharing of situational awareness. In large events the sensemaking capabilities of the EOC can be hindered by the sheer volume of incoming information, which is referenced as a leverage point by Pirolli & Card (2005) in their model of sensemaking (section 2.2). The EOC at 7WTC had geospatial technology, which would have helped visualise and create a common awareness of the event but would not have aided with problems of information overload. It can also be posited that given the lack of inter-agency communications at ground level and the short-comings of the radio technologies possessed by the first responders, that the temporary loss of the EOC might not have had the impact on the immediate response that was suggested. Some reports imply that actually the resilience shown by the EOC as an organisation, to continue functioning in the interim between the evacuation of 7WTC to its re-establishment on pier 92, was impressive (Kendra & Wachtendorf 2003). In many ways the lack of inter-agency communications can be seen as a more troubling issue. One example relates to the inability of the FDNY to get information (visually or verbally) from NYPD helicopters on the extent of the fires in the towers above them (Dearstyne, 2007). Access to these reports may have allowed the FDNY ground command to make assessments on the safety of the towers before they collapsed killing hundreds of firefighters and other first responders.

Given the benefit of years of hindsight, the Author would like to stress that by no means are these observations meant to attribute blame or undermine the acts of heroism that occurred during the response to the 9/11 attacks. These examples are used to highlight the nature of collaborative sensemaking in high tempo events and that traditional models of command and control might not be suitable in such scenarios. In order to address this problem, it is necessary to have a means of measuring how sensemaking is performed and how it varies in its effectiveness. This inspires the first research question in this thesis (section 1.4.1). Also there is a question about the role of technology in supporting collaborative sensemaking. While Geographic Information Systems might be useful in creating a shareable representation of an incident, it does not necessarily help deal with the constraints under which sensemaking takes place, i.e., how the organisation might arrange itself to best deal with the situation. Looking at alternative methods of how technology might aid collaborative sensemaking is the purpose of second research question in this thesis (section 1.4.2).

1.4 Research Questions

The research questions that are to be investigated in this thesis are:

1.4.1 Research Question 1:

Q1. Can collaborative sensemaking be measured in the behaviour of social networks?

Q1.1. Which social network organisations favour collaborative sensemaking?

Q1.2. Are there social network analysis (SNA) metrics that describe this organisation?

Q1.1. Is it possible to develop a metric for collaborative sensemaking?

1.4.2 Research Question 2:

Q2. In what ways can technology aid in the process of collaborative sensemaking?

Q2.1. How might the behaviour of a social system be modified with technology?

Q2.2. How might we indicate the sensemaking activity of the network to members of that network?

1.5 Approach

The work on collaborative sensemaking in this thesis is situated in the field of *cognitive engineering* (sometimes called cognitive systems engineering). Cognitive engineering emerged in the in the 1980s and is a response to the need to understand the intersection of humans, technology and work (Norman, 1986). It is useful to practitioners in a wide range of domains as it combines cognitive science, ergonomics, human-computer interaction and systems engineering (Endsley et al., 2007; Gersh et al, 2005). In this thesis the approach taken to the study of collaborative sensemaking differs from previous research, which has usually fallen into the field of sociology or psychology. However, this author approaches the research topic from an engineering background, which adds in part to the novelty of this thesis. Studies in sensemaking have previously defined models of cognitive processes by recording the observable actions of people and the interactions of social systems. Approaches such as ethnographically informed data collection involve observations of groups, interviews with participants and observations of participants. The research in this thesis however, uses very little direct human observation and instead designs and builds systems to test the sensemaking capabilities of groups under varying conditions. As such, there is emphasis on the technical development of the systems and platforms used to collect data on groups performing sensemaking tasks. Evaluating cognitive processes, such as sensemaking, uses a kind of *black box* approach. This is an engineering term relating to a situation where the inner workings of a closed box can be guessed by assessing the relationship between the box's measureable inputs and outputs. In this case, the box is the person, or group of people, doing a task that requires sensemaking. It is only by measuring the inputs and outputs of the box that theories and models of its inner workings can be developed. As such, this thesis is concerned with developing metrics to measure the outputs of groups doing sensemaking in an attempt to quantify differences observed between them as the inputs are modified. In developing technologies to support sensemaking, these metrics will also be useful in determining the impact made by that technology on the behaviour of the group.

2 Literature Review

In this chapter, sensemaking and related literatures are reviewed and theories presented in this thesis are positioned. Working definitions of collaborative sensemaking are developed for use within the thesis and metrics that are used in later chapters are also explained.

2.1 A History of Sensemaking

Sensemaking involves the activity of imposing meaning (or sense) on to disparate information. Cognitive psychology has long sought to determine the manner in which such meaning is structured and a popular approach uses the concept of *schema*. The concept of schema has its advent in cognitive psychology in the work of Bartlett (1933), although it can be traced to the works of Plato and Ancient Greek Philosophers. Bartlett's use of the term schema was borrowed and refined from Henry Head in the former's attempts to rationalise the "persistent effects of past reactions" Bartlett (1933).

"Such Schemata modify the impressions produced by incoming sensory impulses in such a way that the final sensation of position or of locality rises into the consciousness charged with a relation to something that has gone before"

- (Head et al 1920).

Head's work was primarily concerned with postural change and movement and this enactive approach relates to De Jaegher and Di Paolo's (2007) work on participatory sensemaking that is reviewed later in this chapter (section 2.3). Bartlett extends the schema principle to incorporate, not just movements, but experiences and how they are continually organised to inform response to stimuli in an environment. Bartlett surpassed the notion that schemata are stored chronologically; they are actually interrelated cognitive constructs that are constantly assessed and maintained by an organism's reaction to incoming stimuli. For Bartlett (1933), schemata are a means of organising new information in line with prior knowledge and experience. This juxtaposes Head's idea that schemata are part of a passive retrieval process of fixed memories. Although not as succinctly expressed, it can be seen that Bartlett's work is a predecessor to Klein's Recognition-Primed Decision-Making and also the later work on the Data/Frame model of sensemaking (Figure 3). Anderson & Pearson (1984) suggest that the underlying principles of schema theory can be related to the Gestalt psychologists (Köhler, 1929; Koffka, 1935) who also emphasise the dynamic nature of mental organisation. Interest in this area was re-ignited in the late 1970s with the emergence of Artificial Intelligence, Expert Systems and the ambition of creating computers that could think like humans. During this renaissance of schema theory more emphasis was placed on developing working-models of schemata and how new information can inform change in our existing mental-models or cognitive-constructs (Rumelhart 1980). Minksy (1975) refers to these adaptable memory structures as frames and thus calls it Frame Theory.

The modern use of schema is as an abstract knowledge structure. These structures do not explicitly detail every experience but summarise them into frameworks that generalise or simplify our view of reality. The schemata represent an interrelated network of objects and events, actions and other objects (Rumelhart 1980). Schemata also influence how new information is categorised for later recall. New knowledge can be coded into existing schema or totally novel experiences can generate new schema. The models can be more or less abstract depending on what is being represented and the elements that make up the schema are known as "nodes", "variables" or "slots". Anderson & Pearson (1984) present an example schema (Figure 2), which demonstrates the envisaged interrelated nodes of a typical person's knowledge of a 'ship christening'. In this less abstract example, perhaps only a few of the nodes would be interchangeable i.e. 'done by celebrity' might be replaced by 'done by statesmen'.

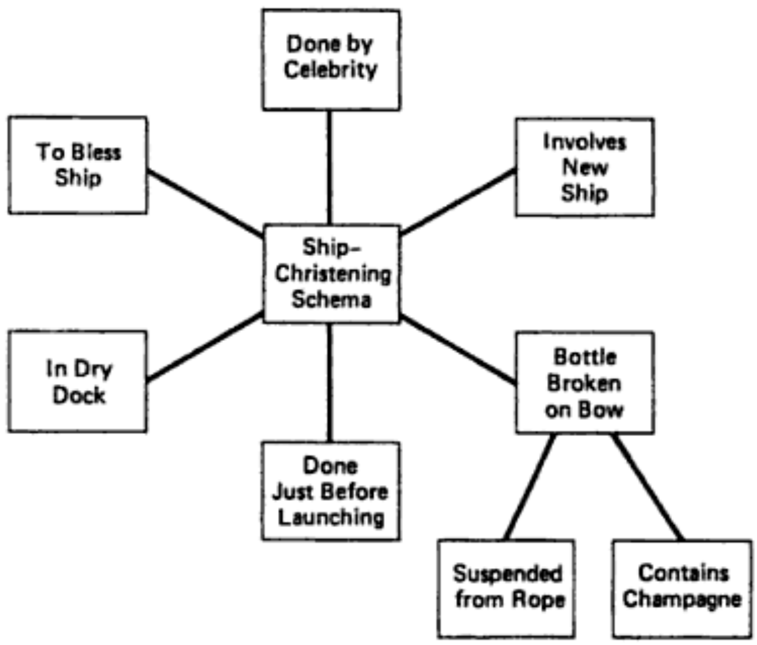

Figure 2 - Ship Christening Schema (Anderson & Pearson 1984)

One criticism of schema theory might be that it fails to emphasise how schema might be shared. In the example above, is it assumed that everyone has the same experience of ship christenings?

Related to Schema Theory is Script Theory. Schank & Abelson (1975) criticised Schema/Frame Theory as being too general so that it required specialisation to be useful in various applications. Scripts are mental constructs akin to schemata but they consist of a sequence of actions that are necessary for a goal to be achieved in a well-known situation.

> *"A script, as we use it, is a structure that describes an appropriate sequence of events in a particular context. A script is made up of slots and requirements about what can fill those slots. The structure is an interconnected whole, and what is in one slot affects what can be in another. Scripts handle stylized everyday situations. They are not subject to much change, nor do they provide the apparatus for handling novel situations, as plans do."*
>
> *– Schank & Abelson (1975)*

The fundamental unit of analysis in script theory is called a 'scene'; a situation with a defined beginning and end. As an example, Schank & Abelson(1975) present a 'restaurant script' which contains the scenes 'entering', 'ordering', 'eating' & 'exiting' and explain that there are certain expectations about going to a restaurant and the sequence of events that normally occur there. Scripts can be considered as specialised schemata called event schemata, which are described temporally or sequentially (Erasmus et al 2002). Scripts are triggered by certain situations where they are used to guide behaviours in accordance with certain expectations in the given scenario.

A criticism of script theory is that it does not elaborate on how new scripts are learnt or acquired. In schema theory new information, events and situations are organised in relation to existing schema, which equates to some sort of understanding. Although it is not suggested that scripts are completely without mindful thought or decision making, there is little elaboration in the theory as to what happens when an agent is forced to go 'off script'. As such, both theories struggle to elaborate on how an agent actively searches for data in ambiguous situations. Klein et al (2007) further suggested that the similarity been schema, scripts and frames is great and the differences too subtle to warrant a distinction; they are all similar attempts to model comprehension. These separate theories are consolidated into 'frames' which constitute part of the data/frame model of sensemaking (Figure 3).

Sensemaking encapsulates the above theories and puts them into the larger, macrocognitive context of what happens when an agent, or organisation, realises the inadequacy of its current understanding of events. As such, sensemaking does not redefine schema or script theory but uses them as building blocks in more holistic process for comprehending situations. Consequently, the cognitive processes involved in semantic sensemaking involve the application of frames (schemata or scripts) to a situation in order to define a structure for the situation. The semantic mode is the more classical form of sensemaking. It is about linking and appreciating aspects of an unknown 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 story or hypothesis. The criterion for what data are allowed to contribute to this hypothesis can be considered a framework or schema. 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. *Experience* becomes important in both modes of sensemaking. In semantic sensemaking, it allows rapid categorisation of the scenario and the criteria for which data should contribute. In pragmatic sensemaking, the feedback from previous action taken informs future action.

An information-science approach to sensemaking was presented by Brenda Dervin(1983) in which she sees *sense-making* as a fundamental necessity for humans who live in a reality full of knowledge-gaps; "Knowledge is the sense made at a particular point in time-space by someone" (Dervin 1998). Dervin attributes Piaget(1926) - whom alongside Bartlett(1933) is credited as a founder of schema theory - as a source of inspiration in her work and appreciates sense-making as an internal cognitive function:

> *"In the most general sense, sense-making (that which is the focus of study in the Sense-Making approach) is defined as behavior, both internal (i.e. cognitive) and external (i.e. procedural) which allows the individual to construct and design his/her movement through time-space. Sense-making behavior, thus, is communicating behavior. Information seeking and use is central to sense-making (as it similarly is seen as central to all*

communicating) but what is meant by these terms is radically different than what is typically meant in the positivistic tradition." – (Dervin 1983)

While this quotation emphasises semantic sensemaking, the focus on 'external' activity calls to mind the notion of pragmatic sensemaking that was raised in the first chapter and also the need for communication. The words 'information' and 'knowledge' are used interchangeably in the sense-making literature under the assumption that all information is dependent on subjectivity and as such represents a person's 'sense' of that information/knowledge. That humans are tainted by previous experience when attempting to bridge new gaps in their reality, is acknowledged and when communicating, information/knowledge cannot be passed on as an unaffected unit into another person without prejudices being applied. The 'sense-making' and 'sensemaking' literatures seem determined not to mention one another yet their similarities are obvious; both acknowledge the importance of experience and how it governs the constraints within which information search will be carried out. Hailing from communications theory Dervin, like Weick(1995), also appreciates the importance of constructing socially-plausible sense. Where agreed upon, this sense is called 'fact' (Dervin 1993). In addition, the idea of external sense-making behaviour, mentioned by Dervin, has links back to script theory and is similar to the idea in sensemaking of the interplay between action and interpretation (Weick 1995). The most glaring difference between 'sense-making' and 'sensemaking' is that the former treats sense as a product of the sense-making activity; suggesting that sense bridges the knowledge gap in a human's reality en route to their achieving some understanding. Alternatively, sensemaking is concerned with generating an understanding that is dynamic and "not a stored, frozen, meaning" (Klein et al 2007). The terminology used in this thesis is 'sensemaking' as opposed to 'sense-making' because it relates to a body of research that is used more widely in emergency management research and encompasses more formal models (section 2.2). Later in this chapter a working definition of sensemaking will be posited (section 2.5).

2.2 Models of Sensemaking

The traditional view of sensemaking as advocated and developed by Weick (1995) is of an on-going process through which meaning is constructed and explanations developed in an effort to establish what is going on. In this version of sensemaking the emphasis is on the approaches by which agents make sense of the unknown, the reasons for doing this, and the properties/attributes of the output of the process. Weick (1995) goes on to suggest that sensemaking is about the organisational consensus of plausible meanings. Approaching sensemaking from the viewpoint of organisational dynamics, he argues that sensemaking is an inherently social and collaborative process since all attributed meanings must be socially plausible and that consensus is key. This idea of 'plausibility' in sensemaking works on an individual and collaborative level. Individually, any understanding or meaning of a novel situation would relate to an existing framework held by the *sensemaker*. The sensemaker might also test this understanding internally with the question; would this be acceptable to others? The sensemaker could then express his/her interpretation of events with the organisation and see if it is accepted. The organisation may contain individuals with varying experience, expertise, beliefs and cultural backgrounds, and thus plausibility would be assessed socially against many frameworks.

This notion raises three important points for the research in this thesis. First, the proposal that sensemaking is inherently social raises the question of how it might be possible to engineer social networks that lead to superior performance on sensemaking tasks. This relates to Q2 (section 1.4.1). Second, the proposal that sensemaking is concerned with creating meaning raises the question of how such meaning could be evaluated (Q1). Weick's suggestion is that evaluation would be through consensus, which implies that a meaning is acceptable if enough people in the network agree with it. While this is attractive as a way of indicating the social aspect of sensemaking, it runs into problems when one considers the ways in which consensus could lead to error, e.g., in terms of 'Groupthink' (Janis 1971). Janis (1971) describes how the members of a network may fall into a mode of thinking where maintaining the cohesiveness of the 'ingroup' becomes the priority and concurrence-seeking, to avoid member conflict, dominates thought processes to the extent that plausible and potentially controversial alternatives are not explored. This suggests that the *process* of achieving consensus (as a social activity) is important, which again relates to the first research question (Q1) but also links to the second research question (Q2). Measuring consensus as a means to evaluate sensemaking is explored in chapter 3. Third, the question remains as to how one might evaluate meaning. In many situations, meaning could be judged against some notion of 'ground-truth' (section 2.4). In terms of the second research question, (Q2) this could allow technology to provide the sensemakers with the elements of 'ground-truth', e.g., presented through various information sources or sensors in the environment, with the requirement to assemble these elements into a 'sense' of the situation. This further suggests that a key aspect of sensemaking lies in the process through which people gather, interpret and share information. Weick's (1995) discussions are less forthcoming on the mechanics of such processes although he does highlight the importance of collaboration in sensemaking.

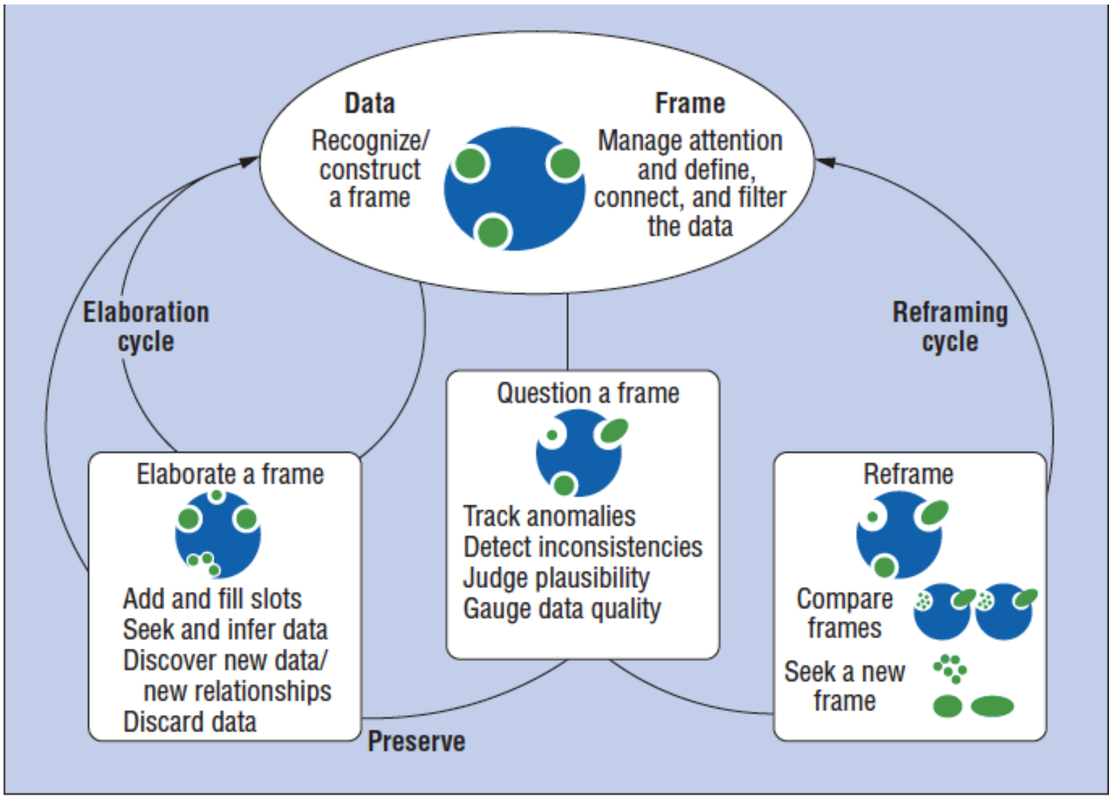

FIGURE 3 - DATA/FRAME THEORY OF SENSEMAKING (KLEIN ET AL., 2006)

Klein et al (2006a) and Pirolli and Card (2005) attempt to rationalise the process of sensemaking as the transformation and manipulation of data. Klein et al (2006a) suggest that sensemaking is a process of simultaneously fitting data into frames (which, as noted earlier, are related to schema or scripts) and redefining the frames to explain these data (Figure 3). A *frame* is an explanatory structure that links available elements in terms of their relation to other elements. It could be a story that describes the chronology of events and the causal relationships between them (Klein et al 2007). Where the frame cannot connect the data, the data can be reconsidered in terms of the frame or the frame can be changed to fit the data. In practical terms, the search for data creates or defines a frame that constrains the domain for further data searching.

"The purpose of a frame is define the elements of the situation, describe the significance of these elements, describe their relationship to each other, filter out irrelevant messages, and highlight relevant messages. Frames can organize relationships that are spatial (maps), causal (stories and scenarios), temporal (stories and scenatios), or functional (scripts)"

– *(Klein et al., 2007)*

It is important to note that this model does not assume that either the search for data or the selection of a frame must happen first; rather both are reciprocal activities. An understanding of the environment is the ability to explain the relationship between data in terms of a framework. Klein et al. (2007) present nine assertions of their data/frame model:

1. **Sensemaking is the process of fitting data into a frame and fitting a frame around the data.**
2. **Therefore, the data are inferred, using the frame, rather than being perceptual primitives.** It is the frame that allows the identification of data as being an important cue in the environment. People with different aims or expertise will acknowledge different things in a situation as being important.
3. **The frame is inferred from a few key anchors.** Generally it is the initial few data elements that are experienced in a novel situation that serve as anchors for creating an understanding and for searching new data.
4. **The inferences used in sensemaking rely on abductive reasoning as well as logical deduction.** Sensemakers will tend to speculate about causes given the effects than to infer effects given the causes.
5. **Sensemaking usually ceases when the data and frame are brought into congruence.**
6. **Experts reason the same way as novices, but have a richer repertoire of frames.**
7. **Sensemaking is used to achieve a functional understanding – what to do in a situation – as well as an abstract understanding.**
8. **People primarily rely on just-in-time mental models.** People tend not to have comprehensive mental models outside of their specialist domains. i.e. a car mechanic has a comprehensive mental model of a car's braking system. The average person has incomplete ideas about such systems and would develop their own "gappy" model of the system as it is needed. Just-in-time mental models refer to the idea that novices and even experts construct these models at the time they are required.
9. **Sensemaking takes different forms, each with its own dynamics.** Seven different forms or activities of sensemaking according to the data/frame model are identified and shown in Figure 4.

(Klein et al., 2007)

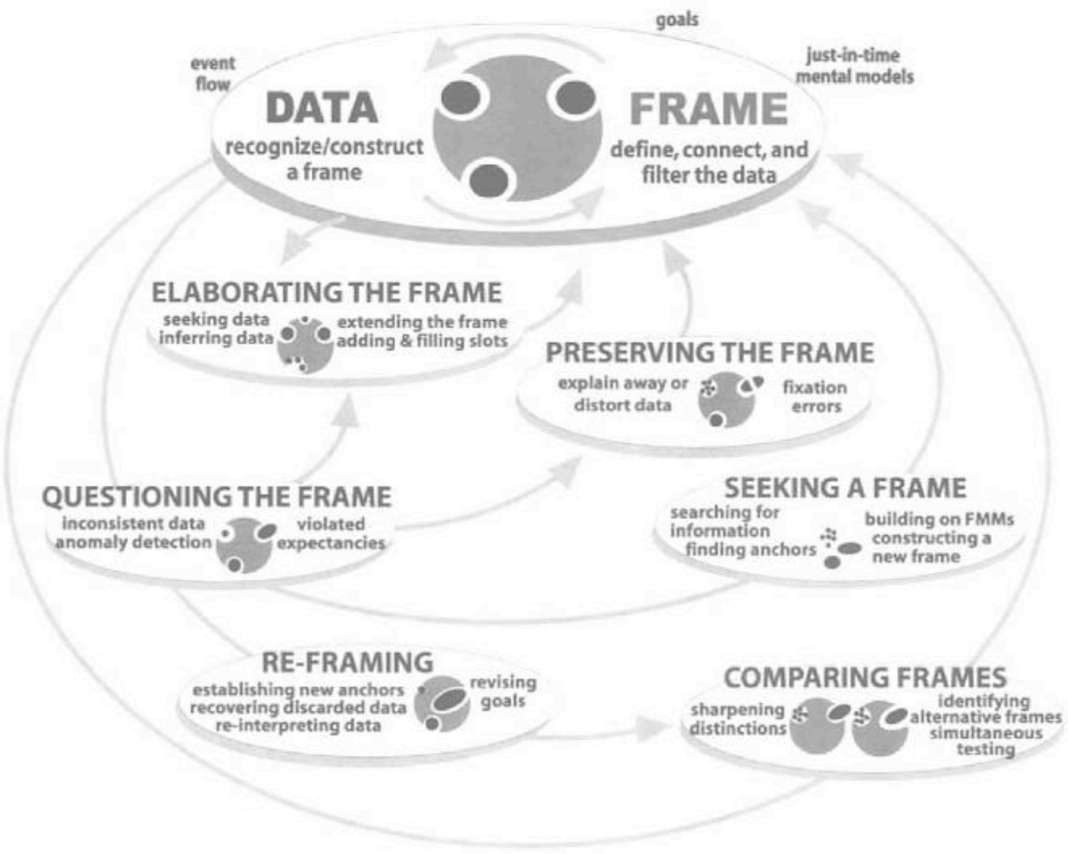

FIGURE 4 - THE SEVEN SENSEMAKING ACTIVITIES (COPIED FROM KLEIN ET AL., 2007)

A potential weakness of the data/frame model is that it is not explicit how this definition of sensemaking can be applied collaboratively. Based on the notions of schema, it difficult to imagine how sense is exported between individuals. This is issue is discussed further in section 2.3.

Pirolli and Card (2005) describe a process (Figure 5) that structures raw information through a process that generates schema, with supporting evidence, that can be the frame of a plausible story, which represents an output to the process. The model was built using cognitive task analysis to study the generalised activities of intelligence analysts as they gathered, filtered and made efforts to understand information. The overall process is arranged in two major loops: a *foraging* loop and a *sensemaking*[1] loop. The foraging loop contains activities for seeking, searching and filtering information

[1] The authors actually use the term 'sense making' in their written description but use 'sensemaking' in

from external data sources into a store ("shoebox"). Evidence is created by reading and extracting information from the shoebox, which supports a theory. This process leads to the sensemaking loop where evidence is fitted into a schema in an attempt to build a case for a hypothesis. The model also contains 'top-down' processes that allow for the re-evaluation of evidence based on the re-assessment of theories. Also taking another step down, the re-evaluation and search for new relationships between extracted information is triggered by changes to the evidence file. Gaps in the information contained in the shoebox can lead to another search through the raw data available.

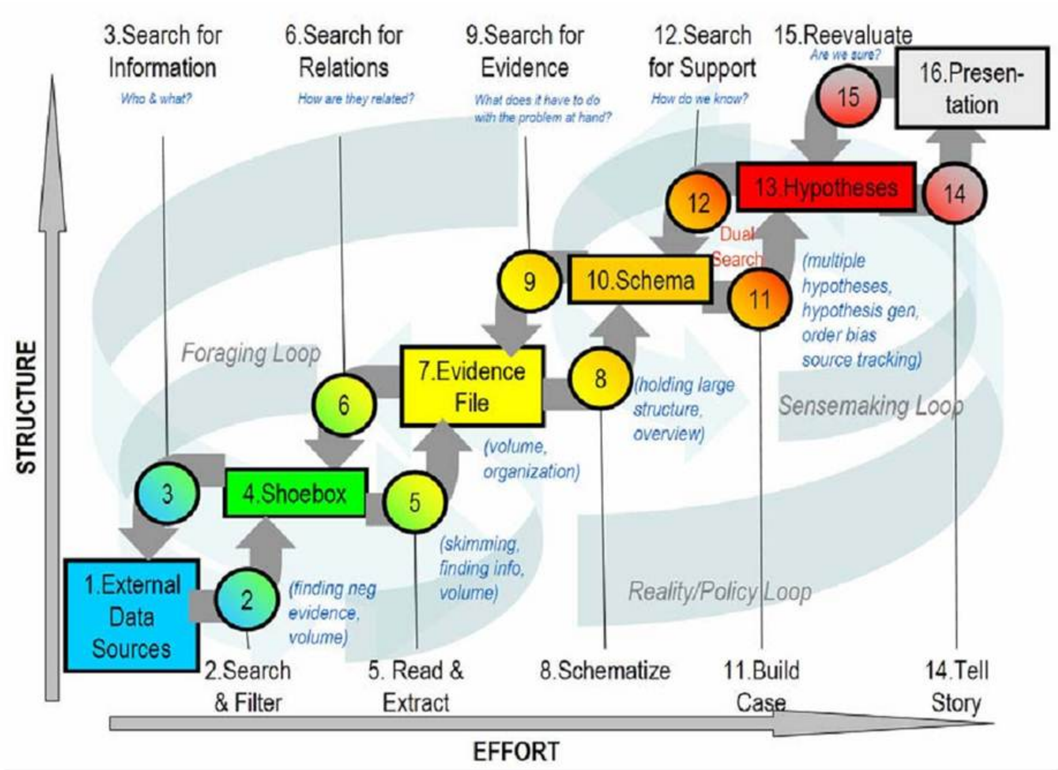

FIGURE 5 - NOTIONAL MODEL OF SENSEMAKING LOOP FOR INTELLIGENCE ANALYSIS DERIVED FROM COGNITIVE TASK ANALYSIS (PIROLLI & CARD, 2005)

Pirolli and Card (2005) identify leverage (or pain) points in their model of information where the analyst might experience problems and which could be areas where new tools or technology could assist. Most of these leverage points are identified in the foraging loop where raw data is *explored*, *enriched* and *exploited*. Exploration of the information space provides the initial data that will be narrowed down through an enrichment process, which creates "smaller, higher-precision sets of documents" (Pirolli and Card, 2005). Exploiting these smaller document sets involves a thorough reading of materials and generating inferences and noticing patterns. The authors posit that there is a trade-off to be made in this exploring-enrichment-exploiting process. In an ideal situation an analyst would prefer to explore as much of the available raw-data as possible. However, this has a knock-on effect in the enrichment process because all of the collected data has to be analysed again. The cost of performing a short or brief exploration is that a crucial bit of data might be missed. From a disaster management standpoint, this model struggles to explain what would happen if the environment that was being explored was dynamic or changeable. Perhaps one of the strengths of the data/frame model is that the data could be changing but the simultaneous process of reframing could capture this. It is unfortunate that Pirolli & Card (2004) are not explicit about their methods for data collection or the presentation of any empirical evidence they used to formulate their conjectures.

The Pirolli and Card (2005) model of sensemaking implies a more linear process in comparison to Klein et al (2006). Despite feedback loops being inserted, they do not seem to be a major consideration in the process model, suggesting that sensemaking is just a case of following the process and arriving at a conclusion. Thinking of this as a collaborative process, it could be inferred that collaborative sensemaking was then a case of ensuring that all members of the sensemaking system followed the process in a similar way. However, at what stage of the process collaboration takes place is less obvious. Alternatively since the process has outputs at each stage, it could be inferred that any of these outputs could be given as artefacts to others to engage their sensemaking processes. For example, the initial search for the 'shoebox' of information could be performed and the information given to others to finish the process. In this instance, this is not strictly collaboration (section2.3) but a similar behaviour is seen in the ELICIT experiment (chapter 4) whereby individuals seem to filter the pertinent information from the environment and share it with others in the organisation. More obviously for this model, the final sub-process ("16. Presentation") suggests that collaboration could become the sharing of hypotheses to ensure plausibility and consensus. Equally, collaboration could be about the sharing and comparison of schema used to create the hypothesis. The question raised is whether, through this model, we are visualising the sensemaking network as a collective of single sensemaking entities or if the network itself is its own autonomous system.

In contrast, the Klein et al (2006) model does not have a starting point and as such we can imagine that any point in the model may act as a jumping off point for the sensemaking process. These considerations become important when contemplating Q2 (section 1.4.2). The implications for the design of supporting technology from the Pirolli and Card model might be to add supporting technology for each step of the process as the sensemaker iterates through the process. Such an approach would be more difficult for the Klein model where it is not clear what support would be necessary since the stages are defined less clearly.

These explanations of sensemaking can be linked on the basis that they explain sensemaking in terms of creating meaning. They emphasise the semantics of data in particular environments. Where sensemaking is a prerequisite to decision-making we would expect to see evidence of a logical process of information collection, comprehension and the generation of comparable option sets. At an extreme, this could involve a detailed process of analysing, interpreting and codifying the data in order to ensure that the appropriate semantics are elicited. One could imagine that situations involving business analysts looking for a new market opportunity or criminal investigators examining a possible case of fraud would follow such an approach.

However, research from the field of Naturalistic Decision Making (NDM) indicates that other situations are far less obviously concerned with the definition of semantics. From studying decision makers in natural settings it was found that people do not seem to be generating and comparing option sets but in fact were using prior experience to rapidly categorise situations that would then suggest a course of action (Klein 2008). These decision making practitioners are still doing some degree of sensemaking i.e. perception and comprehension, but the view of NDM is that any action decided on is primed by prior experience of the type of problem. In such instances, the decision maker is able to pick the most appropriate 'frame' very quickly and use this to guide the response to the situation. This recognition-primed decision-making (RPDM) suggests that skilled decision makers see familiar patterns in data from the environment. This would explain how highly experienced decision makers such as emergency responders can respond quickly, efficiently and usually with some degree of accuracy to, what seem like, chaotic events. They are able to apply previously held mental models to kick-start their decision making without the need for the semantic processing of data. For these reasons technologies that have sought to aid a more formal semantics driven sensemaking have failed to be adopted, as expert decision makers have found them cumbersome and irrelevant to the purpose for which they were designed (Yates et al., 2003). In the NDM research it is difficult to ascertain the exact nature of the inner processes of the decision maker. The results of such studies are qualitative assessments of processes that are not obviously observable. However, the *output* of the process is observable. Once sense is made and a course of action is decided upon, the suitability of the action can be assessed. As such, a more pragmatic and measurable definition of sensemaking can be imagined that considers this output action.

2.3 Collaboration

From the title of this thesis, there is a question of 'collaboration'. The difference between collaboration and cooperation is subtle. Collaboration involves people working together, i.e., co-labouring with the aim of achieving a shared goal. Cooperation involves people performing together, i.e., being in the same space but as individuals they are interested in selfish goals. Reverting back to the example of people driving down the motorway (section 1.1); people are cooperating since they have a common interest in keeping the traffic moving as it helps each individual achieve their goal of getting to his or hers' destination. In this sense they are aligning or coordinating their behaviours, as it is mutually beneficial. The idea that alignment represents a sensemaking effort is explored further in chapter 3. It could be argued that people on the motorway are also collaborating towards the goal of keeping the traffic moving. A better example of collaboration might be an orchestra where everyone is following the same script in an attempt to achieve the same goal of producing the collective sound of the oeuvre. They are working together to achieve something bigger than the individual's part.

It is difficult to see from the explanations of processes provided by Klein et al(2006a) and Pirolli and Card(2005) how the sensemaking process works beyond the single organism. Expanding out from the individual to the group, problems of shared representations, over-lapping mental models and the common knowledge effect are encountered (Gigone 1993). Inherently sensemaking has to be collaborative since any output of the sensemaking process must be socially plausible. As such in collaborative sensemaking, how consensus is achieved is key to understanding the process. Weick(1995) recognises this, yet fails to elaborate on generic processes of how sensemaking in organisations may occur. So our departure point for understanding collaborative sensemaking starts by placing less onus on the cognitive processes of the individual and more on the sharing of information and consensus of group members. One could propose that the 'organisation' could be treated as an individual and that it behaves in a similar manner to the individual. From this perspective, the descriptions of Klein et al (2006a). and Pirolli and Card(2005) simply scale up to something which operates as a collective rather than an individual. From this perspective, sensemaking is something done by a 'system'. While there might be some merit in such an assumption (not least because it makes discussing differences between the individual and social much easier), it runs the risk of obscuring the more important dynamics in collaboration between people in the collective. Therefore, it makes sense to consider whether there might be other ways of considering sensemaking as a collective activity.

De Jaegher and Di Paolo (2007) in their landmark paper, introduce the idea of participatory sensemaking that has fuelled new debate in the field of social cognition. It stems from the enactive approach to cognitive science and focuses on the coupling between agent and world and considers experience as an essential aspect of cognition. Sensemaking under this paradigm is to do with the creation of meaning through an autonomous system's interaction with the world.

> *"Sense-making is the interactional and relational side of autonomy. An autonomous system produces and sustains its own identity in precarious conditions and thereby establishes a perspective from which interactions with the world acquire a normative status" - (Thompson & Stapleton 2009)*

When talking about an autonomous system they are referring to a system that is able to sustain itself independently and regulate its interactions with its environment. Thompson & Stapleton (2009) present an abstract view of what an autonomous system entails. The constituent processes of an autonomous system must:

1. *Recursively depend on each other for their generation and their realization as a network;*
2. *Constitute the system as unity in whatever domain they exist; and*
3. *Determine a domain of possible interactions with the world.*

Examples of what might be considered an autonomous system include single cells, the nervous system, sensorimotor systems, the multicellular body of metazoan organisms, the immune system, and animal and human social groups (Thompson 2007). Of interest to this thesis are the dynamics of human social groups and the coupling, not of agent and environment, but between agent and agent, and their joint sensemaking of an environment. This again is defined by enaction but is reliant on the phenomenon of coordination that exists between coupled systems. De Jaegher and Di Paolo (2007) explain that "*Coordination is a ubiquitous phenomenon in physical and biological systems*"; a claim they support with numerous examples. This has been found to be true even when the coupling between system components is weak. It is highlighted that coordination is easy to achieve and when cognitive beings are involved it does not take any sophisticated cognitive mechanism to achieve coordination. To clarify, coordination is defined as:

"the non-accidental correlation between the behaviours of two or more systems that are in sustained coupling, or have been coupled in the past, or have been couple to another, common, system"

(De Jaegher and Di Paolo, 2007)

Using this model, collaborative sensemaking is about the coordination of action. We know that sense has been made because the output action is appropriate. From this perspective, it becomes less important to know about the semantics of the situation and more important to know about the response that has been 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. From this perspective, it is possible to contrast the semantic point of view (in which sensemaking is concerned with defining meaning in the environment) with a pragmatic point of view (in which sensemaking is concerned with recognising constraints that might affect correct action). In terms of sensemaking as a collaborative activity, these constraints could also relate to the connections that one might have with other people in the network. Thus connecting to people with similar information or views might have a low overhead (in terms of forming common ground) but low information content in terms of semantics (because you do not have to explain what you mean to them but they cannot provide you with anything you do not already know). Whereas, connecting to people with different views or information might have higher overheads but higher information content. This echoes Granovetter's (1973) work on the strength of weak ties but also considers the overheads of coordination between the agents. Communications overheads, dedicated to establishing coordination, are a concern in all forms of networks. For example, packets of information traversing the Internet need the necessary TCP/IP headers to be utilised correctly. In the same way, when people meet, there is an initial establishment of cultural protocols that need to be observed. The importance of enactment in sensemaking was recognised by Weick (1998);

"People often don't know what the 'appropriate action' is until they take some action, guided by preconceptions, and see what happens. Action determines

the situation: The product of action is an orderly, material, social construction that is subject to multiple interpretations." (Weick 1988)

Pragmatic sensemaking can also be found in the fields of teamwork and team cognition in different guises. Carroll et al (2012) describe a theory of *activity awareness*, which is a continuous process that maintains group cohesiveness during a task. They summarise the elements that are important to activity awareness in the figure below:

Facet of Activity Awareness	Description
Common Ground	**A communication protocol** for signalling and enhancing shared knowledge and beliefs.
Communities of practice	**A coordination protocol** for developing and applying community-specific practices through enactment.
Social capital	**A cooperation protocol** of resource exchanges that engender and sustain generalized reciprocity and trust
Human development	**A group regulation protocol** encouraging innovative decisions and approaches in open-system problem solving to evolve group capacities and performance

FIGURE 6 - FOUR FACETS OF ACTIVITY AWARENESS (CARROLL ET AL 2012)

2.4 Ground Truth, Situation Awareness and Sensemaking

Ground truth and situation awareness are concepts that are often associated with sensemaking and as such, are explained for use in this thesis. The concept of *ground truth* is most commonly used in the field of remote sensing where it refers to the relationship between the articles captured from a sensor such as a satellite or aerial photo and the actual confirmed data known to be present in the area where the image was captured. For example, in the case of an aerial photograph this might be achieved practically by sending out people to the area of focus to verify the existence of artefacts captured in the sensor data. Perhaps a more generic definition is that it is a set of measurements that is known to be much more accurate than the system under inspection. Its use within military organisations is taken to mean the actual provable facts of a situation as opposed to what information-gathering resources propose the reality of the situation to be. Similarly in disaster management or emergency response, it is the data that is verified by the trained disaster assessors on the ground that can be used to confirm the reports of the public. In this scenario, ground truth becomes more about the information collected or the sense that is made by a trusted system as opposed to an untrusted system (i.e., the trained disaster assessors are considered better at reporting information about the disaster than the average member of the public). As a definition for use in this thesis, ground truth can be considered to be the things that are known to be true. It is the unquestionable truths in a given scenario used as points of comparison for any system's proposed sensemaking interpretation of that scenario.

Another concept commonly associated with sensemaking is 'situation awareness'. Mica Endsley (1995) gives one of the most comprehensive definitions of situation awareness (SA);

> *"the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future," (Endsley 1995).*

Endsley's model below (Figure 7) highlights the importance of the individual in acquiring SA and lists factors such as previous experience, the individual's goals and capacity of their information processing mechanisms as being important in the process. Hence systems that are designed to aid operators achieving SA must be capable of providing the suitable information and presenting it in an appropriate manner to the user. The psychological school of thought takes the position that several cognitive processes underlie the development of SA (Sorenson & Stanton 2010) and as such maintaining SA is limited by the individual's cognitive capabilities. Generally it is expected that incomplete or inaccurate SA will lead to poor performance in calculating the correct course of action for an identified situation (Endsley 1995). However, the converse is not true; having good SA will guarantee good performance since selecting the right course of action may be determined by factors such as the individual's experience. Endsley's individualistic approach to SA has been criticised for its inability to scale and explain the SA of teams (Sorenson & Stanton 2010). *Shared SA & Team SA* in Endsley's view is made of individuals who each have their own SA is relation to their own goals and the SA of the team can be explained by the sum of the team members' SA. This approach has been challenged as inadequate. *Distributed SA* is an approach to SA that goes beyond the psychological study of the individual and focuses on whole systems (Stanton et al., 2006) Such systems could be teams of the people or socio-technical systems of people and machines. Artman & Garbis (1998) highlight the limitation of earlier SA studies (they cite Vidulich, Dominguez, Vogel & McMillan, 1994; Carretta & Lee, 1995; Wang & Houck, 1995 as examples) as being focussed on an individual's cognitive processes and capabilities. However, since most dynamic decision making scenarios are carried out by teams they expand the discussion to talk about the properties of the system of actors involved. The discussion leads into one of distributed cognition where people, and the artefacts they use, become a "joint cognitive system" (Hollnagel, 2001), where cognition is the output of the coordinated work. The task then "is to describe how cognition is distributed and coordinated (Artman & Garbis 1998).

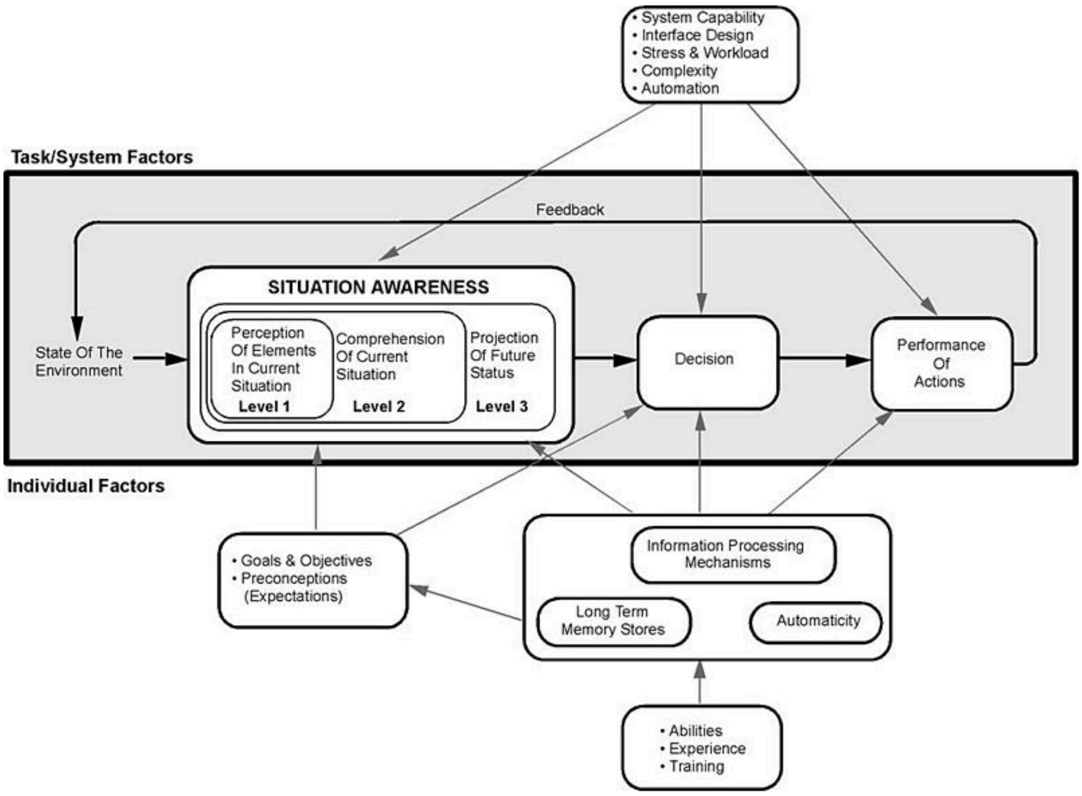

FIGURE 7. MODEL OF SITUATION AWARENESS IN DYNAMIC DECISION MAKING (ADAPTED FROM ENDSLEY 95)

In the same way distributed situational awareness (DSA) is proposed by Stanton et al. (2006) as a way of looking at SA that is system oriented. The approach focuses on how SA is distributed and coordinated which is done through the analysis of interactions of system components (human and non-human). The exchange of information, relevant to the situation, between these components is how we view DSA. Each exchange can be viewed as an exchange of SA and Stanton et al.(2006) go on to name this process transactional SA. In summary, it can be seen that some of the processes that underlie situation awareness theory cross-over with those of sensemaking. To clarify the difference between the two concepts, situation awareness is about the knowledge state that is achieved and the inferences made from this position. Sensemaking comparatively, is the process of achieving this knowledge state (Klein et al 2006).

2.5 Definitions of sensemaking

2.5.1 Individual Sensemaking

Individual sensemaking, following the notions of Gary Klein, operates as an internal cognitive activity, in which data are fitted into frames (schema) while simultaneously reassessing the frame in terms of the data in an attempt to create a coherent account i.e. sense, of a novel situation. Data collection is affected by the individual's experience and expertise.

2.5.2 Collaborative Sensemaking

One of the more significant contributions offered in this thesis is a novel definition of collaborative sensemaking. Seperating the notion into two modes allows us to separate the challenges of studying the creation of meaning and the mechanisms by which meaning is shared and enacted. The author puts forward the term *semantic sensemaking* to not only describe the cognitive processes of individual sensemaking but also the meta-cognitive processes that are concerned with thinking about what needs to be thought about. Semantic collaborative sensemaking is thus about the collective gathering, sharing and understanding of data to produce a shared coherent sense of a situation. The production of a common ground is important to this process and reflects the collective trading of information and understanding. Essential to the Author's notion of sensemaking is Clarke and Brennan's (1991) concept that common ground as mutual understanding can never be perfect but is sufficient to move forward with the conversation or enact a response. Simply giving information to someone is not enough to help their understanding of the situation; common ground suggests another mode of sensemaking, which is less to do with agreed meaning but more to do with an appropriate agreed response.

This second mode of sensemaking is *pragmatic sensemaking*. This reflects the need to appreciate the constraints under which collaborative sensemaking takes place. People working collaboratively in situations need not only to make sense of information but also to make sense of where this information should be shared and how it should be acted upon. Pragmatic sensemaking acknowledges the importance of the organisational dynamics of the people doing the sensemaking. So if semantic sensemaking is to do with thinking about what needs to be thought about, pragmatic sensemaking is to do with thinking about who needs to be involved to make collaborative sense. The Author does not envisage that these two modes of sensemaking happen separately or are subsequent to one another but occur simultaneously as part of the collaborative sensemaking process. However, it is envisaged that at various points in the sensemaking process, activities relating to one mode might be more utilised than the other.

2.6 Sensemaking in Disaster Management

The Author has interests in disaster management and emergency response and these fields will be recurring themes through out this thesis. Sensemaking is a key component in these fields and as Landgren (2005) notes, the complexity and uncertainty of the scenarios encountered make people's sensemaking efforts visible. This makes these fields fertile for the discussion and study of the process.

Crisis, *Disaster* and *Emergency* are terms used interchangeably when they should not be. Klein (1988) defines a crisis as "characterised by low probability/high consequence events that threaten the most *fundamental goals* of an organization" (Klein 1988). A disaster and an emergency 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 preparedness is concerned with providing the training for dealing with routine events (car crashes, fires). Disasters are not always surprise events, i.e., hurricanes are usually tracked for days before they make landfall, however the exact consequences of each reaching landfall are less predictable. Such events put demands on sensemaking. Thus disaster management involves organising the most effective response network of emergency responders, public and private organisations, and volunteers suited to dealing with unknowns. Hence this field offers good examples for observing collaborative sensemaking in the real world.

Emergency responders tend to be highly trained individuals working in teams, the purpose of their training being to maintain the functionality of a community of practice (Wenger et al 2002). Relating to Q1.1 in this thesis, they tend to operate in a hierarchical structure dealing with routine incidents. Conversely, in the field of disaster management, personnel are usually working in exceptional circumstances and new organizations or teams of people form around problems. These ad hoc teams are often only in existence for the duration of that particular problem. When dealing with unknowns, these groups work as networks of exploration (Baber et al, 2008) or adhocracies (Travica, 1999). These networks form around problems, have low formalisation of behaviours, few standard operating procedures and tend to have high capacity for problem solving and creating innovations. Such networks 'do sensemaking' and can start as loosely defined groups with little consensus and differing ideas as 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. That output is not the end of the process, since the process is one of continuous refinement of a plausible working hypothesis. As such, McMaster & Baber (2011) suggest, in addition to emergency responders following Standard Operating Procedures (SOPs), there is a need for them "to practice generic skills related to information sharing and collaborative sense-making".

2.7 Wisdom of Crowds

Surowiecki (2004) has shown that groups of people can be powerful sensemakers and decision makers under certain conditions; a phenomenon he presumes is a collective evolutionary trait. The 4 principle conditions he believes to be essential for the formation of a wise crowd are; *diversity*, *independence, decentralisation* and aggregation. *Diversity*; each member of the group should have their own representation of the information or facts and their own variance in cognitive processes. *Independence*; group members are free to form their own conclusions without the influence of other group members. Independence in essence means that a group member is allowed to build his/her own representation of the problem i.e. do their own individual sensemaking first. *Decentralisation*; group members are able to use their own local or specialist knowledge to form an opinion or answer to the problem; this also encourages independence. Some companies that try to make use of the wisdom of crowds within their own organisation sometimes confuse this idea. The premise is not to find consensus though democratic discussion but to spread decision making power throughout the group and allow them to find their own answers as individuals. However decentralisation comes at the cost that any important information discovered by one part of the system may not diffuse across the rest the system. Sometimes valuable information might not get disseminated. Thus a system has a higher chance of being intelligent if there is a means of aggregating information system-wide. This should be a consideration when contemplating Q2 (section 1.4.2).

The wisdom of crowds notion opens the door to a slightly different form of collaborative sensemaking. The sensemaking literature assumes that people have to communicate to form consensus but the pragmatic approach to sensemaking suggests this is not necessarily true e.g. the example of people driving on the motorway. The wisdom of crowds is about the aggregation of individual sensemaking, or its outputs, and the discussion is about the creating the conditions under which it can occur. As a thought experiment, one might imagine a group of people in a room and the number of ways in which a single collective decision could be attained from them. One might ask them to vote on predefined courses of action and see which is the most popular. One could also ask them to individually suggest a course of action and see which was the most common idea. The latter idea is akin to the monitoring of collective behaviour seen in the wisdom of crowds literature. Another option might be to ask them to organise themselves and discuss until a consensus had been reached on a course of action; akin to a collective semantic sensemaking. This thought experiment reminds us that the crowd can be a powerful decision making collective but only under the right conditions and only when asked the appropriate question.

Approaching from the viewpoint of knowledge management, Oinas-kukkonen (2008) echoes the principles of Surowiecki yet criticises him for not formally describing the operation by which this approach of seeking new knowledge from the crowd can be achieved. He extends the principles and takes into account networking aspects of the approach and proposes 8 conjectures:

1. *It is possible to describe how people in a group think as a whole.*
2. *In some case, groups are remarkably intelligent and are often smarter than the smartest people in them.*
3. *The three conditions for a group to be intelligent are diversity, independence and decentralization.*
4. *The best decisions are a product of disagreement and contest.*
5. *Too much communication can make the group as a whole less intelligent.*
6. *Information aggregation functionality is needed.*
7. *The right information needs to be delivered to the right people in the right place, at the right time, and in the right way.*
8. *There no need to chase the expert.*

Oinas-kukkonen (2008)

Points 5 & 7 are of interest in relation to Q2. Point 5 highlights that a decentralised group might be prone to information overload at some nodes and there is a real need for efficiency of information sharing, especially when the group size becomes large. More information does improve sensemaking performance up to a point but past that point additional information may cease to be useful and might even degrade performance (Klein 2006). Importantly, too much or too detailed communication may kill the desirable trait of diversity of members in the group (Oinas-Kukkonen, 2008). Point 7 highlights there is a challenge to determine what is the "right" information and whom the "right" people are. For the group to share information efficiently there is a need to find out what everyone knows in the network but it may not be necessary for each person to actually have that information. This relates to the concept of *Transactive Memory*, a term coined by Wegner et al (1985) in a paper that describes the nature of cognitive interdependencies in close relationships. Practically it can be imagined as a working index retained by an individual of what knowledge other people in a network have and negates the need for the individual to store that knowledge himself or herself.

Transactive Memory relates to the concept of 'group mind', i.e., the relationship between the mind of an individual working as part of a group and the cognitive operations of that group. This is not dissimilar to the concept of Distributed Cognition (Hutchins, 1995, Scaife & Rogers, 1996, Baber et al, 2006), in which cognitive processes could be allocated between members of a group or the results of cognitive operations could be distributed through the use of shared representations. Cross et al (2001) without explicitly mentioning the term, recognise the importance of Transactive Memory in organisational knowledge management. That is to say that the managers that they interviewed gained much critical information for project success far more frequently from other people rather than any official organisation knowledge management book or database. The managers maintained a knowledge cache of what people in their organisation knew. Fundamentally a transactive memory network operates three core tasks, the encoding, storage and retrieval of knowledge. These functions can be viewed as analogous to an individual's memory system. Since the knowledge stores of a transactive memory system (TMS) are physically separated, it is communication that acts as the conduit for the knowledge-relevant transactive processes that occur between group members (Wegner et al, 1985). In the group, these functions are observed and aided by communication; just as in Distributed Cognition, these functions are mediated through shared representations. The potential is that a TMS can be more effective than the sum of the constituent memory systems of the individuals who form it.

In relating this back to research question 2 (1.4.2) in this thesis there are some key points above. In aiding collaborative sensemaking the main focus should be on the appropriate dissemination of information and how representation is shared between group members. There is cause to suggest that a system that would aid sensemaking would maintain an index of what people in the network knew and would show this information to users as it became pertinent to them. In affect this would allow individuals to forge the communication links that would be most useful to them, aiding pragmatic coordination.

2.8 Technology and Sensemaking

Research question 2 (Q2, chapter 1) refers to the ways that technology might be able to support collaborative sensemaking. Disaster management and emergency response are a continuous theme throughout this thesis and this review of technological approaches to supporting collaborative sensemaking will continue in that vein. In this chapter a working definition of collaborative sensemaking was derived that suggested that sensemaking works in two modes. Firstly, semantic sensemaking is presented as a meta-cognitive process that involves thinking about what needs to be thought about. Secondly, pragmatic sensemaking is described as the realisation of the constraints by which sensemaking takes place. Although this thesis is concerned with collaborative sensemaking, the individual is still at the centre of this process and as such the individual's sensemaking processes must be supported as well as the group's collaborative processes. Given the definition of sensemaking above, technologies could support semantic sensemaking and/or pragmatic sensemaking.

Following the Pirolli & Card(2005) perspective, technology could play a key role in structuring the 'sense' through the sensemaking process. This could either involve the technology in the form of artefacts, which contain and transport the output of semantic sensemaking, or could take the form of agents which structure semantic sensemaking (possibly through 'collaboration', or at least interaction, with people). Supporting semantic sensemaking in groups is predominantly concerned with mechanisms for sharing individual representations, discourse on meanings and interpretations, and construction of shared representations (Umpathy 2010). Various technologies have been developed to help support these processes: Compendium (Selvin et al 2001) seeks to support organisational knowledge capture and shared representation. Jigsaw (Stasko et al 2012) and EWall (Keel 2007) represent two technologies that "foster object focused thinking" through the representation of knowledge as objects. These are just a few of the examples found that support the semantic aspects of collaborative sensemaking. In fields like disaster management and emergency response, such tools might feel cumbersome to personnel who have the capability to rapidly categorise seemingly chaotic events by picking salient cues from the environment based on prior experience. These notions could explain why technologies that have sought to aid a more formal semantics driven sensemaking, have failed to be adopted, as expert decision makers have found them irrelevant to the purpose for which they were designed (Yates et al 2003). The author's impression is that the emphasis on the design of technologies that support collaborative sensemaking has been focused to date on aiding and sharing semantics. The emphasis in this research will be on how the pragmatic workings of the sensemaking network can be supported to enhance sensemaking capabilities.

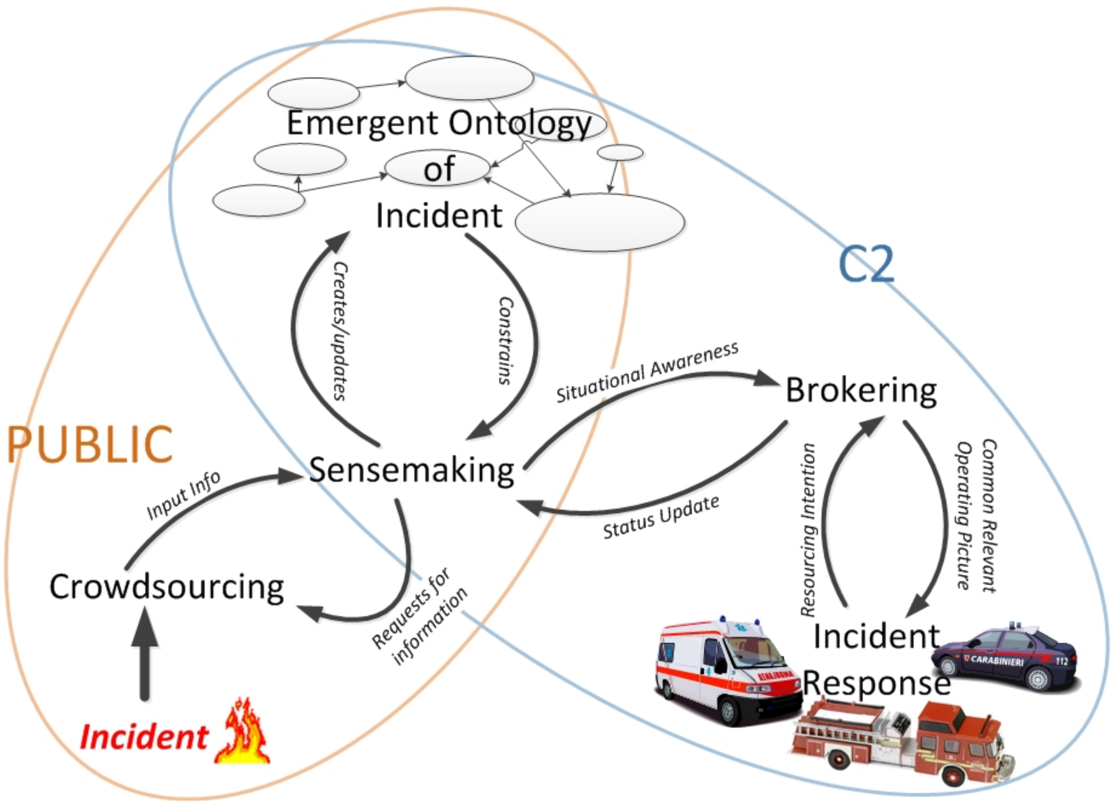

FIGURE 8 - THE META-PROCESSES OF EMERGENCY RESPONSE

In the realm of disaster management, the ability of responders to collaborate during a major incident is difficult because of the instability and changeability of the environment and the high levels of complexity involved. Ascertaining what is going on and maintaining a shared awareness of the situation is important at all levels of the response organisation. A model of perceived meta-processes involved in emergency response is presented in Figure 8. As a thought exercise, the processes of this model and areas for technological support are considered.

Initially an event prompts people to report what has happened (e.g., a major incident). These reports are sent to an incident control room. The initial reports from the public are unstructured and the control room starts a sensemaking process that starts with seeking or building a frame to structure the incoming data into knowledge base for the incident (represented as an incident ontology). The emergent incident ontology constrains the search mechanisms for new information. A plausible hypothesis of the incident environment is developed which is then used as the basis for defining the initial response. This initial response will assess the nature of the incident and report back to control room on the scale of the incident and this initial organisational interaction with the environment will guide future action. Through this, the processes of organisational sensemaking are linked to the brokering of information and resources through status updates and developing situation awareness. Brokering becomes the means through which tactical command is deployed (i.e., by defining which resources that are required or deciding what information is needed to make best sense of the incident in order to maintain situation awareness. Knowledge brokering for shared multi-agency understanding of the incident domain is known as a Common Operating Picture.

McNeese et al. (2006) define Common Operating Pictures as the representation of information in order to generate situation awareness across team members. The Common Operating Picture will generally be managed at the incident control room or EOC, rather than on-scene. It has the potential to facilitate decision making, situation awareness, collaborative planning and assists the various levels of command across the services in achieving shared awareness of the situation. As well as ensuring that only relevant information is passed to the personnel at the scene, a Common Operating Picture with multiple layers would allow commanders to have oversight of the status and distribution of the other services, enabling implicit collaboration (Baber et al., 2007; Keuhlen et al., 2002). Essentially the Common Operating Picture is an output of a collaborative sensemaking effort, in which a plausible explanation of the environment has been established to the extent that a coordinated response to the situation can be attempted. Figure 9 shows a screen shot of GATOR, which is the state of Florida's support tool for Common Operating Picture. The system feeds in information from various official response agencies and visualises the on-going situation using a map and GIS system. Information is accessible on different levels dependant on one’s role within the system. The emphasis with this technology is still on supporting a shared representation without any direct support for pragmatic sensemaking.

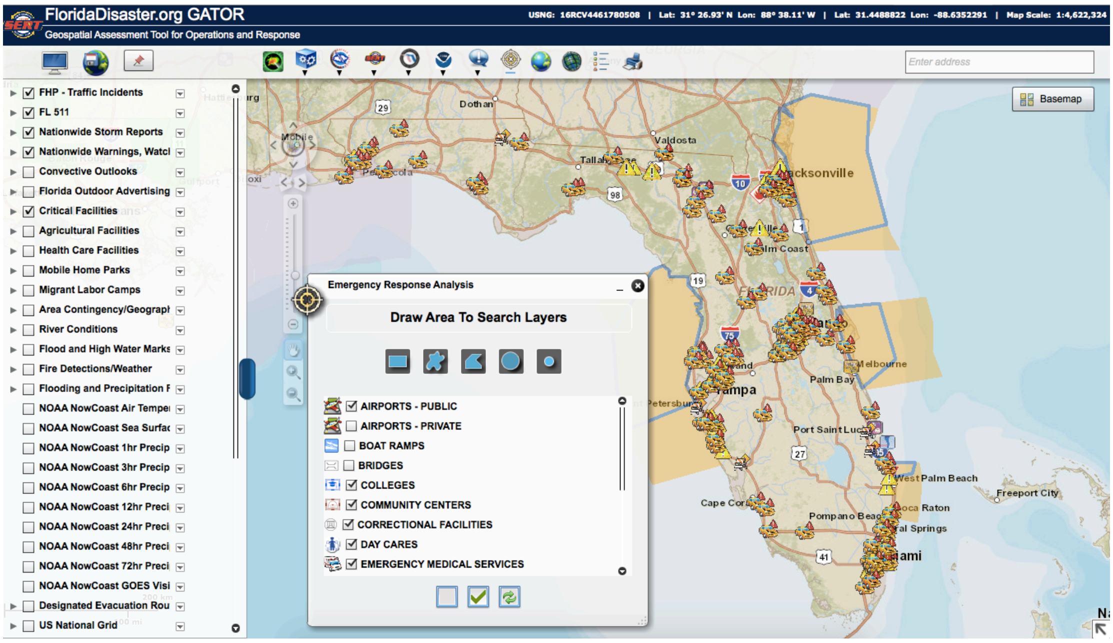

FIGURE 9 – GATOR

A criticism is that the common operating picture seems to work when the situation is defined. When the situation is well defined responders know what the are dealing with and can follow normal SOPs. Information sharing is simple in such a scenario since classical hierarchical command organisations are built for distributing information efficiently. However, if the responders are dealing with high levels of uncertainty it becomes less clear where information needs to be sent as their social structure is put under pressure. The Author's conjecture is that what is needed is a shift from information brokering to communications brokering. In this scenario the brokering is about establishing *who* to talk to and not *what* to talk about. In this way, collaboration is the driving factor in the interaction and, potentially, this could support pragmatic sensemaking (in terms of establishing communication links as action).

2.8.1 Communications Broker

In this thesis, the concept of a C*ommunications Broker* as a supportive technology for collaborative sensemaking is explored in chapters 5 & 6. In Social Network terms, the Communications Broker could take the form of a 'boundary spanner' (Thompson, 1967) across agencies, or a 'weak tie' (Granovetter, 1973, 1983) operating between cliques, or an 'information orchestrator' to manage information exchanges across agencies (Bharosa et al., 2008) or, (in military parlance) a Liaison Officer. As discussed in section 2.3, the potential for this role is to aid pragmatic sensemaking by overcoming the barriers or constraints to finding people with useful information. In a sense this lowers the overheads of coordination between agents. While a person could perform this role, due to the amount of information involved, it would be more powerful as a technical system that could track themes in incoming communications to highlight areas of commonality between responders. In network terms, the envisaged broker would operate as an application layer helping to analyse the network layer of communications below it. To do this, the Broker could dynamically extract themes from the communications that could contribute to an ontology of the situation. By keeping track of the personnel related to each theme the Broker fundamentally acts like a map of who knows what, i.e., as a technological proxy for 'Transactive Memory' (Wegner, 1986).

Many military organisations in the world operate a tactical chat system that allows strategic command to update and maintain a Common Operating Picture of operations. Operational Support within the Company Headquarters transcribes radio communications from Ground Force Commanders in the battlefield. This information is then added to the appropriate chat group. For example, there might be a dedicated chat group for medical evacuation air support that lists the location of all currently known causalities. Such systems are, so far, not installed in the emergency services, but given the potential for the logging of radio traffic, the prospect of a software communications broker can be envisaged. Through text analysis the system would extract and organise an interlinked set of key terms that would be put forward as contributions to the incident ontology. These terms can be used to represent themes (Vilhena et al, 2014). The ontology would then define the limits of the domain of the incident and allow the tracking of themes as the incident develops. This could provide an essential contribution to the dynamic construction of a Common Operating Picture. However, rather than flagging these themes to responders, the broker system would suggest links between responders or create a chat group for responders who were including specific themes in their reports. In this way, the brokering becomes not only a matter of managing the ontology but also a means of creating a Community of Practice across agencies. The Broker could then maintain a network of community members who have ownership of particular ontology components. In this way the envisaged broker would support semantic sensemaking and aid pragmatic coordination between service personnel.

Continuing with the Transactive Memory concept, there is an inference that the Broker would be responsible for the maintenance of this interrelated network of themes and people. 'Maintenance' incorporates not only the addition of emergent themes but also the removal or reassessment of old themes. In this thesis, this notion is referred to as *Theme Decay.* As in the literature on other memory systems, the notion of *Theme Decay* is more akin to *Decay Theory,* which is found in studies on memory (Thorndike, 1913;

Melton, 1963) and should be distinguished from *Information Decay,* which exists in Communications Theory literature (Kåhre, 2012). Information Decay is about the accidental decay of information in communications channels due to external interference or noise. This could be a consideration when studying networks of people communicating around sensemaking tasks. For the purpose of this thesis all communication channels are assumed to be complete and coherent. However, the concept of *noise* is borrowed and adapted in chapter 4 to describe the key terms in the experiment that are not part of the solution. From the perspective of information decay, the engineering challenge is to reject unwanted signals (noise) affecting the original communication. From a psychology perspective, noise is the distracting influences that distort ground truth. Rejecting distractors can be seen as one of the activities of semantic sensemaking and a consideration for theme decay. In the ELICIT trail (chapter 4) we see how some of the groups collectively filter noise from the information provided in developing a correct solution. In a disaster management scenario it is probable that different themes will become prominent in communications as the situation develops. As themes cease to be used a decision should be made as to whether they should be discarded as a component or perhaps archived for later retrieval. These concepts feed into a more advanced discussion on knowledge management during an incident and are beyond the scope of this thesis. However, a rudimentary form of theme decay is used in the Broker system in chapter 6 that forgets all the themes extracted in the previous time period when starting a new time period.

2.9 Metrics

Q1 of this thesis asks; can collaborative sensemaking be measured in the behaviour of social networks? Q1.2 elaborates to ask what social network analysis (SNA) metrics might help describe an organisation doing sensemaking. One of the focuses of this thesis is to find metrics for measuring the behaviour of networks doing collaborative sensemaking. As mentioned in chapter 1, the inner cognitive processes of sensemaking can not be measured directly. Thus, novel methods and measurements must be found that will allow the inference of successful sensemaking from measurements of outputted observable behaviours. If a key aspect of collaborative sensemaking is the manner in which a group of people can share information, it becomes relevant to define and measure the manner in which a network operates. One approach to understanding networks can come from Social Network Analysis, and metrics related to SNA can help define the structure of the network and the manner in which collaboration occurs. In addition to metrics that relate to collaboration, the thesis requires metrics, which can define the semantic sensemaking of the activity. This could be related to the 'sense' that was made and how this related to the ground truth. In this case, the metric would compare the 'answer' that was reached with a 'correct' answer that was defined. For the ELICIT study (in chapter 4), ground truth is defined in terms of the answer that the teams were meant to reach. For the article study (chapter 3), the ground truth is the content of the original newspaper article that was used as the basis for discussion. The content of the newspaper also related to the discussions that were analysed and these could be compared across participants. In order to perform such comparison, the thesis employs a form of signal detection theory. In terms of pragmatic sensemaking, the performance of the teams can be evaluated in terms of how actions are performed or how the teams attempt to solve the problem.

2.9.1 Social Network Analysis

Social Network Analysis (SNA) is the mathematical analysis of human networks, be they kinship, communication or other forms of relationship (Wasserman & Faust, 1994). Oinas-Kukkonen (2008, from cross et al, 2002) notes that "network analysis is a valuable collaboration vehicle in such strategically important groups as top leadership networks, strategic business units, new product development teams, communities of practice, joint ventures, and mergers" Much of this work tends to approach analysis from the perspective of the static structure of a network. This approach can overlook dynamic processes and changes in actor roles because each view of the network structure represents a 'snapshot' of that network (Trier & Bobrik, 2007, Kampis et al, 2009). For the purposes of this work, the aim is to explore techniques in which the *behaviour* of a network can be captured. SNA can identify particular roles in the network, particularly in terms of information dissemination, e.g., a well-connected individual might be classed as an 'opinion-leader' (Valente & Davis, 1999) an individual who connects two groups might be a 'weak tie' (Granovetter, 1973). Given the potential value of these roles in managing the flow of information in networks, it might be useful to determine which individuals or even technologies could be assigned these roles, and whether this assignment might change over time. However, the approach of this thesis takes a more macroscopic approach to the analysis of whole networks and will make less use of SNA to establish the idiosyncrasies of individual network players.

2.9.2 Density

Density is the ratio of the communication links in a network that are being utilised:

$$Density = \frac{Actual\ Connections}{Potential\ Connections}$$

Density can be calculated directionally or unidirectional.

$$D_{directed} = \frac{2E}{N(N-1)}$$

$$D_{undirected} = \frac{E}{N(N-1)}$$

Where E is the number of edges.

Density is a versatile SNA metric. Dependant on the context it can describe various aspects of group behaviour. During collaborative task work, density has the potential indicate preference of group behaviour towards communication. In this context it could be suggested that a team utilising a high percentage of possible communications is focused on teamwork rather than task-work. Density could also be seen as a measure of efficiency when viewed in conjunction group performance. If two similar groups performing the same task get similar performance scores but one has a lower density, it suggests that group was able to complete the task making less use of available connections. It could also correlate with performance where the task dictates sharing. Walker et al (2009) proposes distribution of information in a network can be mapped on to network density. This seems reasonable although it does assume that communication links are continuously utilised in a network to send novel information when it arrives. For example, if a token passed around a network until it had been passed between all combinations of nodes in the network the density would be 1. If new token arrives in that network and is only passed once, the density metric would not describe this. Thus using density as a metric for distribution of information infers that nodes in the network are passing information equally.

2.9.3 Average Communication

Showing average communication per person for a network of people doing sensemaking can give an indication of the emphasis they are placing on teamwork. Used in conjunction with density it can give an indication of the general effort of the group to share information with others in the network. Given by:

$$\bar{x}_c = \frac{c}{n}$$

Where:

$\bar{x}_c$ = Average Communication

c = Total communications.

n = Number of people in network

Equation 1 - Average Communication

2.9.4 Average Activity

Average activity per person is a bespoke measurement defined with relation to each task in this thesis and is loosely defined as the time spent on task. Comparing it to communication rate for a group, gives an indication the preference of that group toward teamwork or task-work. Given by:

$$\bar{x}_a = \frac{a}{n}$$

Where:

$\bar{x}_a$ = Average Communication

a = Total activity

n = Number of people in network

Equation 2 - Average Activity

2.9.5 Signal Detection Theory

Signal detection theory (SDT) is the method by which quantitate measures can be given to the ability of a detecting system, to differentiate between valid stimuli and distracting stimuli in an environment. Pirolli and Card (2005) compares the way analysts filter information to that of a signal and noise problem where by sometimes an expert analysis will look for fainter signals (relevant information) by accepting more noise (irrelevant information). Its use in decision-making studies is recognised but as of yet has not been seen in a sensemaking study.

	Respond Present	Respond Absent
Stimulus Present	Hit	Miss
Stimulus Absent	False Alarm	Correct Rejection

Figure 10 - Signal Detection Theory Matrix

SDT affords the measure of sensitivity of the detecting system. Sensitivity is a function of hit rate and false alarm rate and is represented by the symbol **d’** (referred to as d-prime). **d'** is the difference between the z-transform of the probability of a hit and the z-transform the probability of a false alarm, and is given by the equation:

$$d' = Z\big(p(H)\big) - Z\big(p(FA)\big)$$

EQUATION 3 - D-PRIME

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.

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

$$TPC = \frac{H + CR}{No.Responses}$$

EQUATION 4 - TOTAL PROPORTION CORRECT

3 Article Commentary Experiment

The article commentary experiment investigates the relationship between people's discussions about a newspaper article that was acting as an initial communication. An experiment was developed that allowed people to comment on a topical news article. The domains of the terms used in the article and by the commentators were compared via text analysis. The results suggest that as groups sizes increased so did the preference to use terms introduced in the article rather than the terms that were introduced by the participants. This suggests that participants need to use the ground truth presented in that article as a basis for discussion and the implications for sensemaking are discussed.

3.1 Introduction

In disaster management emergency responders and emergency commanders often need to make sense of a situation during time critical events on the basis of limited evidence. Often the available evidence takes the form of an initial situation report (SitRep), which is then shared with the responding agencies. In order to ensure a coordinated response to an event, it is important for emergency commanders in each participating agency to understand the event in a way that provides some commonality. While this need not mean that all commanders will agree on all of the implications in the report, it is important that they reach some consensus on the nature of the threat they are facing and on the dynamics of the situation. This means that the commanders need to discuss the initial SitRep to agree on its content and implications for response. Such discussion could be verbal, e.g., in face-to-face meetings, but increasingly can take the form of comments made through shared social media. In this chapter, the focus is on the latter. From the perspective of managing collaborative sensemaking, understanding whether and how 'shared understanding' develops through social media can help develop advice on how best to manage collaborative discussion or how to introduce information to such discussion. In addition to the question of whether the contributors to a discussion reach consensus, or shared understanding, it is also interesting to consider when and if contributors rely solely on the information in the initial SitRep or whether they introduce additional information or opinion. A reason for introducing additional material could be to help clarify points in the initial SitRep or to help move the discussion towards consensus. However, such additional material could also lead to distortion of the SitRep's content, which in turn, could result in commanders forming an erroneous view of the situation.

To clarify, the concern is not with everyday emergencies like an automobile accident or a house fire. Although these are events that may involve a multiagency response, they do not necessarily involve a great deal of collaborative sensemaking since emergency responders tend to be highly trained individuals whose experience allows them to rapidly categorise situations and hence they can then carry out trained procedures. Large-scale disasters involving multiple agencies and responders are a good place to observe collaborative sensemaking assuming that it is possible to gain access to timely data wherever the response is being coordinated from. Obviously this is not always possible and the alternative is to assess the practices of the commanders in hindsight following the events. Both approaches pose significant challenges. In hindsight there is a risk that you will miss the data that is not captured on official channels, i.e., the words that are spoken between members in the control room. In capturing data as the event unfurls you would need large resources in terms of people to capture the data that is flowing between people in close proximity (together with appropriate levels of clearance, permissions and health and safety policy to allow attendance of the incident in the first place). Refocusing on the aims of this thesis, the practices observed in emergency management provoke two questions:

1. How can you capture the process of multiple people developing a shared understanding, in a way that allows some experimenter control in a risk-free environment?
2. How can this process be measured?

In response to question 1, by presenting people with an initial communication and asking them to state their reactions as written comments, it should be possible to capture the sense they had made of the content of this communication. It was decided to use news articles as this initial communication in the following experiment. In a practical sense the initial communication could be a report sent about an emerging major event or disaster. It is reasonable to draw this comparison since both news articles and reports contain an initial layer of analysis by the respective authors. In addition to adding comments, it was decided that this could be made a collaborative process by allowing the commenters to reply to each other's comments; in affect giving the opportunity for a discussion to develop.

In answering question 2 above, it is hypothesised that two people would show a shared sense of the communication when they could be seen to be using similar vocabulary to one another. If collaborative sensemaking is about the alignment and coordination of actions, this assumption seems reasonable. The hypothesis is that participants in the experiment would coordinate their responses to the initial communication and that this could be measured through the overlap in the *domain* of language they were using. By aligning their use of language with respect to the initial communication they are coordinating with each other. *Domain* in this instance is referring to the set of key terms that that are extracted from a source. i.e. the domain of the article is the set of key terms that are extracted from the article.

This raises the question, to what end are they coordinating? In this initial experiment, the goal was to contribute to a discussion of a news article. Contribution was entirely up to the discretion of participants, but it was implied that the discussion would stop when the major themes and implications of the article had been explored. This meant that there was no explicit instruction as to when to identify the end-state. It was felt that leaving the definition of end-state open would produce behaviour that was analogous to the network of exploration that one might find in the initial stages of an incident, when commanders are sharing their views, opinions and interpretations of the situation report. This is not to say that participants might not have had their own goals in taking part in the experiment and responding to the communications. The point is that for this experiment they were not given the goal of trying to achieve consensus with one another or to purposely disagree or to make no response.

3.1.1 Hypothesis

The aim of the experiment was to establish how participants aligned their contributions with the content of the article and with other participants' contributions. It is hypothesised that a strong correlation between the article and the participants' language will show that the participants were reliant on the terms in the article in their communications. Key terms extracted from the content (section 3.2.1) can be said to be representative of themes within the content (Vilhena et al, 2014). There is a reliance on common ground to form discussion and as such common ground in this instance can be seen as a set of shared key themes between participants. Looking at the alignment of key terms from the article with key terms used by the participants in conjunction with participants introducing key terms between one another, it can be measured how participants are forming common ground. One conjecture is that participants in larger groups will be increasingly reliant on the article content to situate their own contributions. This is because it was thought there would be a cost in introducing new terms (themes) as is seen with the formation of common ground (Clarke & Brennan, 1991). If key themes are a proxy to common ground and there are social overheads to achieving this, it could also be observed that participants strive to form smaller discussion groups in the comment thread to counteract this.

3.2 Experiment Design

This experiment was carried out in two stages, both with slightly differing methods of data collection. Firstly the author created an experimental platform that participants could use to comment and reply to comments for a given article. Secondly, comments and replies were mined from a popular online news website in an effort to gain larger data sets.

In the first part of the experiment, it was decided not to mine online news websites for the collection of data for two reasons. Firstly, there was an assumption that the Author's platform would not attract the same number of users as the large online news websites. This would allow for the analysis of smaller group interaction that could be compared to the larger groups of users typically found online. Secondly, it was uncertain if there was an editorial agenda for each particular news site. Figure 11 shows an example of a 'comment removed' notice from the Guardian news website. It is imaginable that such comments are removed because of the use of unacceptable language, such as profanities, or unacceptable behaviour, such as aggression or bullying; however this could not be ascertained despite efforts to contact the various news institutions that run such discussion sites. The Guardian website's community standards do suggest that moderation occurs to keep comments relevant to the original article and this justifies the creation of the Author's platform, which had no restrictions on the content of contributions. To counter any moderation in the second part of the experiment, it was thought that selecting online articles with that had a large number of users ($N > 50$)[2] would mask the affect of comments that were removed by moderators. These factors persuaded the creation of a platform where there could be complete control of the data inputted. An additional factor for starting with a self-made platform was that there were not always comment streams available for all of the articles that were deemed appropriate as initial communications for this experiment.

[2] The author's platform had 49 users

FIGURE 11 - COMMENT REMOVED NOTICE TAKEN FROM THE GUARDIAN NEWS WEBSITE.

A second question remains; why not mine data from social media sites? Posts to social media sites tend to be quite short. Sometimes this is because they are restricted by the platform. Twitter, for example, limits its posts to 140 characters. Facebook has a more or less unlimited post length restriction (more than 60,000 characters as of November 2011) although the average number of characters per post is approximately 104 characters (Nierhoff, 2013). This would be an insufficient amount of text to act as an initial report to use text analysis on to assess the domain of the initial communication. Later in this thesis (section 6.2.5), 140 character participant contributions are used where there is not a requirement for measuring alignment between contents.

With this in mind a bespoke platform was built for the experiment that allowed more control and access to all content submitted by participants. The investigation was to see if participants aligned their content with the article and/or other participants. Using Figure 12 we can visualise the possible intersections between the participants and the article or initial communication.

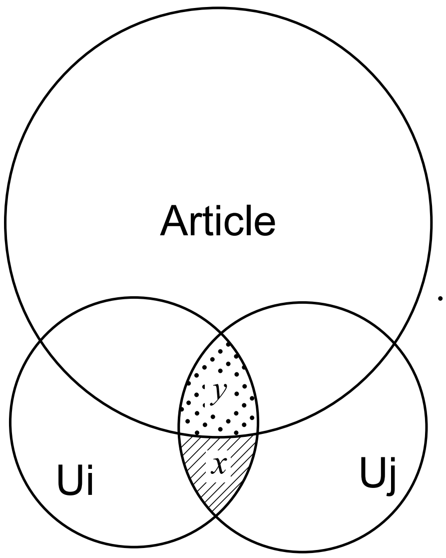

FIGURE 12 - INTERSECTIONS OF CONTENT

To put this more formally we can say:

The relationship between a participant's content and the article content is given by:

$$U_i \cap A$$

The relationship between two participants' content and the article content is given by:

$$y = U_i \cap U_j \cap A$$

The relationship between two participants' content that is not a part of the article content is given by:

$$x = U_i \cap U_j \cap \overline{A}$$

To summarise, '*y*' represents the set of terms that are introduced in the article and utilised by a pair of participants. '*x*' represents the set of terms that are used by a pair of participants but are not introduced in the article. These sets were found using text analysis techniques (section 3.2.1).

3.2.1 Author's Experimental Platform

The platform consisted of a website built on a MySQL server using HTML and PHP scripts (Figure 13). Topical new articles were copied from mainstream news websites (with referencing) and stored in a table in the MySQL database. Participants could browse a list of articles on the client-side website. On opening an article the participant was afforded a text box to submit comments to the bottom of the comment thread. There was also an option at the end of every comment to 'reply directly' to that comment.

FIGURE 13 - CONFIGURATION OF WEBSITE/SERVER

The article was to act as the initial communication and to define the starting word domain of each trial. The articles were taken from various online news sources and were selected on the premise that they would be interesting to a wide range of people while at the same time avoiding any personally contentious discussion i.e. religion. The purpose of this was to make sure that people were not discouraged from taking part in the experiment by articles that might seem contentious or unappealing to them. Participants were also able to reply directly to other participants' comments. The purpose of this was to capture direct engagements of one participant to another as opposed to participants just commenting which is akin to broadcasting information to the network of participants.

Figure 14 shows an example screen shot of participants' comments and replies to an article on the website. The slightly indented text box from 'User: 12' represents a reply submitted to the comment above from 'User: 1'.

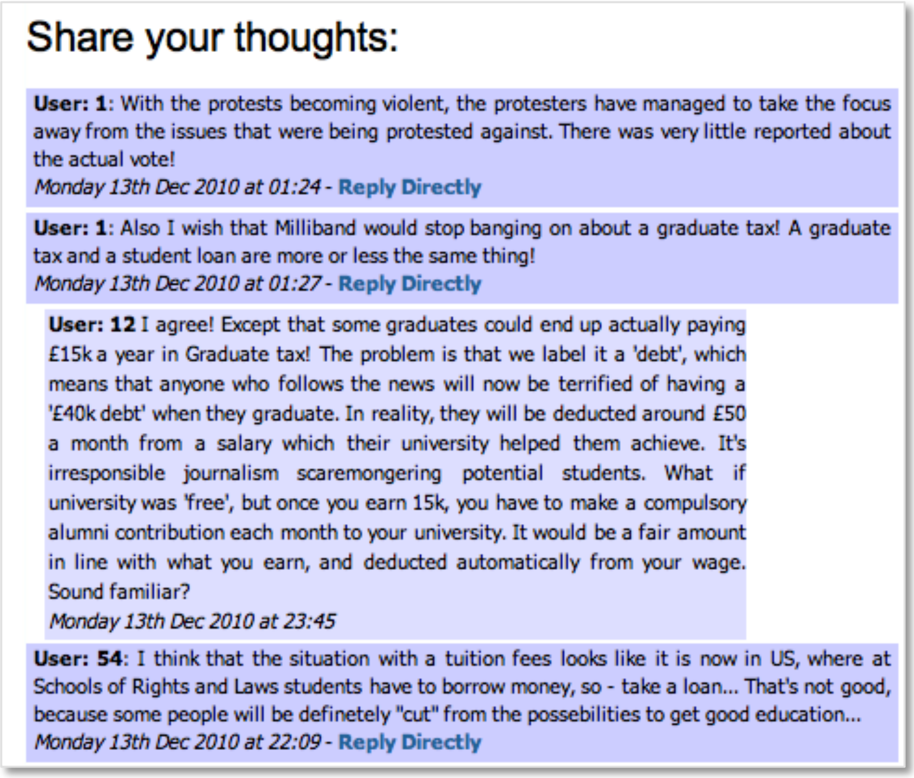

FIGURE 14 - EXAMPLE OF PARTICIPANTS COMMENTS AND REPLIES

3.2.1 Text Analysis: Extracting key terms

A method for identifying key terms in the text was developed that involved statistical analysis of content to find terms that occur with certain frequencies. The aim was to remove noise from the content of the website while maintaining terms that were characteristic of the content. Luhn (1958) proposes that such word frequencies within documents are a useful measurement of the word's significance relative to the document.

The text analysis processing was coded in PHP language and hosted on the same server as the website showing the articles and associated database; www.socialnetworknews.co.uk. The process by which the text analysis was carried out is detailed in the following steps:

Step 1: The first stage involves a search query language (SQL) query on the website's database to extract the desired content. The articles and respective comment threads are stored in interrelating database tables and are analysed separately.

Step 2: The text extracted from the database is in HTML format and all HTML language tags must be removed. The text then exists as a string stored in a variable. All punctuation is stripped from the string using a regular expression. Regular expressions are a sophisticated method of identifying patterns in a string. The following regular expression was used to match and remove any punctuation:

```
'/\s*[\s+\.|\?|,|!|\!|(|)|\-+|\'|\"|=|;|&#0215;|\$|\/|:|{|}]\s*/i'
```

The remaining string is passed through a built in PHP function that turns all upper case characters into lower case that leaves only the raw content string remaining.

Step 3: The whole string of content is then split into a storing array using the white space on either side of every word as delimiters. At this stage the complete text will be represented as an array equal in size to the number of words in the text.

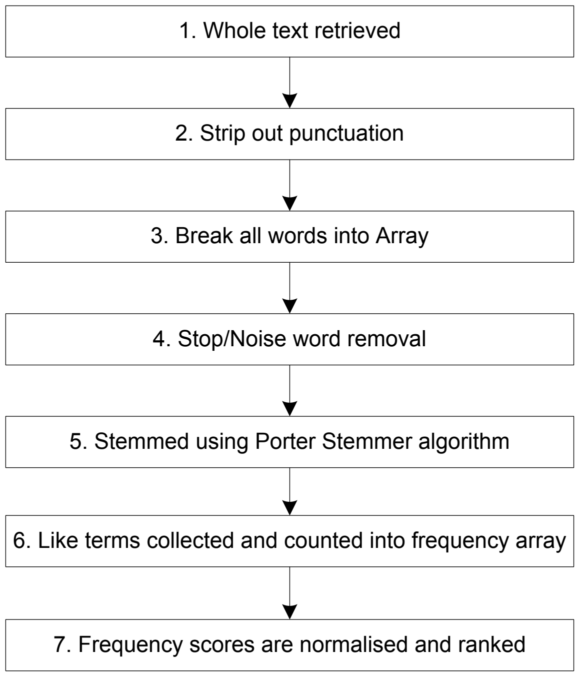

FIGURE 15 - TEXT ANALYSIS PROCESS

Step 4: The storage array is then compared to a list of stop or noise word. If a stop word is identified in the storage array it is removed. The stop words are a list of the most frequently occurring terms in the language. They can be obtained for English by extracting approximately the top 250 terms found in the British National Corpus (BNC) (Kilgarriff, 1998, BNC, 2007). Figure 16 shows the inversely-proportional nature of the frequency of terms found in a large corpus as noted by George Zipf (1935) and marked on top are the cut-off points that indicate the threshold in which between key terms can be found. Words that occur above the upper cut-off are too common in the English language and are not useful for this analysis as they are characteristic of most English written language. Words such as ‘and’, ‘of’, ‘or’ etc, occur with very high frequencies and do not help to define the domain of the content being analysed. Similarly, words that occur with low frequencies (below the lower cut-off point) are usually not useful in characterising a text. However, Zipf's observations were made on larger corpuses of text compared to the articles and comments used in this experiment. As such, it was decided not to set a lower cut-off frequency threshold for fear that terms which were characteristic of the text would be discarded. The upper cut-off was initially set to remove the 250 most frequently occurring words in the BNC and then altered manually through an iterative process of inspection of the storage arrays until all stop words were absent. This manual inspection was necessary since setting the upper cut-off too high would potential remove key terms from the storage array, which would result in different correlation values between contents. The final stop word list had 242 words.

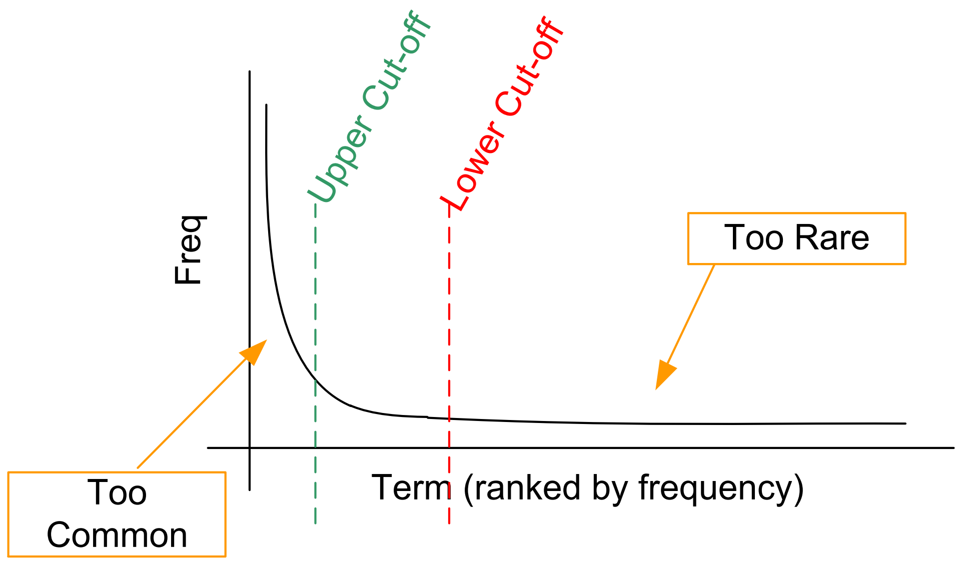

FIGURE 16 - ZIPF'S LAW AND THRESHOLD OF CHARACTERISTIC TERMS

Step 5: Once noise words have been removed from the holding array. The remaining words are stemmed. The point of stemming is to consolidate all the terms that have the same stem. For example the terms 'education' and 'educate' have the same stem, 'educat'. In this way we combine terms in a frequency count that inherently have the same meaning. The Porter Stemmer algorithm was used for the stemming process (Porter, 1980, Heyes, 2005). This was first developed by Martin Porter in 1980 and was subsequently refined over the following decades. It is now the defacto stemming algorithm used when stemming the English language. A PHP version of the stemmer was retrieved from Heyes (2005). The algorithmic steps of the stemmer are details in appendix 4.

Step 6: Duplicated stems that remain in the storage array are a summed and placed in a frequency array under a common index. The array is then indexed by the key terms and the stored values are their respective number of occurrences within the content.

Step 7: Frequency scores are normalised against the total number of terms in the array, which gives an individual weighting for each term relative to the content with noise removed. In effect this gives us a measure of that word's importance relative to the content of the article. For each article, frequency arrays for each participant's contributed content were created in conjunction with the frequency array for that given article. While the system could have been developed further to incorporate more sophisticated techniques for analysing the text i.e. a synonym library could have been added to bind terms with similar meanings, it was decided to keep it simplistic and light weight. This was decided because the system was only created as a means for measuring the alignment of participants. Also it was thought that by discounting more idiosyncratic language constructs that the system could be used on other Latin based languages given the appropriate stemming algorithm.

3.2.2 Mining Online Content

A second data collection method was also developed to use as a comparison to the Author's experimental platform. The motivation for this was to compare the behaviour of larger user groups obtained from data mined online with the smaller groups of participants in the Author's initial study. This method involved mining comments and reply threads from articles on the Guardian News Website. The Guardian website was chosen because it had a similar mechanism for commenting and replying to comments as the Author's system. Also due to its layout and functionality, it was simpler than other sites to expand and copy the entire comment thread from an article. In practice this was achieved by copying and pasting the entire comment thread from an online article and pasting it in its entirety into a text file (Figure 17). This text file was then parsed with a PHP script into individual comments and replies whilst also capturing user IDs. To illustrate, Figure 17 shows a sample of a text file containing the comments and replies from a given online article.

```
1  pollystyrene
2  23 February 2014 8:06pm
3
4  Recommend
5  20
6  You can't possibly be as pissed off as I was when I discovered they'd disco
7  Reply Report
8  Share this comment on Twitter
9  Share this comment on Facebook
10
11 Suleywoman pollystyrene
12 23 February 2014 9:02pm
13
14 Recommend
15 2
16 You can still get the old units at the moment, though you have to order the
17 Reply Report
18 Share this comment on Twitter
19 Share this comment on Facebook
```

FIGURE 17 - SAMPLE OF MINED COMMENT AND REPLY TEXT FILE

In the above sample, line 1 shows the initial comment that was posted by 'pollystyrene' and then line 11 marks the beginning of the first reply to this comment by the user name 'Suleywoman'. This data was parsed into the same database that was created for the author's experimental platform from which it could be accessed for analysis via the text analysis steps described above.

3.3 Task

Following University Ethics approval and guidance, participants who chose to take part had to sign up to the website to attain a username and password, agree that their information could be used for analysis, and then peruse the articles that were posted on the site at their leisure. After reading the article they were encouraged to leave a comment by the affordance of a submission box provided at the bottom of the page. If any comment thread already existed on a given article then that would be available for the user to read. In leaving a comment they were given two options; They could add a comment to the article thread. This would then be displayed in date order with the rest of the comment thread. They could also respond directly to another user's comment. This reply comment would then be displayed underneath the respective comment (as seen in Figure 14)

These two options were designed to give the participants the choices of either adding to the existing comment thread i.e. broadcasting to all participants, or actively engaging with another participant by replying to their comments. These options were included to gage the level of interaction of the participants i.e. a direct reply to a comment was scene as the direct engagement between participants. The participants could comment as many times as they liked and other participants were notified via email when the thread or their particular comment had been responded to.

3.4 Participants

Table 1 shows the number of participants that contributed content and how many interactions there were in total in relation to a given article. 49 participants signed up to the experimental platform but only 22 of these contributed usable content for analysis. This could be due to the lack of a specific goal in the instructions for this experiment; participants were asked to contribute their thoughts to any article they found engaging. They were not explicitly told to write something on each article. Perhaps it should not be unexpected that less than 50% of users actually contributed since reportedly lurking numbers can account for 90% of users in some online discussion groups depending on the domain (Nonnecke & Preece, 2000). Levels of lurking in this exercise might be attributed to the fact that the users' names appeared next to the comments that they posted, although admittedly there was no safeguard to prevent them signing-up with a fake name. The system had been purposely designed in this way to avoid any rude or vulgar comments that can occur on Internet discussion groups when people are allowed the shield of anonymity and distance (Wesch, 2008).

TABLE 1 - PARTICIPANTS AND INTERACTIONS BY ARTICLE; AUTHOR'S EXPERIMENTAL PLATFORM

Article ID	No. Participants	No. Interactions
2	2	3
3	5	11
4	2	2
5	5	6
6	3	3
7	8	9
8	4	5
9	5	6
10	1	1
11	6	7
12	2	2
13	5	5
14	12	20

3.5 Performance Measurement

3.5.1 Assessing Correlation & Group Membership

A means for quantifying the alignment between participants was derived. In this case, 'alignment' is a measure of the overlap between two participants' domains. A participant's domain is defined as the set of terms that are used by a given participant. For the purposes of this analysis, it was decided that Pearson's Product-Moment Correlation would be an appropriate test to apply. This test assumes a line of best fit between two sets of data (although the statistic reports the spread of data around this line rather than the slope of the line per se). While the data need to be on the interval or ratio scale, the test does not require data to use the same units of measurement or to distinguish between dependent or independent variables. This robustness means that it can be applied as an exploratory form of analysis without the need to make any assumptions about the distribution or type of data being compared. The correlations between each participant's contributions were assessed on two different bases. Firstly, the correlation between each participant was assessed in relation to the key terms that were introduced by the article. Secondly, the correlation between each participant was assessed in relation to the key terms that were introduced by the participants. These correlation calculations will show if the participants are aligning with the language introduced in the article or language that was introduced by other participants. Excel offers an appropriate function for assessing the Pearson's Product-Moment correlation between two sets of values (Equation 5).

$$Correl(X,Y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

EQUATION 5 - EXCEL CORRELATION FUNCTION

The process for calculating the correlations involved two stages. Firstly, two spreadsheets were created for each article; spreadsheet 1 for content introduced by the article and spreadsheet 2 for participant-introduced content. The stemmed key words (from the text analysis process) in each instance were listed vertically and participants who contributed to that article were listed horizontally. The cumulative score for each participant's usage of a given key term was added to the respective cell indexed by the participants ID and the key term. Figure 18 shows an example of one of these spread sheets. In this instance we can see that participant 15 has used the term 'fee' twice while participant 29 has used the same term three times. This spread sheet is an interim step in calculating the correlations. Each vertical line represents a term vector for key terms introduced by the article (spreadsheet 1) or the key terms introduced by the participants (spreadsheet 2). An issue that might be anticipated is that the participants contributed differing amounts of content to each article. As such it could be imagined that the correlation stats would be skewed. However the given correlation function accounts for this possible imbalance as the denominator of the equation (Equation 5) contains as a normalising function.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Key Terms:**	**Participant IDs**											
2		**1**	**12**	**15**	**29**	**54**	**60**	**61**	**62**	**64**	**65**	**67**	**68**
3	fee	0	0	2	3	1	2	1	0	5	2	2	1
4	univers	1	3	1	1	1	2	1	1	1	1	0	2
5	mp	0	0	0	0	0	0	0	0	0	0	0	0
6	govern	0	0	0	1	2	0	0	0	1	0	0	0
7	dem	1	0	0	0	0	0	0	0	0	0	1	0
8	lib	1	0	0	0	0	0	0	0	0	0	1	0
9	student	1	2	5	0	3	0	0	0	3	1	1	1
10	vote	1	0	0	0	0	0	0	0	0	0	1	0
11	39	0	0	0	0	1	0	0	0	0	0	0	0
12	fair	0	1	0	0	0	0	0	0	0	0	0	0
13	conserv	1	0	0	0	0	0	0	0	0	0	0	0
14	polici	0	0	0	0	0	0	2	0	0	0	0	0
15	cut	0	0	0	1	1	0	0	0	0	0	0	0
16	increas	0	0	1	0	0	0	0	0	1	1	0	0
17	against	1	0	0	0	0	0	0	0	0	1	0	0
18	tuition	0	0	1	0	1	2	0	0	2	1	0	1

FIGURE 18 - DATA SAMPLE FOR ASSESSING CORRELATION

The second stage calculates the correlation coefficients for all pairs of participants. The correlation function compares all the accumulative term-usage scores between two participants and evaluates a correlation coefficient. The correlation coefficients between participants in each instance were recorded in a matrix. A sample of the matrix can be seen in Figure 19. From this sample we can see that the participants' IDs populate the vertical and horizontal indices. A correlation coefficient of 0.3 or greater was assigned as a measure of alignment between participants. The correlation coefficient matrices for these articles were best visualised as network graphs since in latter trials, where comment streams from online news sources were mined, the results matrices became too large to fit on a single page (Appendix 1).

	1	12	15	29	54	60	61	62	64	65	67	68
1	0											
12	0.370	0										
15	0.228	0.313	0									
29	0.154	0.119	0.321	0								
54	0.394	0.158	0.519	0.364	0							
60	0.388	0.391	0.240	0.401	0.413	0						
61	-0.006	0.130	0.081	0.256	0.054	0.195	0					
62	0.242	0.236	0.161	0.207	0.426	0.199	0.101	0				
64	0.147	0.297	0.549	0.571	0.343	0.440	0.440	0.020	0			
65	0.270	0.251	0.469	0.516	0.313	0.351	0.351	0.361	0.546	0		
67	0.283	0.271	0.347	0.372	0.245	0.291	0.291	0.150	0.401	0.250	0	
68	0.482	0.243	0.429	0.285	0.638	0.350	0.350	0.407	0.179	0.371	0.245	0

FIGURE 19 - SAMPLE OF CORRELATION MATRIX

3.5.2 Additional Measures

In addition to correlations statistics, a number of other statistics were calculated for the groups of participants interacting with the article. The total number of interactions for each article was calculated and broken down into the constituting number of comments and replies. The ratio of replies to comments bears interest because it indicates the amount of discussion taking place. A reply to comments ratio greater than one suggests that there is at least one comment with two or more replies.

The average number of key terms used per participants was also calculated by calculating the total number of key terms used in all contributions to the comment thread divided by the number of contributors.

3.6 Results

Table 2 shows results from the author's experimental platform. Table 3 shows the results from data mined from online articles. The tables show participant numbers (N), interactions (Ints.); broken down into comments (Coms.) and replies (Reps.). The criterion for inclusion was a minimum of two or more participant contributions on an article. The table also shows the average number of key terms used per participant (Avg. KT/P), which gives an impression of the average length of each contribution posted to that article. This was calculated to see if the average amount of content submitted by each participant had any impact on the overall strength of alignment of the participants on each article. Standard deviation (SD) is reported along side AVG. KT/P. There was no maximum or minimum length limit for contributions and as such a high standard deviation shows that there was a wide disparity in contribution length between users on a given article.

The example results from this experiment are presented as a series of network graphs (Appendix 1) and show the correlation between participants who contributed to a particular article. The existence of an edge between two nodes in the networks signifies a correlation coefficient of 0.3 or greater. The weight of the edge between two nodes is proportional to the coefficient between two participants. Each node represents a participant who contributed to a given article and the label on each node is the anonymous user ID assigned to that particular participant. Articles are identified by an assigned ID, which matched the record id from their respective rows in the database table.

TABLE 2 - RESULTS FROM AUTHOR'S EXPERIMENTAL PLATFORM

Article ID	No. Alignment pairings (Article content)	No. Alignment pairings (Participant content)	N	Ints.	Coms	Reps.	Reps. / Coms.	Avg. KTs/P (SD)
3	1	0	5	11	8	3	0.38	40.0 (25.4)
5	4	0	5	6	6	0	0.00	20.4 (7.91)
7	10	0	8	9	9	0	0.00	16.8 (8.69)
9	1	0	5	6	5	1	0.20	26.0 (7.85)
11	3	0	6	7	5	2	0.40	14.0 (6.43)
14	32	3	12	20	15	5	0.33	69.7 (30.4)

TABLE 3 - RESULTS FROM MINED CONTENT

Article ID	No. Alignment pairings (Article content)	No. Alignment pairings (Participant content)	N	Ints.	Coms	Reps.	Reps. / Coms.	Avg. KTs/P (SD)
18	55	6	59	90	13	77	5.92	34.2 (47.1)
19	66	5	41	58	26	32	1.23	38.4 (43.8)
21	1168	71	144	257	98	159	1.62	46.6 (58.5)
22	120	9	59	100	41	59	1.44	50.3 (59.1)
23	501	27	122	206	59	147	2.49	45.4 (53.4)
24	241	21	109	186	59	127	2.15	46.6 (55.7)
25	143	30	81	109	64	45	0.70	27.2 (28.0)
26	54	4	50	73	37	36	0.97	30.5 (48.1)
27	715	52	103	192	75	117	1.56	63.4 (77.0)
28	315	33	98	180	75	105	1.40	62.9 (78.3)
29	829	79	115	217	63	154	2.44	52.7 (79.2)
30	102	14	68	102	51	51	1.00	14.2 (19.6)

Where the following data exhibited a trend between two variables, regression lines were drawn using the regression function in Excel. While formal curve fitting algorithms could have been used it was felt to be unnecessary for this work. The line model (linear or polynomial) with the highest value of R^2 was selected.

Figure 20 shows the relationship between the replies to comments ratio and the number of alignment pairings with regards to participant-introduced content. It shows a positive correlation ($R^2 = 0.49$) between the two variables. The results from article 18 were removed as an anomaly; this is explained in section 3.7.1.

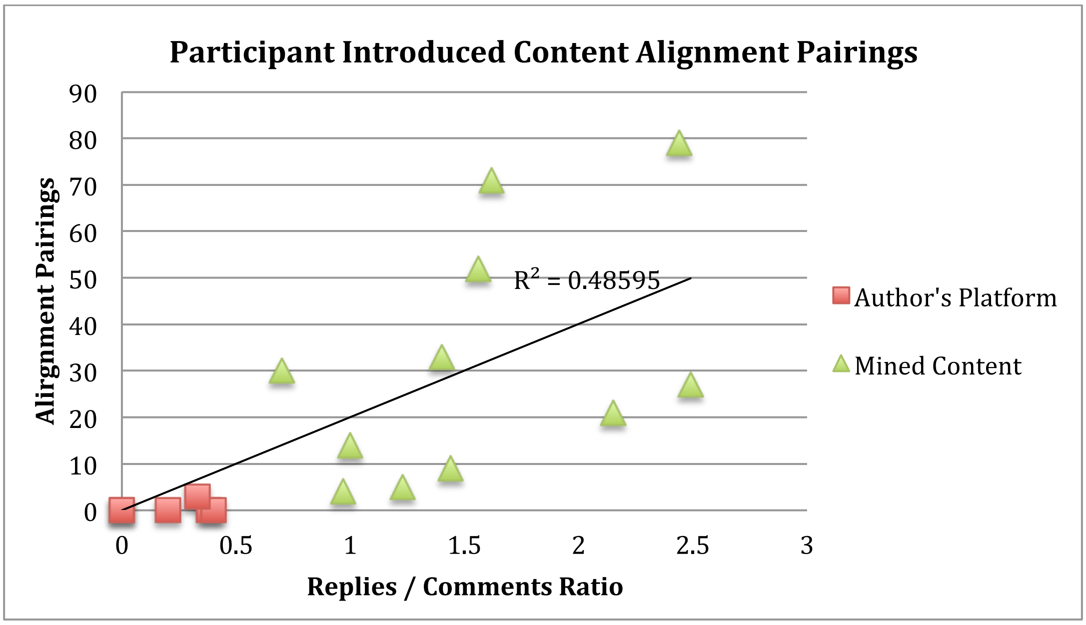

FIGURE 20 - PARTICIPANT INTRODUCED CONTENT ALIGNMENT PAIRINGS

Figure 21 shows the total number of interactions plotted against the replies to comments ratio for each article. The linear regression shows a positive correlation (R^2 = 0.795). The results from article 18 were removed as an anomaly.

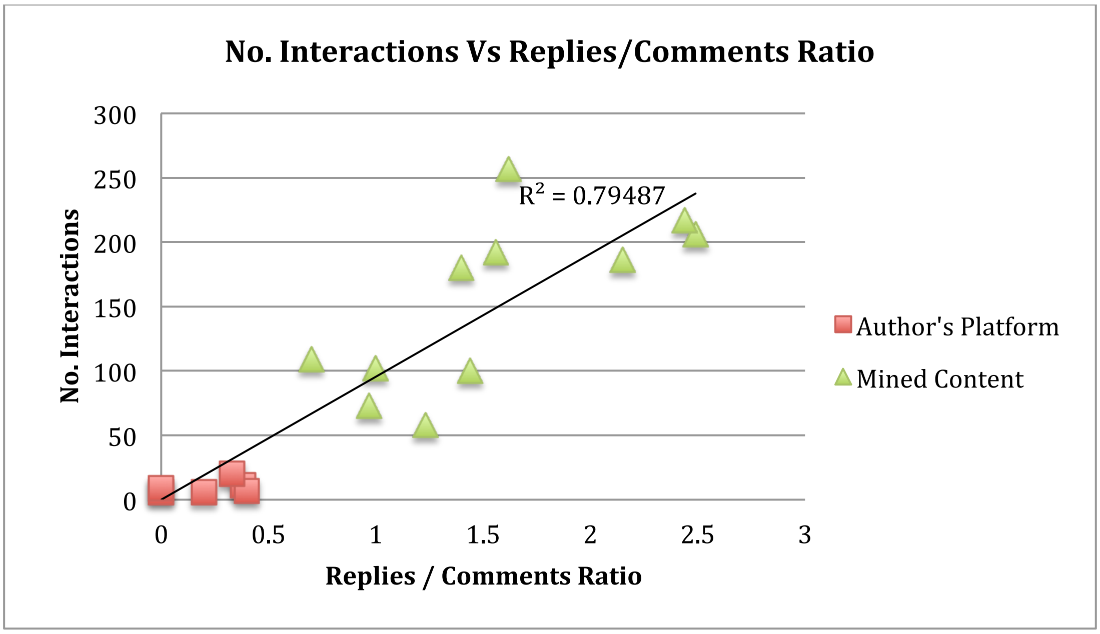

Figure 21 - No. Interactions Vs Replies / Comments Ratio

Figure 22 shows an exponential relationship between the number of alignment pairings for article content and the number of participants who contributed to the article.

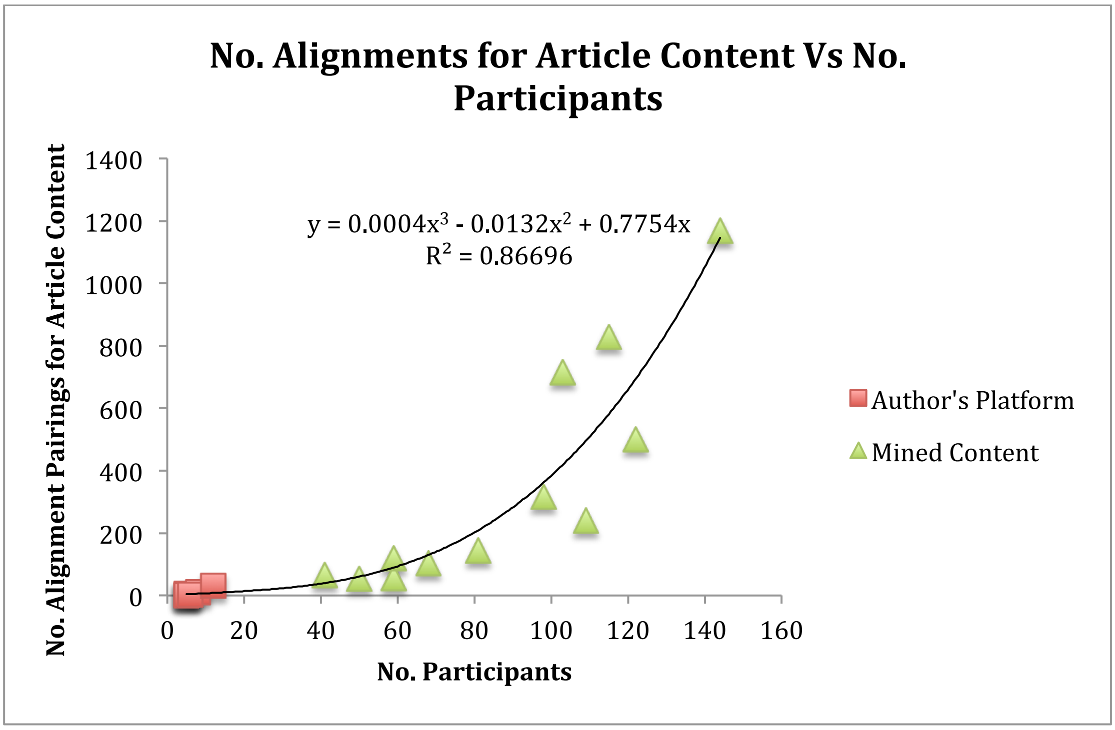

Figure 22 - No. Alignment Pairings Vs No. Participants for Article Content (R^2=0.87)

Figure 23 shows a weaker exponential relationship between the number of alignment pairings for participant content and the number of participants who contributed to the article.

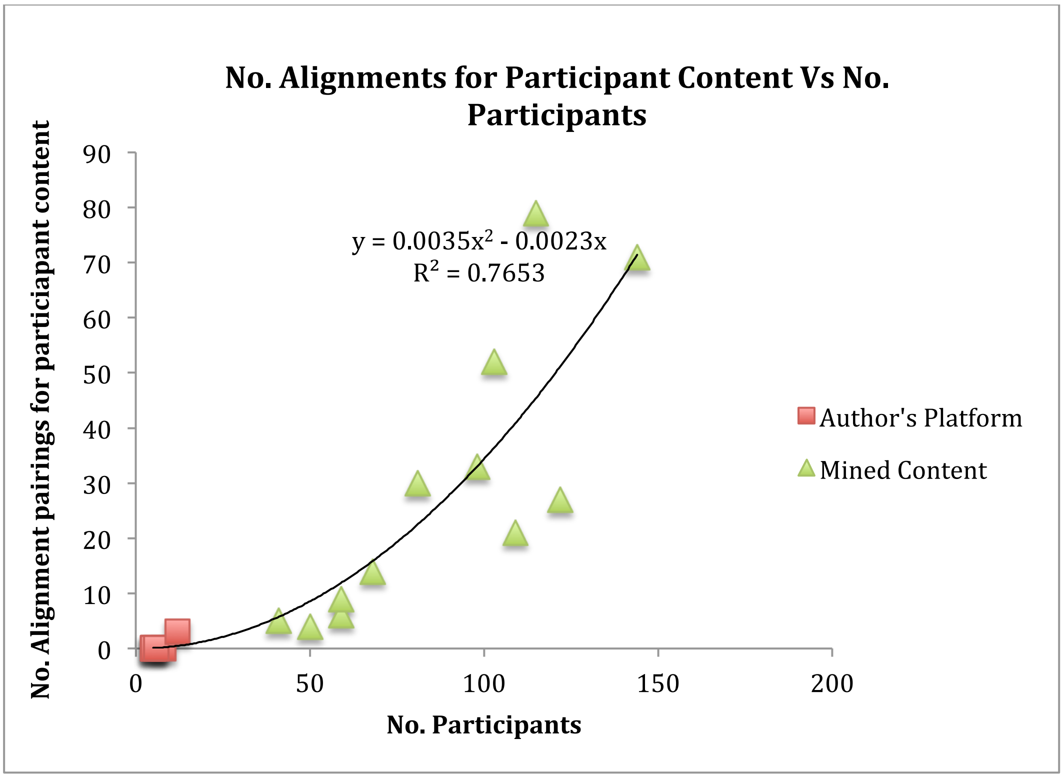

FIGURE 23 - NO. ALIGNMENT PAIRINGS VS. NO. PARTICIPANTS FOR PARTICIPANT CONTENT (R^2=0.77)

Table 4 shows approximations of the gradients of the upper and lower portions of the trend lines of shown in Figure 22 & Figure 23. The point is to show the reader that although the regression lines on the graphs look similar, the article content alignments is a lot larger.

TABLE 4 – AVERAGE GRADIENTS

Lower Portion	dy/dx
Article Content	1.60
Participant Content	0.22
Upper Portion	
Article Content	53.3
Participant Content	0.83

There is an interesting relationship between the number of participants who interact with an article and the average number of interactions per participant (Figure 24). The average number of interactions per participant trends upwards gradually as the number of participants per article increases.

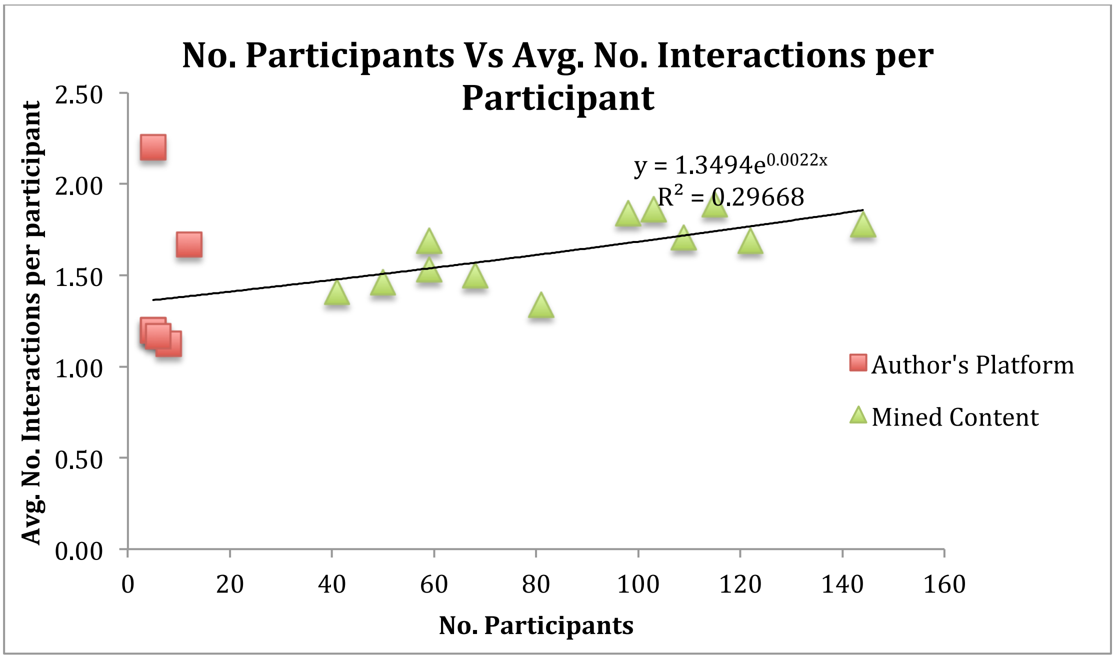

FIGURE 24 - NO PARTICIPANTS VS AVERAGE INTERACTION PER PARTICIPANT

3.7 Conclusions

3.7.1 Content Alignment

There is noticeably more alignment between participants around content introduced in the article than content introduced in the participants' comments. This shows that the participants are mainly using the key terms that were present in the original article in their comments and replies. This suggests that the participants are staying on topic while making comments and replies, but on the whole are not adopting new terms that other participants are introducing. From a sensemaking perspective, this shows that the participants were generally reliant on the content of the article to provide common ground for comment and discussion. There is an assumption that the article provides all participants with the same information and the ensuing discussions can be seen as an effort to ascertain a common sense of the article.

Where there is alignment around participant introduced content it would suggest that some participants were trying to create their own common ground and they were moving their conversation in a different direction. Figure 21 shows the total number of interactions against the replies to comments ratio of each article. It shows the probability of a new contribution being a reply rather than a comment is greater as the number of interactions increases. Figure 20 shows a positive linear regression between the replies to comment ratio on an article against the number of alignment pairings. This reveals that the participants are increasingly aligning their content with one another as the number of replies to comments increases. Both these findings in conjunction suggest that: Firstly, that on some articles with a larger amount of interaction, there are discussions or subgroups forming within the comment thread. This becomes more apparent where the replies to comments ratios are above one. Secondly, it suggests that within these subgroups the use of language is diverging from the article to some extent. Since the number of interactions increases with the number of participants, it is reasonable to suggest as the size of the groups of participants increases as do the social overheads of introducing new language and/or opinions. This might suggest why participants tend to accept and use terms introduced in the article. Where there are low amounts of alignment this also correlates with low replies to comments ratio. This means that people are just commenting without replying to each other and hence in these instances there is no discussion developing and as such no collaborative efforts.

3.7.2 Other Measures

The main difference between the author's experimental platform and the data gained from mining online documents is the disparity in the number of participants interacting with the article. There is a question of what motivates people to engage with the article. In writing a comment on the article, you are fundamentally broadcasting your message. However the commenter is unaware of whom he or she is broadcasting to and to how many people. It can be assumed that when people were commenting on the online news site that they were more likely to be broadcasting to larger audiences in comparison to people commenting on the author's experimental platform. Although when participants were commenting on the author's experimental platform they could not be confident they were broadcasting to a large group of Internet users. As such, the motivation for commenting can be considered the same for both sets of commenters.

This means that the greater the number of participants on an article, the more likely a participant will engage on the comment thread on that article. Establishing an explanation for this correlation is difficult. It could be that an article is on a topic that is of particular interest to certain readers. It could also be the case that the initial comments, which we can view as a means of broadcasting, are engaging other readers to the extent that they post a reply. If the latter were true, we would expect to see a higher ratio of replies to comments in those articles. Of course, both cases could also be true simultaneously. In relation to sensemaking, one contention is that for large groups of people there is a requirement for more interaction to take place for a shared understanding to develop. This is not a far-fetched idea since we can imagine a process whereby comments act as new ideas being added to the content by opinion leaders. The replies to these comments act as a means for people to question, add to or agree. We don't see lots of repetition in the content because people read the previous comments first and only engage and add content if they find discrepancies or provocation in the existing content. In this `scenario, the hypothesis is that the relationship between the number of participants and the average number of interactions per participant is a reflection of the participants' efforts to organise and define the domain of the topic area. Thus in relation to sensemaking the suggestion is that the ratio of replies to comments indicates a discussion taking place. The higher ratio of replies to comments the more participants are engaging with one another. This is justified by aspects of the literature review that show sensemaking as a discrete inherent process that all organisms do when facing unknowns (section 2.3).

An exponential relationship exists between the number of alignment pairings (correlations between participants) and the number of participants in a given trial when looking at article introduced content (Figure 22). The number of alignments increases as the number of participants increase. In more simple terms, the exponential trend line is showing that there is a fairly dramatic increase in the use of terms from the article by each participant as the number of participants in the study increases. This is expected since there are more people with which to have correlation. The fact that the number of pairings increases exponentially suggests that there is a heavier reliance on provided information for larger groups. This is plausible because there is a need in larger groups to establish some common ground on which to base their discussions. These findings suggest that sensemaking, in the semantic sense, becomes more difficult as group size increases. The groups are relying on the common ground presented to them in the article; they mimic the terms used in the article to establish a common language for discussion. The novelty of these finding is that they were found directly from the analysis of collected data as opposed to through direct observation of the social dynamics of the group. Through some semantic analysis between participants we can see how they are prioritizing the use of certain terms from the article content. The author's conjecture is that the participants are doing this to create/maintain group cohesiveness.

Figure 23 shows the same relationship for content that was introduced by the participants. Once again there is an exponential relationship but when this is compared to the relationship found in Figure 22 we see it is considerably weaker. Approximations of the gradients of the upper and lower portions of both trend lines were taken from each graph (Table 4). There is a dramatic increase in the number of pairings per additional participant with respect to article-introduced content and a lesser rise for participant-introduced content. This supports the conjecture that as the number of participants in the trial increases, they are not as reliant on the terms introduced by other users in forming their discussions. In contrast this suggests that smaller groups seek to create their own common ground. It seems there is a cost for introducing new content, which increases as group size increases. Smaller groups approach sensemaking by deciding between themselves which terms to use and will build their own common ground. They are developing common ground as a sensemaking activity because there is a low social cost to introducing new terms or ideas. Larger groups seem to accept common ground presented to them and any new terms become diluted by the presiding terms introduced by the article. What this tells us is that common ground as a facet of sensemaking has differing functions dependant on group size. If the group size becomes too large there is a battle against the Common Knowledge effect; the group will reject new terms to maintain the cohesiveness of the group. Relating to the emergency management scenario that inspired this chapter, it can be imagined that when you broadcast information to many people there is a cost involved for those people to discuss it. Within these larger groups, cliques will tend to form, like the replies thread below certain comments. As such the suggestions is that sensemaking is done collaboratively in smaller groups where there is a low cost in introducing terms and ideas, and it is easier for the group to discuss a range of things. The next question is whether technology can support this process.

3.7.3 Further Thoughts on Semantic Sensemaking

In this experiment the primary concern was semantic sensemaking. Although, it should be affirmed that meaning is not being measured directly, the emergence of a mechanism by which meaning can be discoursed and shared is measured. i.e. the alignment of article key-term (theme) usage by the participants shows consensus on the language to be used in discussion. The adoption of key terms demonstrates the participants' ability to filter the terms that are considered by the group to be characteristic of the article. Content analysis suggests that with these key terms adopted by the participants, the approximate meaning or 'gist' of the article could be reconstructed. Thus, it could be suggested that semantic sensemaking in this experiment is not necessarily to do with the formation of common ground but is actually about the collaborative efforts of the participants to recognise the key terms that represent the 'gist' of the article. If this is the case, then it makes sense to ask; how the manner in which groups share information can affect the production of 'gist'? As such, in the following chapters the objective will shift to developing ways to study pragmatic sensemaking and looking at sensemaking behaviours in goal driven experiments and under organisational constraints. The emphasis will be on smaller groups of people (approx. $N<30$) since, as this chapter suggests, large groups of people have high overheads when forming discussion which in turn inhibits sensemaking capabilities.

4 ELICIT Study

This chapter presents the results from analysis of experimental findings of the effect of group structure on the ability for a given team to complete an intelligence analysis task. It was found that in general having a flat cross-organisational structure promotes better sensemaking capabilities for a team and that hierarchical command structures inhibit sensemaking processes.

4.1 Introduction:

This chapter starts with an overview of the experimental protocol used to collect the data analysed, as it is beneficial to have an understanding of the experiment used in this chapter before embarking on a discussion about the aims and objects of analysing the experimental data.

4.1.1 Experiment Overview

The Experimental Laboratory for Investigating Collaboration Information-sharing and Trust (ELICIT)[3] is an experimental platform developed by the United States Department of Defense Command and Control Research Program of the Office of the Assistant Secretary of Defense for Networks and Information Integration (OASD/NII). It was developed for running studies into the capabilities of different command and control structures with a view to moving from traditional hierarchy-based command and control practices to more agile command structures with less centralised power and decision rights (Alberts & Hayes, 2006). The platform provides a software environment for participants to take part in an intelligence analysis task. The task in question involved groups of participants, organised in a pre-ordained command structures, sharing information snippets (factoids) in an attempt to collaboratively find the Who, What, Where and When of an anticipated terrorist attack. No single person starts with all the factoids and in order for a successful conclusion to be reached, players must share with one another in a timely manner. An individual ELICIT trial involves a group of at least 17 people, each situated at their own networked computer, operating in one of the two possible conditions; a hierarchical network or an edge network. In the hierarchical condition each of the participants are assigned a role to play within a hierarchical organisation before the trial begins. They are assigned roles as either a team member, a team leader or a cross-team coordinator. There were 4 possible teams;

[3] http://www.dodccrp.org/html4/elicit.html

Who, What, Where and When, each with 3 or 4 team members who communicate with a team leader and in turn these team leaders communicate between one another via the cross-team coordinator. This condition is shown in Figure 25. Additionally there is a set of central websites acting as information repositories that participants may pull and post factoids to. It should be noted that the communication constraints of this experimental condition were organisational and not technical, i.e. members could still send information to anyone else within the organisation. The edge network condition consisted of all members in a flat and equal organisational structure. This is shown in Figure 26.

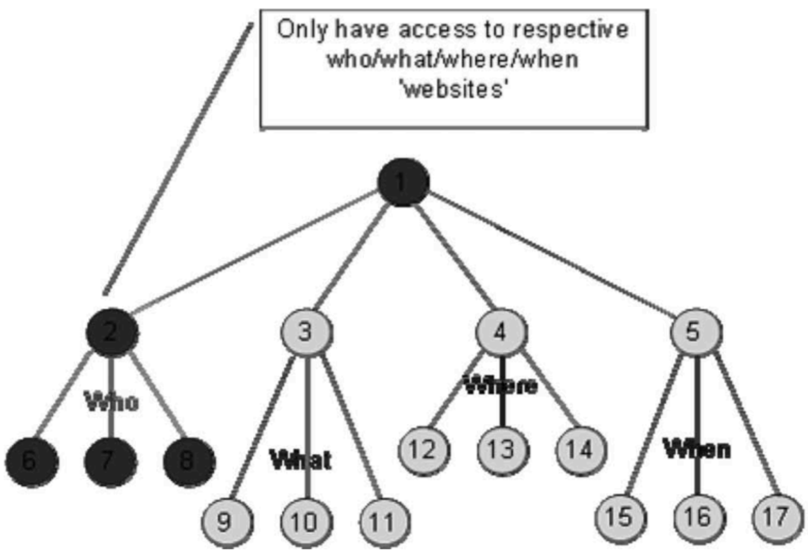

FIGURE 25 - CONFIGURATION OF HIERARCHICAL NETWORK CONDITION IN THE ELICIT STUDY (STANTON ET AL, 2012)

There is no mechanism in place for the participants to freely chat during a trial. At the beginning of the trial 2 factoids are given to each participant in the network. Every 5 minutes from the start of a trial, an additional factoid is distributed to each member of the network. The factoids and their distribution are designed in such a way that no single person has all the information necessary to complete the task and thus sharing must take place.

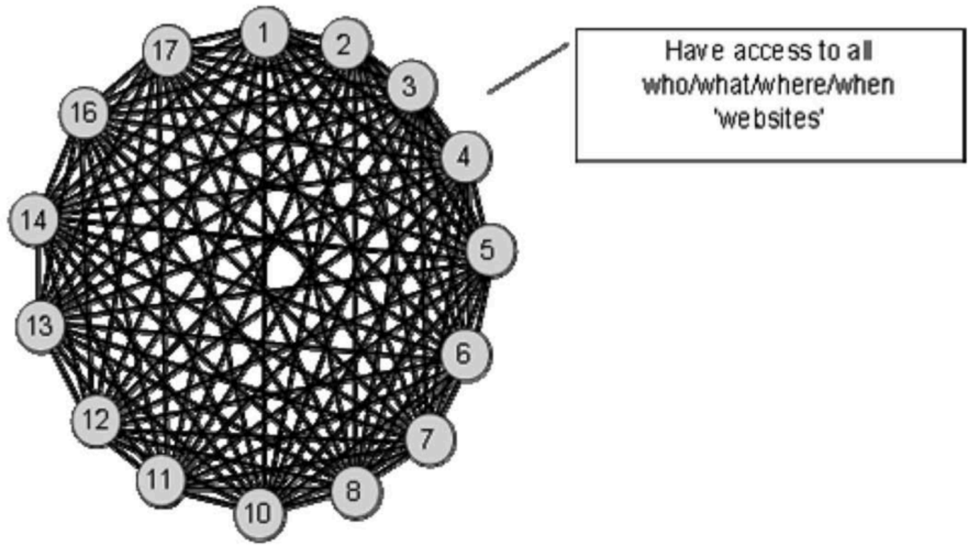

FIGURE 26 - CONFIGURATION OF EDGE NETWORK CONDITION IN THE ELICIT STUDY (STANTON ET AL, 2012)

4.1.2 Discussion

The primary focus of chapter 3 was Semantic sensemaking; measuring how participants aligned their use of content with one another in an effort to form common ground for discussion. The formation of common domains is part of the process for establishing shared meaning. The previous chapter acknowledges the constraints that social dynamics have on performance in terms of common knowledge affect and the effort required in terms of information sharing to overcome these constraints. In this chapter's experimental work, the participants are sharing fixed knowledge tokens (factoids) and there is no allowance for free text chat with other participants. As such, the experimental platform is inhibiting an activity that allows collective semantic-sensemaking to take place. This allows us to explore the possibilities that pragmatic sensemaking is occurring via some sort of group coordination i.e. the unstructured groups start behaving differently in the way they coordinate their communications as they make sense of the task at hand. Since the hierarchical teams have a fixed organisation it will be challenging to see any change in the way they are coordinating themselves to achieve the task. Thus, it is primarily with the teams organised into edge networks that observable pragmatic coordination should occur. It is also expected that the edge network groups should perform differently at this task because they are operating as *networks of exploration.* From the literature on this concept (section 2.6), the higher levels of connectivity in these types of groups *should* lead to better information sharing and decision-making. Describing better performance is dependant on the performance measure that is used and relates back to the initial research question; can metrics be assigned for measuring sensemaking? The performance metric used for this experiment is detailed later in this chapter (section 4.3).

In chapter 3 it was observed that smaller groups of people do semantic sensemaking more efficiently than larger groups based on the notion that larger groups tend to have more overheads in terms of establishing common ground and introducing new concepts to the group. This seems to be true, at least in a non-goal driven exercise. As such, moving forward, this thesis will be concerned with sensemaking in *smaller* groups. Using the analysis from the last chapter, an approximation can be made that small groups contain up to about 12 people and a large group has more than 40 people. For the purposes of this thesis, a medium size group can be considered to be between these two ranges; approximately 12 – 40 people. Interestingly these numbers are similar to army unit sizes where there are squads or sections, which operate as 4 to 12 people, platoons that operate as 16 or more, and companies that will range from 60-200 people [4],[5]. These are all approximate sizes and will vary for non-infantry units but it can be imagined that these group sizes have evolved over time in recognition that there is a natural number of people that can be commanded and/or work together on a given task. From these findings and observations it would appear that approximately 12 people is a natural group size that can be commanded by a commander/controller. This reflects findings by Dunbar (1998), who acknowledges in his work on naturally occurring sizes of groups of humans, that groups size clusters tightly around the values {5, 12, 35, 150, 500 and 2000}. Although this group size works well for organisational command and control it is a good approximation for a group required to do collective sensemaking. The previous chapter's results suggest that larger groups ($N >\approx 40$) struggle with collaborative semantic sensemaking because there is too much effort involved in attaining common ground. At the opposite end of this spectrum it might be possible that a group has too few members to be an adequate network of exploration. That is to say that there might not be enough people in the network to generate plausible explanations or challenge existing common knowledge effects.

In the hierarchical condition in this experiment there are 4 teams separated by function (figure 1). Each team has a team leader and there is also a cross-team coordinator node that oversees these team leaders. Reviewing the two network types, it can be seen that lines of communication in the hierarchical structure have the potential to act as

[4] http://www.army.mil/info/organization/unitsandcommands/oud/

[5] http://www.secondworldwar.co.uk/index.php/army-sizes-a-ranks/86-army-units-a-sizes

information bottlenecks that could slow the speed of salient information diffusing through the network and perhaps inhibiting sensemaking. Also as data traffic becomes higher, there will be increasing strain on the top node in the network in terms of its ability to process and pass on information. In the edge network the ability for all nodes to communicate directly with all the other nodes removes these information bottlenecks but at the risk that any individual node could become overloaded; especially if the group size is too large. It also means that each node in the network is responsible for acting as a filter for salient information dispersal. The temptation for members in this type of organisation might be to broadcast to all nodes; a behaviour seen during the pilot study of chapter 5. In doing this they are drowning out useful information in noise and also losing focus on task work in favour of teamwork. So the goal of members in this organisation is to associate themselves with other members who have useful information for them and to only pass on, what they believe to be, salient information to other group members.

Here we can begin to see the strengths and weaknesses in terms of sensemaking capabilities of both network types. The hierarchical network, because of its strict communication protocols, may struggle to efficiently disperse information through the system, however it has no overheads in terms of establishing *where* it needs to send the information. The pragmatic workings of the network have already been established and the responsibility of sensemaking befalls on the individual network node to categorise the factoid and forward it to the appropriate department (i.e. who, what, when, where). This is akin to a sort of transactive memory network (Wegner, 1986). Transactive memory is a mechanism by which groups encode, retrieve and store knowledge artefacts (section 2.7). The hypothesis is that an individual does not store all the knowledge of the group but instead maintains a meta-memory of who knows what, or what expertise they possess, within the group. In this instance the structure dictates the location of the expert group. i.e. Factoids containing location information should be sent to the 'Where' subgroup in the command structure. Contrastingly, for the edge network to perform efficiently and successfully, the challenge for the network is to realise its pragmatic workings. In one scenario this might involve each node maintaining a record of factoids possessed by the other nodes in the network to ensure information sending isn't duplicated and to avoid the dilution of salient information in noise.

With the above in mind, the expectation might be for the hierarchical groups to perform more efficiently during this experiment. The task in the ELICIT study can be seen as a game like CLUEDO where one is trying to collect the who, what, where and when artefacts of a murder. However, ELICIT is more nuanced than CLUEDO because the factoids given in ELICIT are more detailed:

- Factoid 1: The attack will be at 11:00.
- Factoid 2: The Violet group is planning something big on the 5th.
- Factoid 3: There is a lot of activity involving the Violet group.
- Factoid 4: The Jackal has been seen in Tauland

Factoid 1 contains purely 'when' information but it is incomplete, as it has no date. Factoid 2 contains 'who' and 'when' information but the when date is still short of a relative month. Factoid 3 is purely 'who' information. The difficulty is seen that factoid 2, as a fixed unit of information, could be sent to the 'who' and the 'when' branches within the hierarchical structure. This complicates the pragmatics of the network and places more emphasis on individual semantic sensemaking to decide if a factoid contains enough information relevant to one or more departments.

So actually, the goal becomes harder for the hierarchical group because of the compartmentalisation of the command structure and the nature of the information with which it is working. In the hierarchical structure – at least in the real world - it would be anticipated that people in the position of team leader or the command node would have sufficient expertise and experience to recognise relevant information and pass it to the appropriate team leader.

In a real-world context, ELICIT can be considered a simplification of the emergency response structure where the functions would be fire, police and ambulance. Using this comparison, the team members might be bronze level (operational) commanders responding at the scene of an incident and the team leaders could represent silver level (tactical) control room commanders. The hierarchical command system between control room and responder works when an individual branch within the organisation is dealing with something that is routine or at least not uncommon. When the response organisation as a whole is faced with a scenario that is outside of its comfort zone, where events are occurring at a high tempo, demand for resources is high, information is unclear and the environment is changing rapidly, its ability to make sense of the scenario will diminish. As the silver level command becomes bogged down with issues of interoperability and forming a common operating picture, the established structure of the organisation leaves team members (bronze level) isolated and delays response time.

To summarise, in a changing environment the ability of an organisation to do semantic sensemaking is important and the process by which this take place is impacted by the pragmatic sensemaking of the group. An existing well-defined organisation might struggle with problems of semantics and equally a new poorly-defined organisation may struggle to carry out a routine task as its efforts are spent establishing whom to send information to with less emphasis on the content of that communication. An Edge organisation (as in this study) may struggle initially with the task as they establish the pragmatic workings of the organisation. In larger edge groups this problem is exacerbated and it is likely that the formation of cliques and communities of practice will occur before a coherent collective response takes place. Roles will evolve in this flat organisation dependent on the situation being faced. In comparison the clearly defined roles in a hierarchical structure make it more efficient at delegating activity.

4.2 Experimental Process

The results for this chapter were obtained from a publically available data set from experiments that were carried on the ELICIT platform in various locations around the world. The experiment is an investigation into the differences in extremes of the NATO Approach Space, i.e., a traditional hierarchical command and control structure and an edge organization in which, everyone can communicate with everyone in the network. These are illustrated by figure Figure 25 and Figure 26 respectively. The study had a total of 544 participants from five countries; Portugal, Singapore, Canada, United Kingdom and United States (Stanton et al, 2012). The author of this thesis did not collect the experimental data used in this chapter. This has two implications. First, the data are being analysed in ways that are beyond the original aims of the people who collected them. This has meant that there might be situations in which the analysis has made assumptions on the data in order to interpret them. Second, while the instructions for ELICIT experiments were detailed in the preceding section, it cannot be guaranteed that all versions of the experiments have been run with equal rigour, and so there might be some noise or distortion in the data as a result of this.

4.3 Performance Criteria

The ELICIT trial is played with one of three sets of factoids. For each factoid set there is a 'correct' answer. This answer is the time, date, location and persons responsible for an upcoming attack. For example, one solution reads:

"The Violet group plans to attack a financial institution in Psiland on April 5 at 11:00 AM"

At regular points in the experiment, each member of the group submits a hypothesis, informed by the factoids that they have obtained up to that point in the experiment. For the purposes of the author's analysis, the 'correct' answer represents a ground truth against which these hypotheses can be assessed. Previous studies using ELICIT have created performance metrics for the 'Correctness' of a given hypothesis. Chan et al (2011) break a submitted hypothesis into 4 constituent components: Who, What, Where and When. The first three constituent parts of the hypothesis are given a score of either 0 or 1 (correct or not correct) and 'When' is given a score of either {0.25, 0.5, 0.75, 1.0} which allows for partial correctness. The performance score for a hypothesis is then given by:

$$C = 0.25\ (\ \mathtt{WHO}\ +\ \mathtt{WHERE}\ +\ \mathtt{WHAT}\ +\ \mathtt{WHEN}\ ).$$

This metric is considered inadequate since it does not account for *noise* in the participant's hypothesis. 'Noise' (discussed in section2.8.1) in this experiment is considered any term submitted in a hypothesis that is not part of the trial solution. As such to obtain maximum accuracy only one response for each of 'who', 'what', 'where' and 'when' should be submitted. To illustrate, a hypothesis submitted during one of the trials is given below:

'The Lion' will attack the 'coalition member embassy, visiting dignitary or financial institution ' in 'Tau, Epsilon, Chi, Psi or Omega-lands' on the 'April 10'

Considering the above hypothesis, the participant has submitted multiple options for various components of the solution. From the sensemaking perspective of the Pirolli and Card(2005) model (section 2.2), generating a hypothesis with multiple options for each component would require a further re-evaluation of evidence or another iteration of information searching which could be costly. As such from the perspective of the evaluator, these additional guesses must be penalised in the performance condition. In Chan et al's (2011) method there is no way of representing that the hypothesis has been 'diluted' by additional terms. Stanton et al (2012) also present performance results, reporting; "the Edge organisation correctly identified the solution, on average, on more occasions (X = 2. 38 times / mean rank 10.69) than the C2 organisation .(X=0.62 times / mean rank 6: 31)". However, the paper fails to elaborate on how these results were calculated.

Arguably, the Chan et al's (2011) method of correctness could have been altered to divide each facet of an attack hypothesis ('who', 'where' 'what', 'when') by the number of each guessed respectively. i.e. if the participant had guessed two possible locations for the attack and one had been correct, the score for 'where' would have been ½ or 0.5. This might have made it possible to compare Chan et al's (2011) method with the new adjusted method. Two problems exist with this approach. Firstly, it does not offer any more insight into the behaviour of the groups as the comparison would be between two invented performance scores. Secondly, each hypothesis would have to be manually assessed under both measures to extract the individual facets, which given the amount of data would not have been feasible. As such, it was decided to develop a bespoke metric for performance in this task that would incorporate noise as a factor. Initially it was thought that the relationship between ground truth and hypotheses can be considered in terms of signal detection theory where:

- Hit = Term used in hypothesis that matches solution.
- Correct Rejection = Term not used in hypothesis and not in solution.
- Miss = Term not present in hypothesis but should be.
- False Alarm = Term used in hypothesis and not in solution.

This would then allow sensitivity (section 2.9.5) to be used as a metric for individual performance. However, it was decided that sensitivity as a measurement did not fit the scenario well. Firstly it was felt that performance should be expressed as a percentage because the ambition is to see to what extent a participant's hypothesis matches the trial solution taking considerations for noise in the hypothesis. Secondly, the miss rate would always be the same as the false alarm rate because any term used in the hypothesis that was not part of the trial solution could be considered as a miss; by virtue that is wasn't a hit, and a false alarm. Finally, correct rejections could only be calculated because the false alarm rate and correct rejection rate should add up to one. However, given the previous point, this would mean that correct rejection rate and hit rate would be equivalent.

The performance metric developed needed to take into account hit-rate and noise. Hit rate in this instance is measured by the number of correctly identified terms in a hypothesis as a ratio of the total number of terms in the solution. The performance measure must also take into account any noise in the submitted hypotheses. Terms submitted in a hypothesis that are not part of the solution represent 'false alarms' and dilute the hypothesis. A noise coefficient is created by calculating a ratio of the correctly identified terms (Hits) as a fraction of the total number of terms in the submitted hypothesis. By multiplying the hit rate by the noise coefficient it will give a performance score:

$$H = \left({}^{h}/_{T_s}\right)$$

$$N = \left({}^{h}/_{T_{hyp}}\right)$$

$$P = H \times N$$

$$P = \left({}^{h}/_{T_s}\right) \times \left({}^{h}/_{T_{hyp}}\right)$$

$$\boldsymbol{P} = \frac{\boldsymbol{h^2}}{(\boldsymbol{T_s} \times \boldsymbol{T_{hyp}})}$$

Where:

P = Performance.

h = number of hits.

H = hit rate.

T_s = All terms in the solution.

T_{hyp} = All terms in the hypothesis.

N = Noise Coefficient.

EQUATION 6 - ELICIT PERFORMANCE MEASURE

For practical purposes, while calculating the performance measure only the key terms in the solutions and hypotheses were used. As in the previous chapter, this was achieved by stop/noise word removal. An example is presented in Figure 27.

Full Trial Solution:

*"The **Violet** group plans to attack a **financial institution** in **Psiland** on **April 5** at **11**:00 **AM**."*

Trial solution Key Terms: **violet** - **financial** - **institution** - **psiland** - **april** - **5** - **11** - **am**

Sample hypothesis 1:

*"**VIOLET** will attack the **embassy** in **epsilonland** on the **APRIL** **10** at **11** 00 **AM**."*

Key Terms: **violet** - **embassy** - **epsilonland** - **april** - **10** - **11** – **am**

Performance = (4 / 8) x (4 / 7) = 28.67%

Sample hypothesis 2:

*"**VIOLET** group will attack the **FINANCIAL INSTITUTION** in **omegaland** on the **APRIL 5** at **11** 00 **AM**"*

Key Terms: **violet** - **financial** - **institution** - **omegaland** - **april** - **5** - **11** - **am**

Performance = (7 / 8) x (7 / 8) = 76.66%

FIGURE 27 - EXAMPLE PERFORMANCE SCORES

In the above samples the key terms are in bold. The correctly identified key terms are highlighted in red. It can be seen in sample hypothesis 1 the performance score of

28.67% is calculated from the hit rate as the ratio of correctly identified key terms and the total number of key terms in the solution; (4/8). The noise coefficient is the ratio of hits and the total number of key terms in the hypothesis; (4/7). The performance measure is expressed as a percentage where 100% indicates a submitted hypothesis that contained all the correct key terms and a noise coefficient of 1.

Since the focus of this thesis is with sensemaking as a collaborative process, the concern is assessing the group's performance as a whole. Therefore, the median performance score of each group is assessed as each member of the group submits a hypothesis. The median is used as the summary statistic as the performance scores are not normally distributed within groups and it avoids misleading mean scores that can be skewed by a group with a few good performers. In relation to sensemaking, a group's current collective understanding should be dependent on all group members. If a group has a big disparity between a few high performers and the remaining members, a mean summary would suggest that the group is collectively performing better than it really is. This allows the collective correctness of the group to be observed as a trial progresses. It is a bespoke measurement for this experiment, which offers us a valuable means of exploring the ways teams of people collaborate to make sense of a problem space in which they need to gather pieces of information and then decide the most appropriate way to combine these into a meaningful answer. The performance measure is not measuring sensemaking per se but does provide an indication of how successful any measurable behaviours of the group were to producing a good result. The analysis reported considers the performance of groups as they work towards their maximal performance and to what extent communication links are utilised to pass information (factoids) between individuals in the network. One might, for example, expect the hierarchical groups to demonstrate highly directed factoid sharing, which might involve relatively low exchange of information reflected in relatively low density network. Alternatively the groups arranged in an edge structure may have high levels of undirected sharing where factoids are broadcast to many members of the organisation.

4.3.1 Other Measures

Additional indicators that will give an insight into group behaviour are the number of factoids that are pulled and posted to and from the websites in the experiment. Interaction with the websites in the experiment shows individual efforts to search for information to achieve the task. A high level of factoid sharing between participants, conversely, shows more focus on teamwork. While both groups will display website interaction and sharing, it is anticipated that the edge groups will show more sharing since they have no organisational constraints on doing so.

4.4 Results

4.4.1 Performance

Table 5 shows the median final performance scores of all groups at the end of each experiment. The general observable trend is that the edge networks have a higher performance score than the hierarchical networks. This is true in all groups, except for the two trials that took place in Canada. No group's median performance reaches 100% meaning that there isn't unified understanding of the trial solution within the group.

TABLE 5. - MEDIAN PERFORMANCE SCORE OF EACH GROUP.

	Performance (%)	
Group Name	**Edge**	**Hier.**
Boston	62.3	7.24
Canada	25.1	44.7
Cranfield[6]	64.4	G1: 7.24
		G2: 6.35
NPS	76.7	31.1
Singapore	70.1	37.6
Southampton	64.4	31.1
Westpoint	76.7	28.7

4.4.2 Bass Curves

On plotting the median performance of each group against time, 'S' shape trends were observable. It was decided to make these trends more obvious by adding smoothed Bass trend lines to the existing performance plots. This is illustrated in Figure 28.

The Bass trend lines were calculated via the given formula:

$$N(t) = N_{(t-1)} + p(m - N_{(t-1)}) + q\left(\frac{N_{(t-1)}}{m}\right)(m - N_{(t-1)})$$

EQUATION 7 - BASS CURVE FORMULA

[6] Cranfield had two hierarchical groups; G1 and G2, that completed the trial.

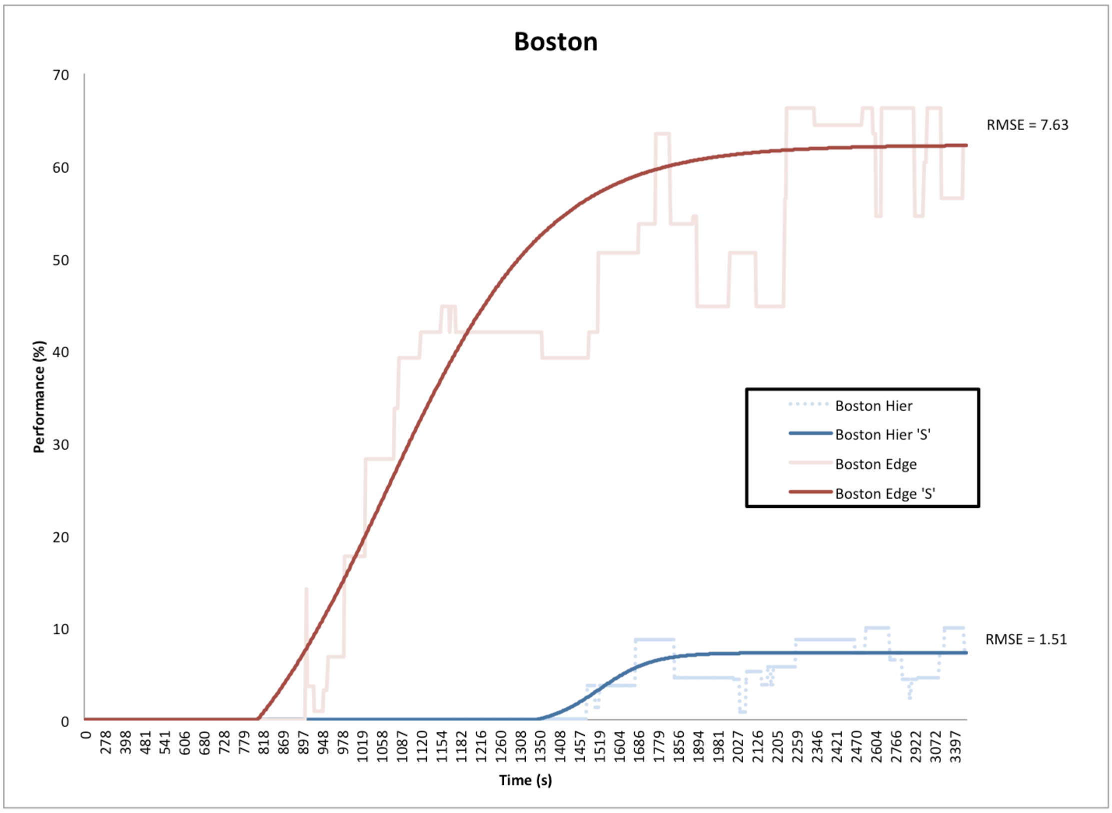

FIGURE 28 - MEDIAN PERFORMANCE FOR BOSTON GROUPS WITH FITTED BASS 'S' SHAPE CURVE

'p', 'q' and 'm' are coefficients of the curve. Traditional Bass curves are used in diffusion models where 'p' and 'q' are the coefficients of innovation and imitation respectively, and 'm' is the ultimate final market potential. In this case, 'm' is the final median performance score of the group. To obtain the best fitting trend line, values of 'p' and 'q' were adjusted to find the lowest Root Mean Square Error (RMSE) of the curve. RMSE is given by the following equation (Equation 8):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(sy_i - y_i)^2}$$

EQUATION 8 - ROOT MEAN SQUARE ERROR

Where '*sy*' is the value of the bass curve and 'y_i' represents the calculated median group performance score at a given point. The shape of the Bass curves is indicative of how tendency to the correct hypothesis is diffusing through the groups during each trial. The mechanism by which this diffusion is happening is of interest since in this experiment the people in the groups are not free to discuss hypotheses with one another. As such, it can be assumed that it was through the sharing of, what were perceived to be, pertinent factoids that a consensus towards the correct solution occurred.

The classical use of Bass curves is to demonstrate diffusion of innovation in networks. As such, it seemed reasonable to explore how the 'p' and 'q' coefficients of innovation and imitation might describe the groups in the ELICIT trials. The coefficients cannot indicate the exact number of opinion leaders (innovators) or sinks (imitators) in a group but may give an indication that a group has a strong or poor ability to filter and share the most pertinent factoids to the group. The coefficients for each trial are presented in Table 6.

TABLE 6 - BASS CURVE COEFFICIENTS

	Diffusion Coefficients					
	Edge			Hier		
Group Name	**p**	**q**	**perf.**	**p**	**q**	**perf.**
Boston	0.35	3.50	62.3%	0.22	1.05	7.24%
Canada	0.60	2.50	25.1%	0.01	1.85	44.7%
Cranfield	0.37	2.50	63.4%	G1: 0.50	G1: 5.70	G1: 7.24%
				G2: 11.0	G2: 6.00	G2: 6.35%
NPS	1.40	5.00	76.7%	2.00	8.00	31.1%
Singapore	0.48	3.30	70.1%	0.38	2.44	37.6%
Southampton	0.16	3.00	64.4%	0.10	2.70	31.1%
Westpoint	0.60	2.70	76.7%	0.13	1.00	28.7%

Plotting 'p' and 'q' against performance, Figure 29 and Figure 30 respectively, shows two clusters on either graph. For 'p' coefficient the graph shows a cluster with a coefficient value range of 0.16 to 0.60. For 'q' coefficient the cluster has values of 2.50 to 3.50. The hierarchical groups contrastingly show no clusters of results and can be view in Appendix 2.

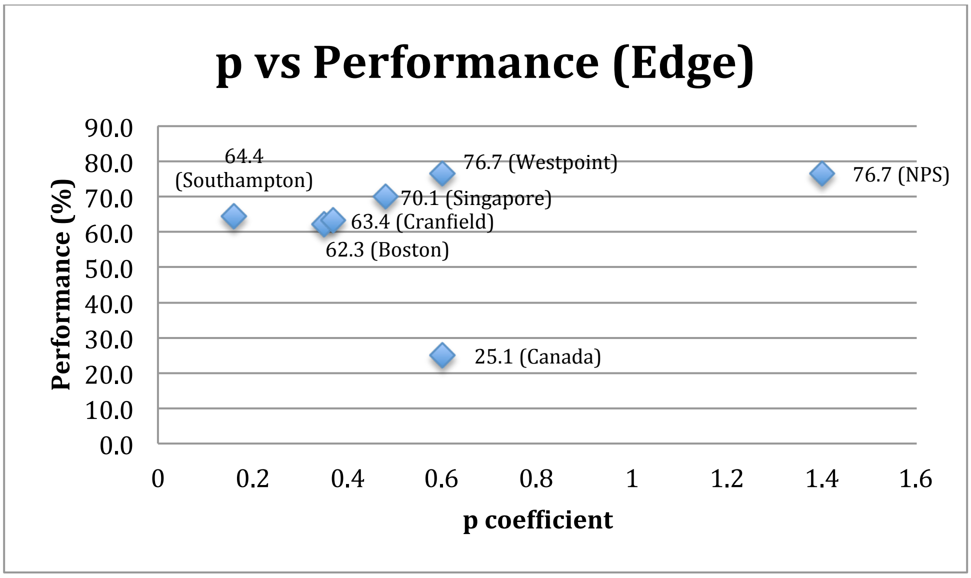

FIGURE 29 – 'P' COEFFICIENT VS PERFORMANCE FOR EDGE GROUPS

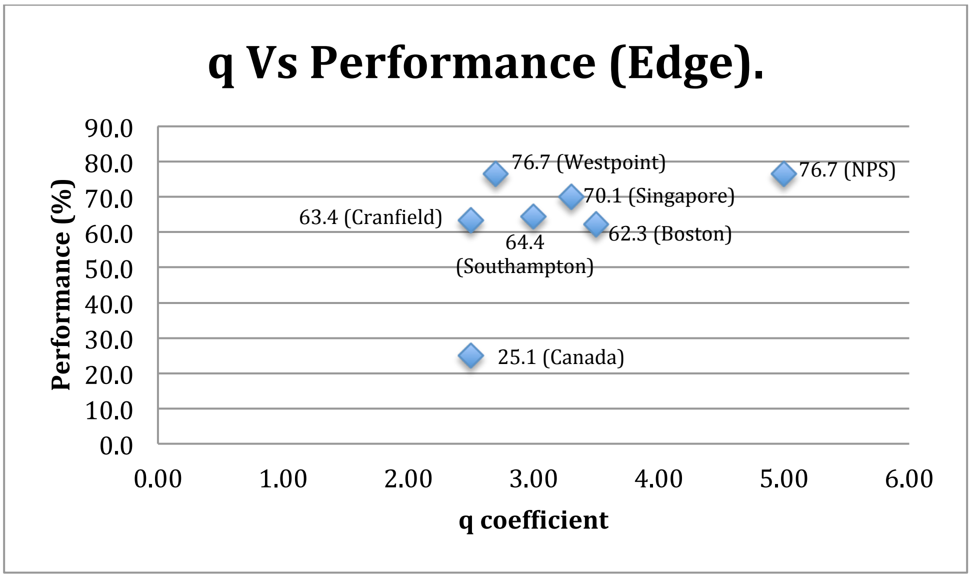

FIGURE 30 – 'Q' COEFFICIENT VS PERFORMANCE FOR EDGE GROUPS

Figure 31 shows all 'S' shapes for all the groups' median performance scores. Broadly, the Edge groups have curves that reach their asymptote earlier than Hierarchical groups and these asymptotes tend to represent much higher performance for the Edge groups. The results for individual trials are plotted in Appendix 2.

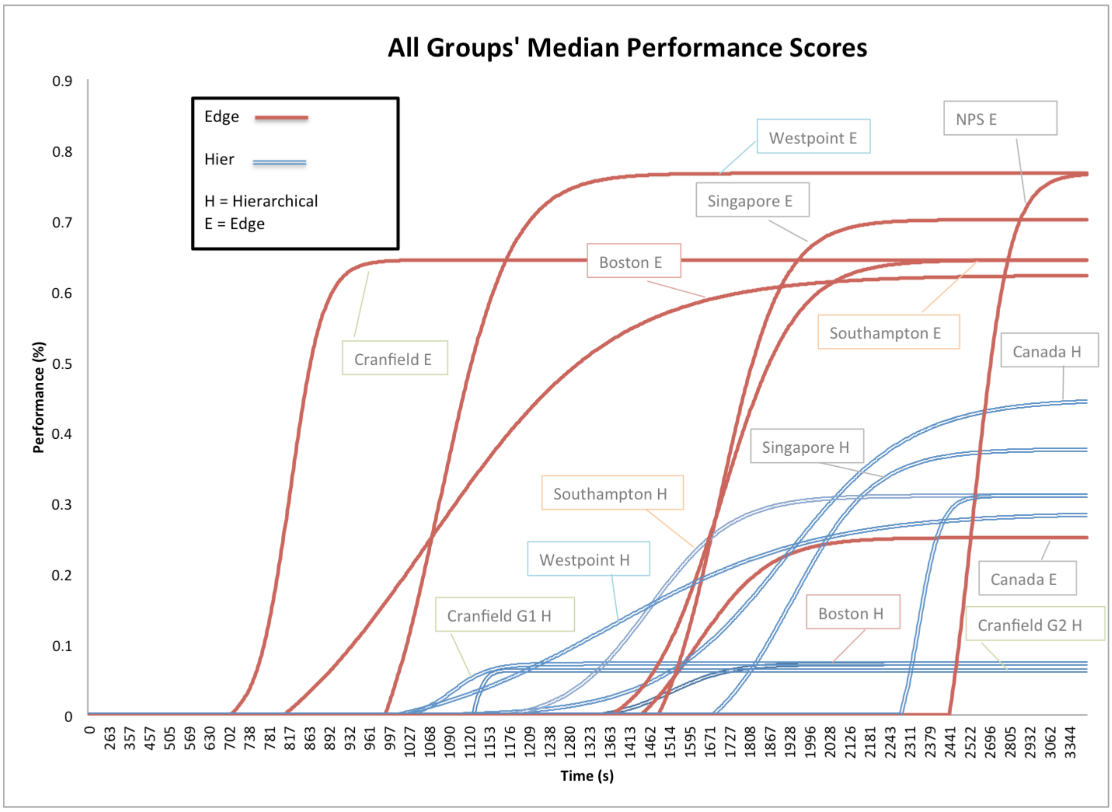

FIGURE 31 - THE MEDIAN PERFORMANCE SCORES FROM ALL GROUPS OVER TIME AFTER FITTING BASS CURVES

4.4.3 Density

Density in social networks represents the number of connections actually being utilised in the network as a proportion of the number of possible connections. Figure 32 shows the density plotted against median performance for each group. The regression lines show a slightly steeper gradient for the edge groups compared to the hierarchical groups. Figure 33 was also included to show what would happen if the results from Canada was treated as an anomaly. The Canada trial was the only trial where results were disparate to the trials carried-out elsewhere, in that the edge group performed worse than the hierarchical group. The Author did not carry out the data collection for this study and it is unclear if this outlying data was a result of generally poor performance or if there were other contributing factors. The regression model for the edge group fits the data considerably better ($R^2 = 0.46$ compared to $R^2 = 0.09$) with the result from Canada omitted. However, without knowledge of what explains this inconsistency the results cannot justifiably be removed from this analysis. Figure 33 is included so the reader can benefit from the visualisation of this insight. In this figure we see more of a difference between the gradients of the regression models. The hierarchical groups' performances seem not to be affected as much by higher density.

The edge groups' performances contrastingly do seem affected positively by higher density values.

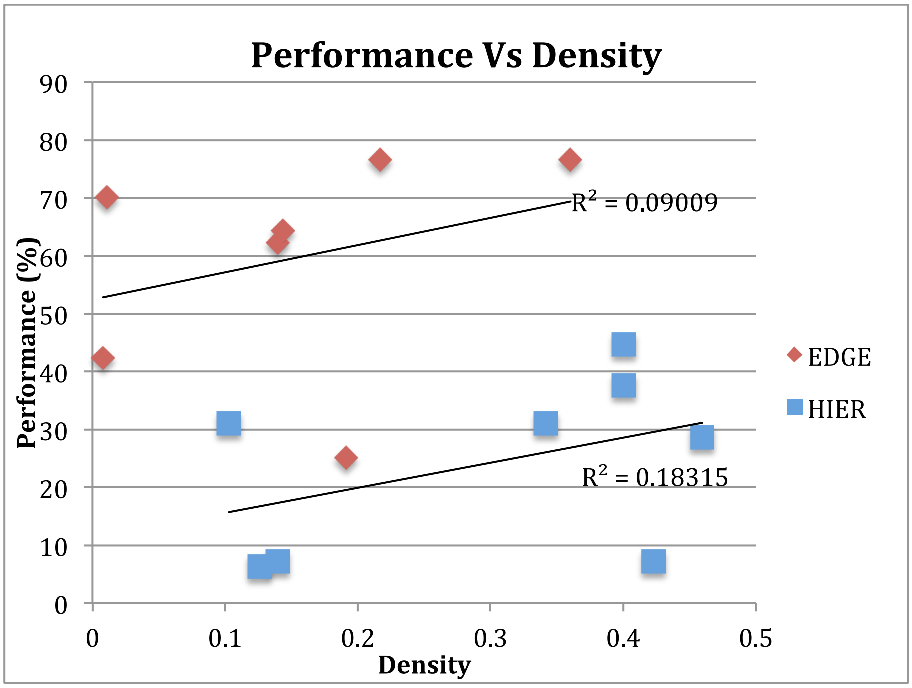

FIGURE 32 - PERFORMANCE VS DENSITY

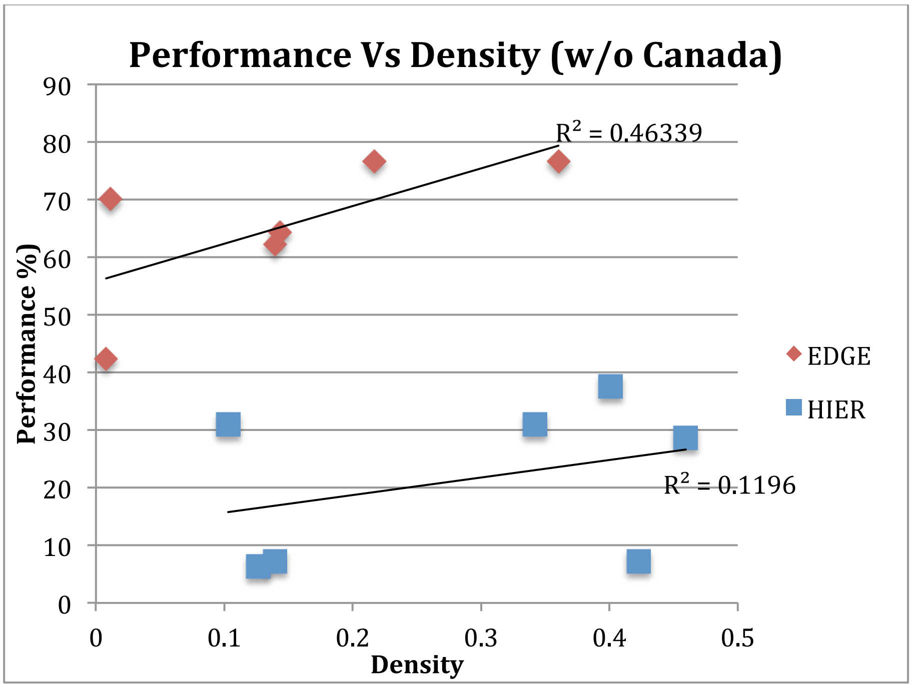

FIGURE 33 - PERFORMANCE VS DENSITY WITHOUT CANADA ANOMALY

The hierarchical groups are asked to behave as a hierarchical structure during the experiment i.e. constrain their communications through a rigid hierarchical command structure. As such the possible connections that each node should be able to utilise compared to the edge groups should be lower. In a scenario where a hierarchical group is utilising all its possible constrained connections, there should be a maximum density value of 0.32. However, Figure 32 indicates 5 out of 8 hierarchical groups have a density score greater than this value. As such it be assumed that these groups did not adhere to the communication constraints of the command structure.

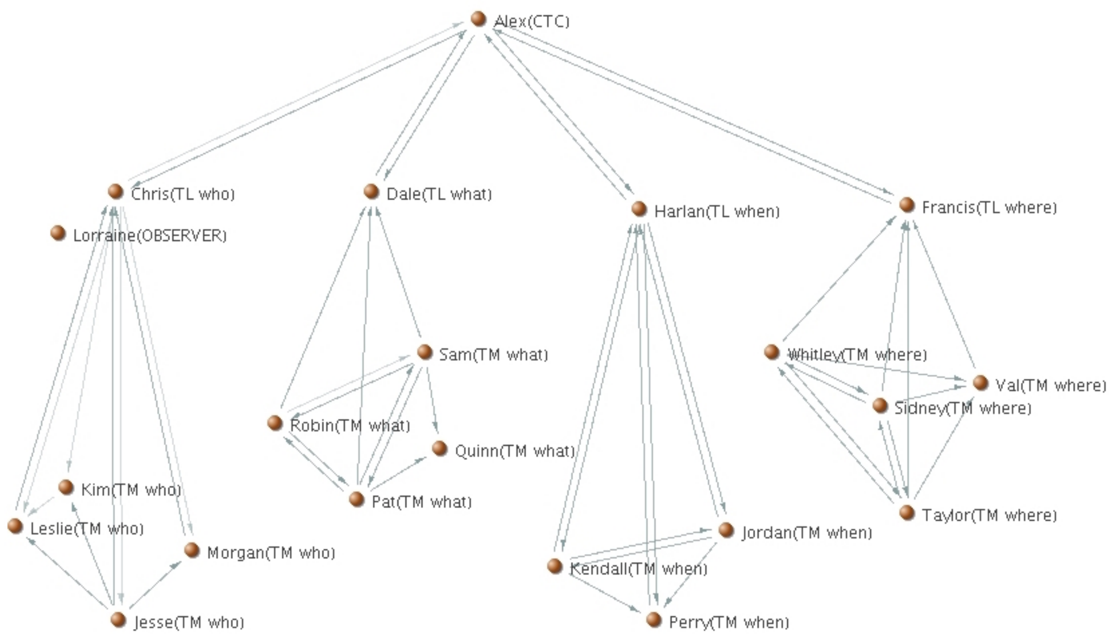

FIGURE 34 – SOCIAL NETWORK: CRANFIELD GROUP 1 HIERARCHICAL

Three hierarchical groups did have density metrics in the range that would indicate good adhesion to the imposed command structure. Figure 34 and Figure 35 show social network graphs from hierarchical groups from the Cranfield trial. The social network graphs[7] are generated from factoid sharing between group members. These groups accounted for two of the three poorest performing groups from all the trials, scoring 7.24% and 6.35% respectively. Within the same range of density, the NPS hierarchical group performed considerably better, 31.1%. However, the social network graph (Figure 36) for this trial shows considerable divergence from the command structure of the hierarchical condition. Figure 33 also shows that the hierarchical groups that ignored the sharing constraints imposed by the experimental condition still did not perform as well as the edge counterparts. As such, within the hierarchical condition there are additional factors inhibiting performance.

[7] CTC = Cross Team Coordinator. TL = Team Leader. TM = Team Member

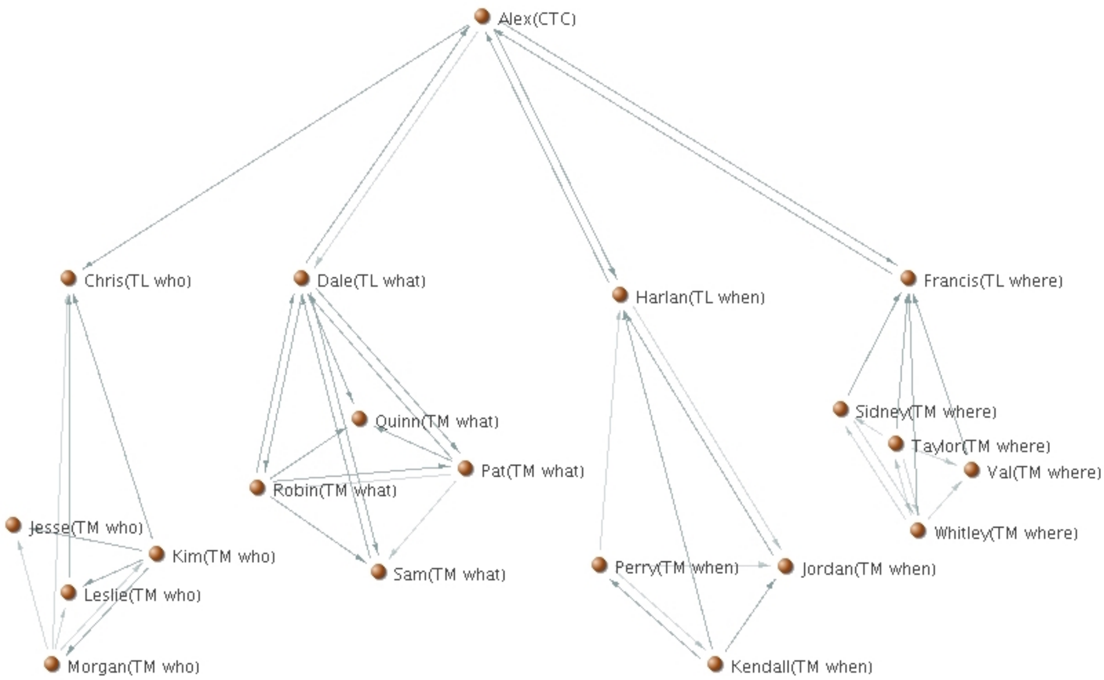

FIGURE 35 - SOCIAL NETWORK: CRANFIELD GROUP 2 HIERARCHICAL

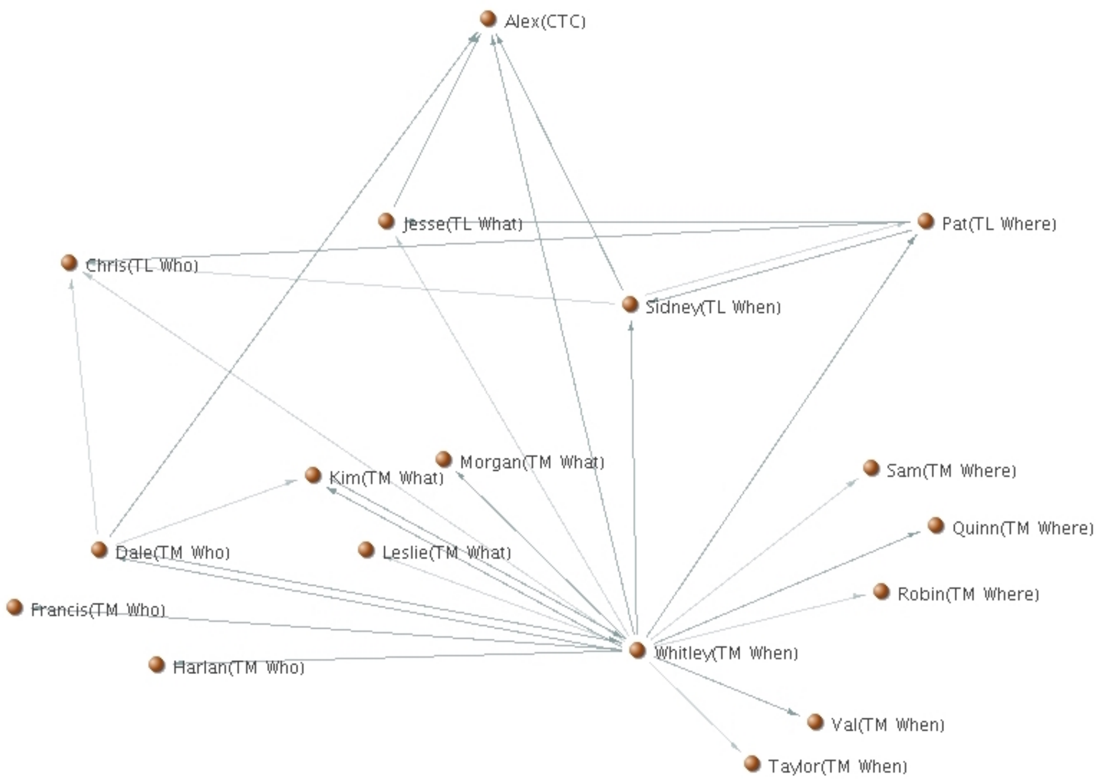

FIGURE 36 - SOCIAL NETWORK: NPS HIERARCHICAL

4.4.4 Interactions

In addition to the ability to share factoids participants are also able to pull and post factoids from central website. Under the hierarchical condition participants can only pull from the websites that relate to their functional group, e.g., members of the 'who' team can only access factoids on the 'who' website. In the edge group all members can access all sites.

TABLE 7 - TRIAL STATS ORDERED BY DESCENDING PERFORMANCE

Trial Name	Performance (%)	Density	time (s)	identify	share	post	pull
NPS Edge	76.66	0.217	2915	16	267	133	1305
Westpoint Edge	76.66	0.360	3612	62	1378	172	2232
Singapore Edge	70.1	0.011	2366	36	21	145	1368
Southampton Edge	64.39	0.143	2292	40	119	131	747
Boston Edge	62.25	0.140	3576	119	66	156	1769
Canada Hier	44.74	0.401	3616	80	499	160	1583
Cranfield Edge	42.42	0.008	1439	70	21	180	461
Singapore Hier	37.6	0.401	2462	44	484	203	1499
NPS Hier	31.05	0.103	3130	9	130	134	510
Southampton Hier	31.05	0.342	2440	42	326	154	167
Westpoint Hier	28.67	0.460	3649	66	653	180	1859
Canada Edge	25.1	0.191	3600	63	145	84	1311
Boston Hier	7.24	0.423	3622	76	358	138	1000
Cranfield G1 Hier	7.24	0.139	1237	32	223	97	84
Cranfield G2 Hier	6.35	0.126	1237	8	122	76	42

Figure 38 shows total factoid sharing per group against performance. Factoid sharing is, on average (Table 8), more frequent in the hierarchical groups than it is in the edge groups. Despite this, the average level of performance is much lower. A 'Pull' represents a factoid being retrieved from a website. On average the edge groups are pulling more information from websites than their hierarchical counterparts, which is to be expected since the hierarchical group members can only access one of the websites. The average number of posts to websites is identical for both conditions. As such, it would seem that the groups are behaving differently under the two conditions in the way they gather information. The hierarchical group is more dependent on gaining information through the sharing mechanism. The information available to them through the 'pulling' process is limited by access to the sites and in the hierarchical groups that adhered to their communication structure, would have then been reliant on information being passed to them through their team leader and information periodically given by the ELICIT platform. The hierarchical groups that chose to break their communication constraints are forced into doing more sharing since they cannot post to all websites thus the only way to spread information is in an ad hoc peer to peer network. The factoids that are distributed to a team member may stay locked within that functional group otherwise.

TABLE 8 - AVERAGE PULLS, POST AND SHARES

Average Sharing Edge:	288
Average Sharing Hier:	349
Average Pulls Edge:	1313
Average Pulls Hier:	843
Average Posts Edge:	143
Average Posts Hier:	143

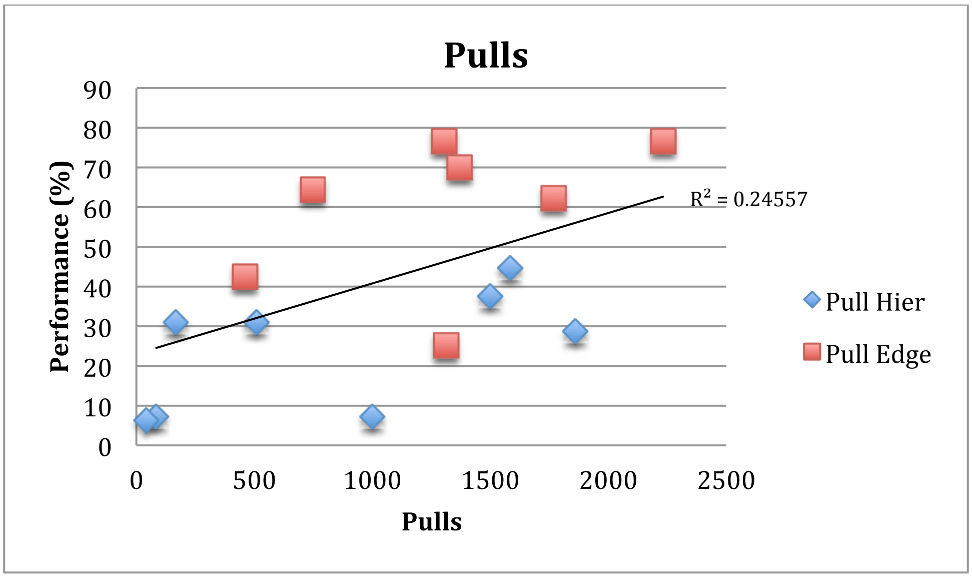

FIGURE 37 - PULLS VS PERFORMANCE

Contrastingly the edge groups, on average, are doing much more pulling and less sharing of information. This suggests that the participants in the edge networks are gathering and processing information more independently than the hierarchical groups. It might also suggest they only share when it is necessary. As individuals they can explore the available information and make initial hypotheses before passing on the supporting factoids. An argument could be made that, in this scenario, the edge network is a more effective system for filtering useful factoids from non-useful factoids since the tendency would be to share only the factoids that support and initial hypotheses. In sensemaking terms they are creating a frame around the information available to them and then they can continue to search the websites for supporting information whilst sharing information that supports their hypothesis.

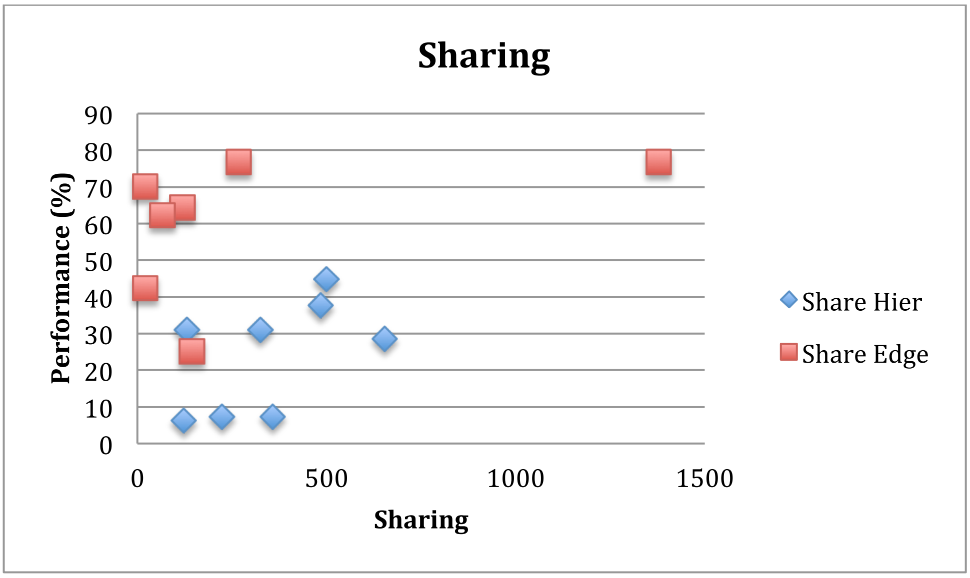

FIGURE 38 - GROUP SHARING VS PERFORMANCE

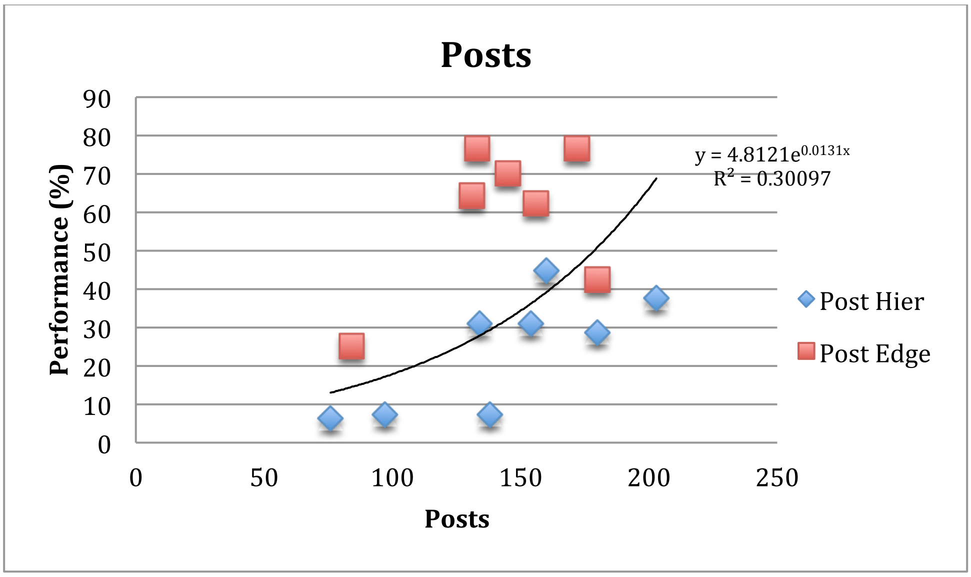

FIGURE 39 - GROUPS' POSTS VS PERFORMANCES

The total number of interactions per group is calculated by adding the number of posts, pulls and shares of a given group. This is shown in Figure 40 where total shares are shown in relation to a group's performance. The groups that have higher total interactions tend to have higher performance measures. In essence, the groups that are working harder, in terms of moving and manipulating information are able to formulate better hypotheses. In the hierarchical structure team members are only allowed to post and pull to their respective websites. As such information sharing between functional teams becomes necessary to prevent Team Leaders and the Cross Team Coordinator nodes becoming overwhelmed.

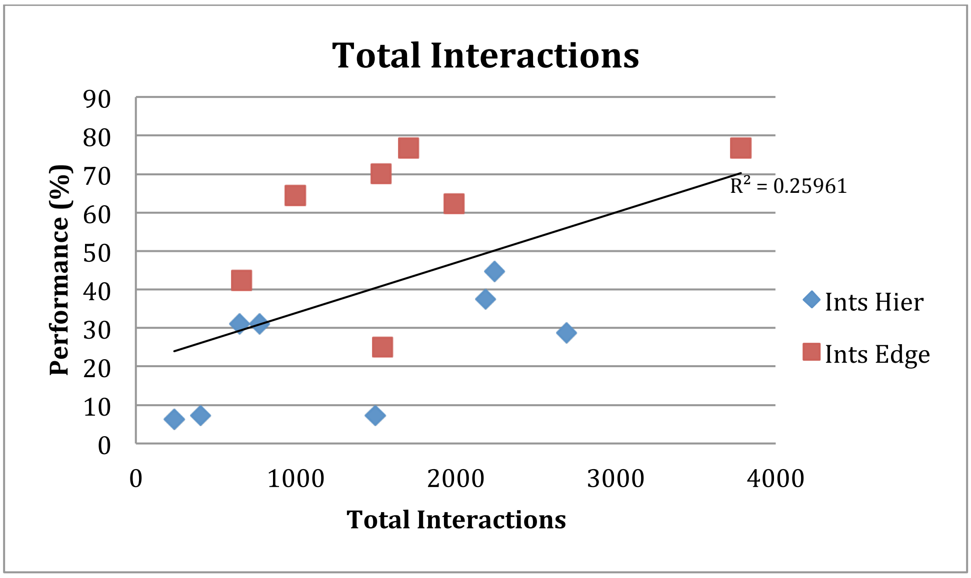

FIGURE 40 - PERFORMANCE VS TOTAL INTERACTIONS

4.5 Results Discussion

The results reveal that alongside the normal test conditions in this experiment there is a 3rd condition in which hierarchical groups are constrained by their ability to post and pull information from the websites but have ignored the communications constraints of the hierarchical test condition. As such, counter to the original aims of this chapter, we have observed a kind of pragmatic reorganisation in these non-compliant hierarchical groups. It could be assumed that this deviancy occurs due to the frustrations of working on a task requiring semantic sensemaking in an organisation that doesn't support this kind of behaviour.

It was predicted that the edge groups would have more information sharing because the there no constraints on communications within this group. However the results show the Edge groups have less information sharing and make less use of available connections than those in the hierarchical network while still managing overall better. Thus, the Edge groups seem to be more *efficient* than the Hierarchical groups in the way they are performing the task. To elaborate, the results from the majority of the groups show a 'S' shape characteristic, which was highlighted by the addition of Bass diffusion curves (Figure 28). These curves are showing the diffusion of the correct hypothesis. However, the ELICIT platform doesn't allow participants to communicate hypotheses, only factoids. So to achieve better efficiency and performance, the edge groups must be only passing on factoids with high relevance to the solution. In essence they are filtering out noise in the information stream as a network. The most plausible explanation for how this is being achieved is by the individual participants in the edge groups doing a lot of individual sensemaking. The premise is that the participants are prototyping hypotheses individually and using these as frameworks to filter out information being pulled from the websites. This is supported by the correlation of 'pull' rate and performance in the edge groups (Figure 37). Once an individual has a reasonable hypothesis, he or she can then pass the supporting factoids to other participants in the network that allows them to hopefully conclude the same or a similar hypothesis.

Assessing performance in relation to total interactions asks questions about the competency of the individual participants to do the task asked in the experiment. It can be seen that some of the hierarchical groups, despite the imposed limitations of their experimental condition, perform better than some of the edge groups. This tends to occur where hierarchical groups have high levels of interactions or are working harder to complete the task. It is inferred that high 'sharing' scores show a focus towards teamwork during the experiment. High 'pull' scores contrastingly, show a tendency towards task-work. Therefore the participants in the edge groups are working harder at the task of collecting and analysing information on an individual basis because they don't have to be concerned that other nodes in the network are reliant on them for information. As such, the behaviour of the Edge groups is directed at establishing consensus on the 'sense' for the problem; although their only way of doing this is by forwarding factoids that support their hypothesis.

Interestingly, none of the median performance scores for the groups ever reaches 100%. This means within each trial, members of the groups are submitting hypotheses that still contain some noise or error. To some extent this shows the real-world nature of collective semantic sensemaking. 100% consensus in groups is not always possible or desirable; there is a possibility that participants in a group could all reach the same wrong hypothesis if they didn't undertake some individual task work. However, a certain level of consensus indicates a point at which a decision for action could be made. In this experiment, where group members can only send factoids to each other and cannot discuss hypotheses, the median performance scores are a way of representing the individual sensemaking outputs collectively. From this perspective, the ELICIT platform, when applied in Edge condition, emulates the conditions under which, Suroweicki (2004) and Oinas-Kukkonen's (2008) (see chapter 2) suggest, a group of people can be wise. The ELICIT environment allows for the aggregation of individual sensemaking into a collective result where individuals are not directly influenced by other group members. All of which are precursors for 'wise' groups. The independence of individual group members to access information and formulate their own opinions is important in this instance for a group to perform well. It is also important that group members are able to share artefacts that have contributed to their sense of the situation. Interestingly, it seems less important that individuals have the ability to freely communicate with others in their network. Relating back to emergency response, there is existing research to suggest that the sharing of structured sensemaking artefacts with the control room is a more efficient way of collectively sharing the sense of the situation. McMaster et al (2012) observe that within police emergency response, the passing of formal artefacts supports the sharing of a common picture of the scenario encountered. These shared artefacts are an output of individual sensemaking efforts, which are supported by private (informal) artefacts.

This chapter has shown that quantitative measures can be used to assess the outputs of groups performing collaborative sensemaking. These measures related to three broad areas: information sharing, network structure and utility of information. While the latter measure can be defined in terms of ground truth and hypotheses for this study, the author accepts that this might be more difficult in real incidents. However, in terms of information sharing and network structure, messages or artefacts passed between individuals in networks and social network structures can feasibly be analysed on the fly. Referring back to Q2 in this thesis, the next question is how can technology be used to support collaborative sensemaking? This chapter suggests that supporting individual sensemaking as well as the mechanism for group collaborative work are key factors in the design of any supporting technology and will be explored in the following chapters.

5 Noisy Map Experiment – Pilot Study

This chapter builds on lessons learnt in the previous two chapters and starts an investigation into the use of technology to support sensemaking. The pilot study introduces and highlights strengths and weakness of a Communications Broker developed to aid sensemaking in a small group of participants collaborating on a task. It was found the designed technology doesn't support the individual's need to search for information in the problem space and was enhanced further for the following chapter's work.

5.1 Introduction

The previous chapter showed how organisational constraints can affect the semantic sensemaking of a group of participants performing a task that required collaborative sensemaking. The degree of independence that an individual participant had in searching and forming hypotheses was important for the overall performance of the group. However, in the reality of an emergency management scenario, knowing whether the outcome of a collaborative sensemaking effort is correct is not possible. As such technologies that support real-world collaborative sensemaking efforts must support the individual and group's information and coordination needs. This chapter starts by looking at the pragmatic workings of a group involved in a sensemaking task and how technology might influence or alter the performance of this group. Networks of people performing a task that requires sensemaking have varying degrees of semantic and pragmatic overheads. Semantic overheads, as seen in the Chapter 3, involve establishing a common ground for discussion and agreeing a common meaning of knowledge artefacts and plausible explanations. Pragmatic overheads involve establishing the principles by which the group will operate, pass information and interrelate with one another. As reviewed in chapter 2, there are different ways in which technology might be deployed to help a group of people engaged in a collaborative sensemaking effort. In this chapter, the aim of the Noisy Map experiment is to see if the introduction of a technology in the form of a *Communications Broker* (section 2.8.1) can benefit a group from a pragmatic sensemaking point of view.

The Noisy Map experiment is broken into two parts; the 'pilot study' and the 'zombie study'. It was inspired by the nature of communications in multi-agency emergency/disaster response. In the UK a large event requiring a multi-agency response would typically involve the Police, Ambulance and Fire services. In the US, it may additionally include several civilian charitable organisations and official response units from differing jurisdictions. In such scenarios, utilising a common communications technology is an important step in aiding cross-organisational coordination. Within the UK, the ability to use TETRA-standard (TEerrestrial Trunked RAdio, the European Standard for digital emergency services communications) networks and handsets across all three emergency services is now in place (Airwave, 2010). Different organisations operating compatible digital radio networks, have the potential to greatly enhance the ability to communicate within and between services and allow the coordination of communications and information. The TETRA technology enables the creation of 'chat groups' in different channels and also, all radio communications are automatically recorded and can be played back instantly via the communication software. However so far, the emergency services have struggled to make best use of the new technology for cross agency collaboration (McMaster & Baber, 2006).

They require the ability to carry out collaborative problem solving, which in turn requires close coordination between the emergency services at all command levels (Figure 41). When these incidents occur the ability for the emergency services to act as Communities of Practice (section 2.6) is affected because the escalating incident will often challenge the pragmatic workings of the group. In these circumstances what is required is a process for developing a shared multi-agency understanding of the incident domain, known as a Common Operating Picture. McNeese et al. (2006) define Common Operating Pictures as the representation of information in order to generate situation awareness across team members. The Common Operating Picture will generally be managed at the incident control room, rather than on-scene. It has the potential to facilitate decision making, situation awareness, collaborative planning and assists the various levels of command across the services in achieving shared awareness of the situation. As well as ensuring that only relevant information is passed to the personnel at the scene, a Common Operating Picture with multiple layers would allow commanders to have oversight of the status and distribution of the other services, enabling implicit collaboration (Baber et al., 2007; Keuhlen et al., 2002). Essentially the Common Operating Picture is an output of a collaborative sensemaking effort, in which a plausible explanation of the environment has been established to the extent that a coordinated response to the situation can be attempted.

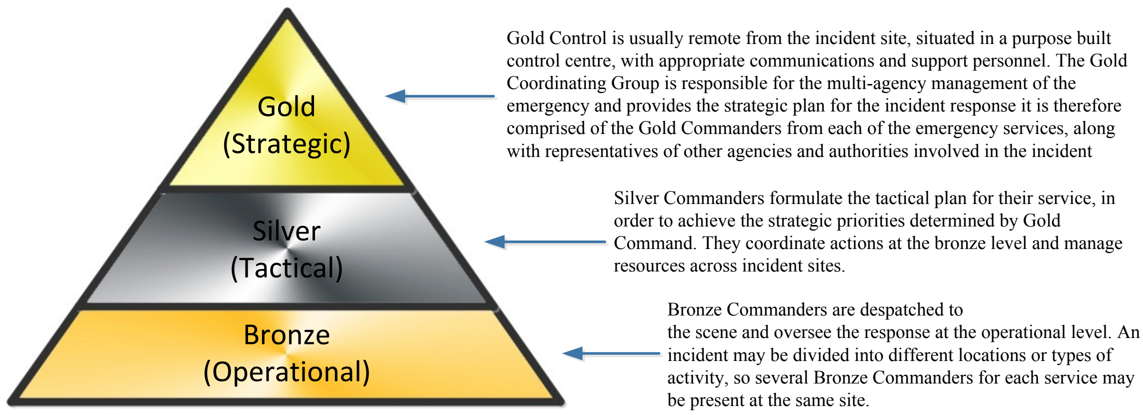

FIGURE 41 - EMERGENCY COMMAND AND CONTROL STRUCTURE IN THE UK

The case study below demonstrates the difficulties of a real-world cross-agency response to a large-scale event.

5.2 Case Study: 7/7 bombings

On the morning of 7th July 2005, four bombs were detonated on the public transport system in central London; three of the explosions took place in quick succession on London Underground trains, with the fourth detonating on a double-decker bus. 52 commuters were killed and over 700 were injured. The attack was designed to cause maximum disruption and London's emergency services quickly activated a large-scale response. Despite their best efforts, the emergency services encountered problems in organizing their initial response in terms of a collaborative cross-agency effort. These difficulties were arguably attributable to problems of organizational control and information sharing (7 July Review Committee, 2006). Whilst service control rooms and resources at the scene were able to share information, and pre-existing response-plans agreed by the services indicate that they should liaise with each other from the start of an incident (LESLP 2007), the emergency services still faced problems in coordinating the collection and assessment of incident information and in providing appropriate resource levels at each scene. Response data presented in the Report of the 7 July Review Committee suggested the reason for this may lie in the fact that the services were organising their responses using their individual command structures and information management systems, with little incident information being passed between them.

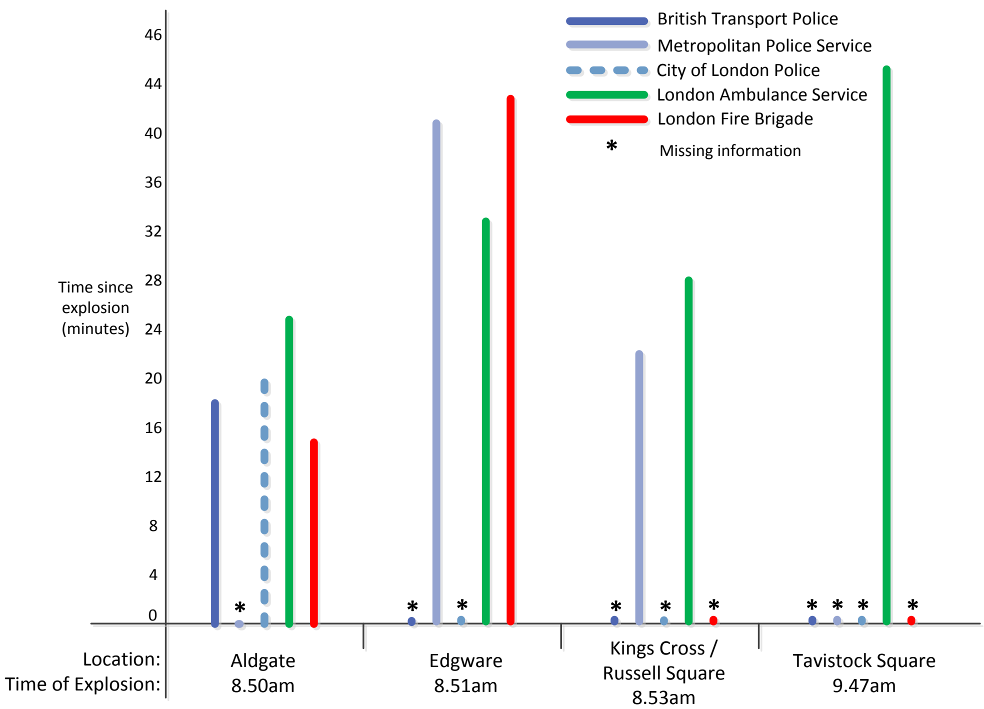

Figure 42 - Time since explosion for emergency services to declare a major incident at the scenes (based on timings in The Report of the 7 July Review Committee)

Declaring a major incident in a timely fashion is important because strategic command (Gold level) doesn't convene until this declaration has been made. Figure 42 shows the elapsed minutes before each organisation involved in the response effort declared a major incident at the various scenes and emphasises the variation in response across organisations. The large disparity between the declaration times suggests that the services were unable to share information adequately in order to reach consensus on the nature of the incident during the initial response and were operating as three separate services. Additionally, as the number of major incidents increased there is a decrease in response efficiency. This is hinted at by the increasing time to declare each major incident and the number of services that failed to make or acknowledge that declaration. This may have been partly due to information-overload in the emergency service control rooms.

The issue of communications-overload is implied by Figure 43 (Duffy et al, 2012), which shows the social network diagram for all incidents on the 7th July. The central nodes in the network are the various response agencies' control rooms. The diagram shows the control rooms are heavily engaged in communication with their respective response units; in essence stove-piping information. This is to be expected, as the control room coordinates operations centrally during the initial response phase of an emergency. In contrast, the communication between the various control rooms is quite sparse, which suggests that the services were organising their responses to the emergencies separately.

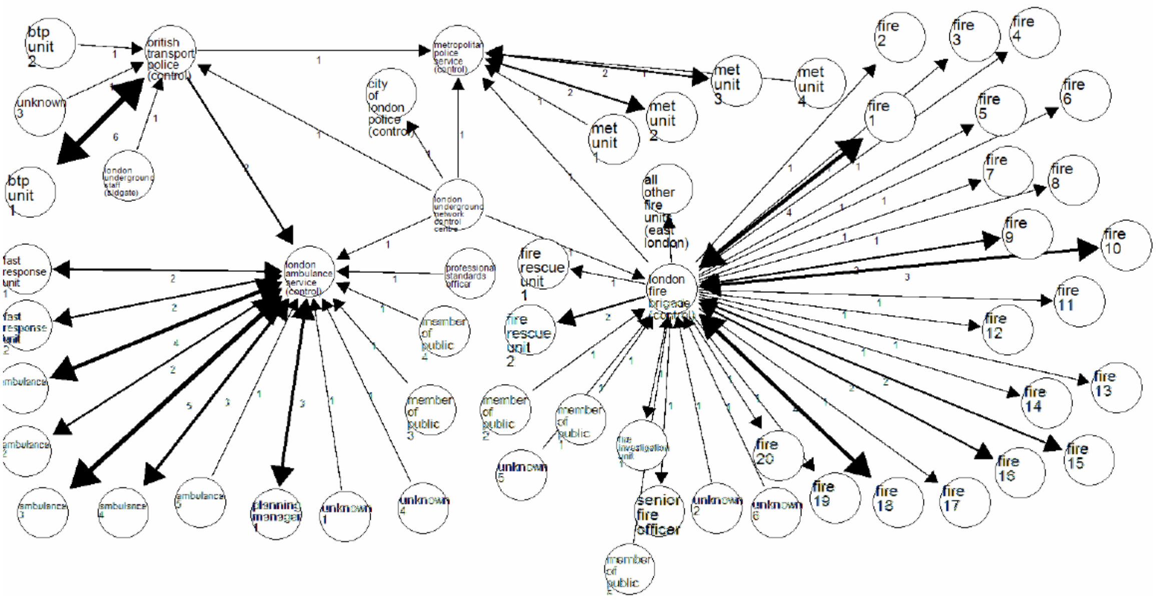

FIGURE 43 - SOCIAL NETWORK FOR THE INITIAL RESPONSE TO ALL INCIDENTS ON THE 7TH JULY (DUFFY ET AL, 2012).

Major incidents require close coordination between the emergency services, due to high levels of uncertainty (especially during the initial response); the presence of hazards within the inner cordon and the potential for the situation to rapidly change, requiring the use of different tactics. Given the demands of major incidents, it might therefore be expected to find that the emergency services had employed some form of Common Operating Picture to facilitate information collection, analysis and sharing and therefore aid rapid, coordinated response planning. However, the account of the immediate response in the Report of the 7 July Review Committee indicates that this was not the case; each emergency service investigated, analysed, resourced and classified each of the incidents separately.

FIGURE 44 – CROSS-AGENCY MEETING/HUDDLE 7TH JULY 2005

This issue of stove-piping was resolved at the bronze level on 7th July by the formation of cross-agency huddles at the sites of the incidents (Figure 44). This shows a social solution to the challenge of resolving cross-agency situation awareness. This huddle can share knowledge, set goals and start operating as an inter-agency Community of Practice. However, this solution is not always ideal, as the huddle can only be formed at a point when the battle rhythm of the incident dictates there is time to do so. The emergency services also have information liaison officers whose role is to carry information between services. However, as the communications load increases, this role can quickly become over-burdened. The issue of stove-piping at silver level occurs because incident control rooms are geared towards managing their own people rather than managing and sharing incident information across control rooms. The lack of cross-agency communication at bronze and silver levels in the initial response inhibits collaborative sensemaking. The Common Operating Picture relies on the collection and organisation of incident information that supports the strategic dissemination of this information to the appropriate units. This is a collaborative semantic sensemaking task at the tactical level. Contemporary approaches to the development of a Common Operating Picture often draw on the notion of ontology as a means of structuring the underlying knowledge. In order to better support Joint Service response, there is a requirement to develop a Common Operational Picture. There are attempts at creating complex, structured ontologies for emergency response (Babitski et al, 2009). However, the use of complex ontology structures may not be desirable or necessary since the emergency services, in an effort to create cross-agency Common Operating Pictures, have already started standardising the way in which they report information back to control. It is appropriate for the first resources that reach the scene to use the acronym 'CHALET' (Casualties Hazards Access Location Emergency-Services Type) for structuring their initial reports. From a sensemaking perspective this allows the emergency services the ability to gather incident information that fits into a common framework of understanding. These initial reports are structured artefacts that can be shared with the control room to help spread a shared understanding of the scenario. Attempts to enhance the cross-agency cohesiveness of the UK emergency services for major incidents, take the form of large multi-agency training exercises. However, with these exercises being few and far between their usefulness could be called into question. As such, it seems sensible to explore avenues of alternative operational philosophies and technological aids.

The case study above shows a real-world scenario of well-defined organisations struggling against the challenges of an uncertain and changing environment. The UK emergency services personnel are highly trained individuals and on an average day perform excellently in responding to routine emergencies e.g. car crashes and building fires. In such emergencies there is frequently a multi-agency response however, experience and training of the individuals allows them to rapidly categorise situations and then act upon them. As such, there is little requirement for collaborative sensemaking. The service personnel in this instance are working as a well-defined Community of Practice. In a large crisis the capabilities of the Community of Practice are likely to become overwhelmed and larger strategic response governed by a Common Operating Picture is required. As discussed in *section 2.8.1,* a proposed solution in this thesis to the challenge of strategic brokering and dissemination of information is a *Communications Broker* that will propose communication links between emergency responders. The goal is to support the development of a Community of Practice by inviting people with common issues to collaborate. In the scenario above, the primary task would be to analyse the information flowing through the emergency services communications channels and build themes based on the content. These themes could then become the topics of dedicated chat rooms.

5.3 Noisy Map Experiment: Pilot Study

With the above as motivation the noisy map experiment set about trying to establish if and how a communications broker might influence the performance of a group performing a sensemaking task. This study was developed to explore how aspects of a communications broker system might work. The aim was to see if a group of people performing a task, which they could only solve collaboratively, was aided by the introduction of a tool that gave them a real time indication of the amount of information they had in common with other users in the system. The pilot study involved small groups of participants working collaboratively to resolve information plotted on a map.

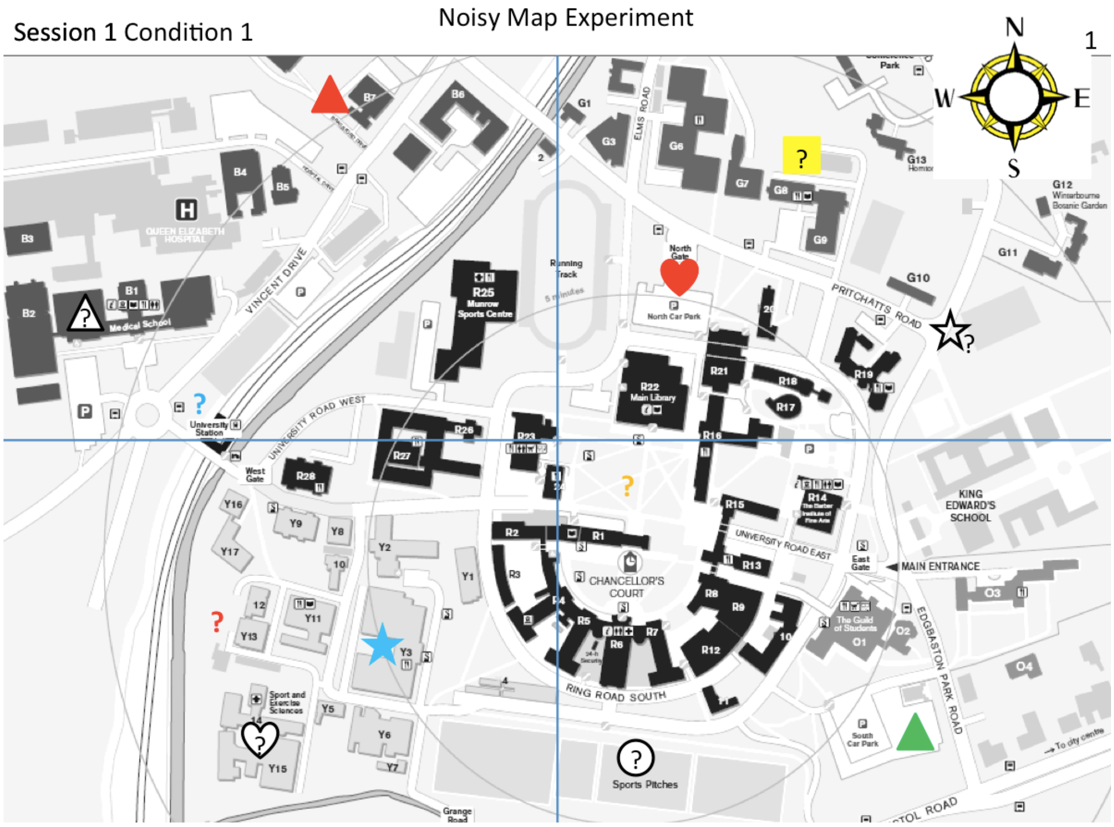

FIGURE 45 - EXAMPLE SCENARIO MAP

5.3.1 Experiment Design

Figure 45 shows an example of the scenario map with which each participant was presented. On the maps were a number of icons. Each icon had three properties; a shape, a colour and a location. However, one of these pieces of information was missing or erroneous. Between all the maps given to participants, the complete resolution of the information on the maps was possible. The scenario given to the participants was that they were the first responders to an incident and a map had been electronically sent to them but the communication had been distorted and some of the maps information was missing or wrong. Using only the system provided (Figure 46) they had to communicate with other responders and resolve their own map. Each session had 30 minutes to complete the task.

Participants were sourced from students taking an introductory course in human performance at the University of Birmingham, UK.

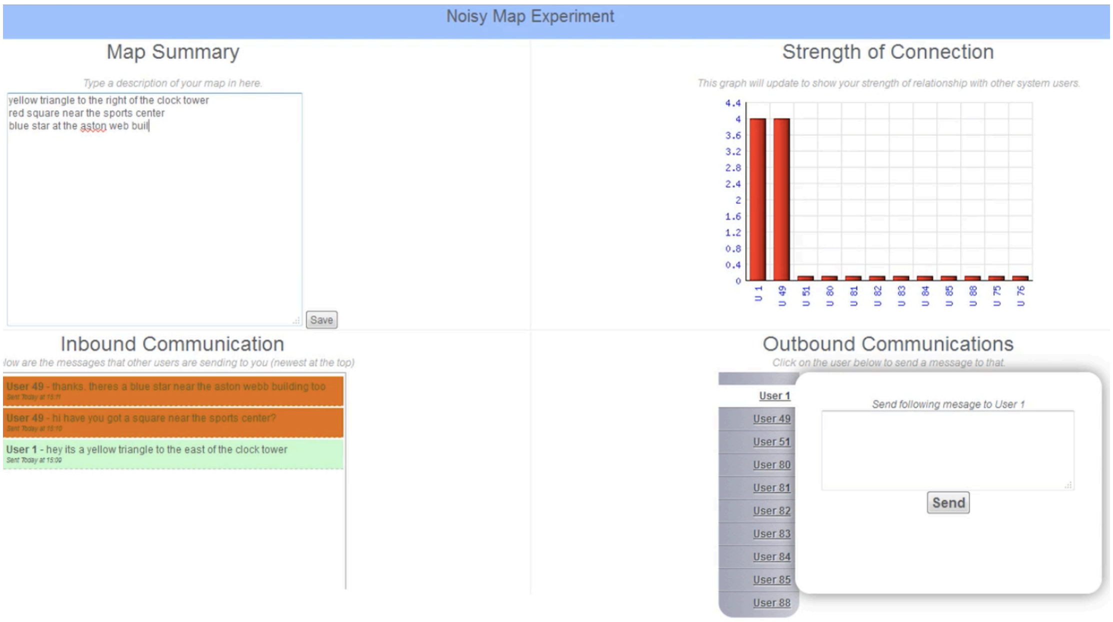

FIGURE 46 - MAP STUDY SCREEN SHOT

The system contained a free-text map summary box (shown top left of Figure 46). Part of the task given to participants was to keep an up-to-date description of their map in this box. The map summary is a statement of a given participant's current situational awareness and in an emergency response scenario would represent the information that is reported to the control room that would be fed into the incident ontology. The system also contained an inbound and outbound communications box. This allowed participants to send messages between one another. The final part of the system was a 'strength-of-connection' bar chart (top right of Figure 46). It showed the relative number of key terms that each participant had with the other participants of the system. The key terms were found, as in chapter 3, by removing stop or noise words from the participants' map summaries, stemming the remaining terms with the porter-stemmer algorithm and then consolidating like stems in a frequency array. The final stage was to calculate the intersection of each participant's frequency array with the other participants in the system. This score of the number of terms in common was used to dictate the height of the strength of connection bar chart and was updated every few seconds.

The experiment had two different conditions. Under the first condition the participants were aided by the Communications Broker and were able to see a strength-of-connection that indicated to them other participants in their network using similar key terms. The second condition did not have access to the chart. The hypothesis was that the strength of strength-of-connection chart would help participants see who else in their group were talking about similar things and it would help their performance in the task of resolving the map information quickly by leading them to the information they needed. The trial was separated into 5 sessions where 2 groups of participants took turns in attempting 2 different scenarios. The first session was a training session where participants were familiarised with the system. Sessions 2 and 3 were held simultaneously immediately following session 1. There was then a short break before the groups switched conditions and sessions 4 and 5 were conducted side by side. At the end of each trial, participants were asked to return their maps with the resolved information completed and this was compared to the complete scenario map and given a score. They were also asked to complete a NASA Task Load Index after each session. Results for the training session (session 1) are not used for analysis.

5.4 Results

TABLE 9 - PILOT STUDY STATISTICS BY SESSION

Session	Group	Broker	N=	Avg Comms[8]	Density	Avg Activity	Perf (avg)[9]
2	1	No	4	20	0.92	27.5	13.1
3	2	Yes	5	13.4	0.60	30.2	16.0
4	2	No	5	6.6	0.75	28.2	18.5
5	1	Yes	4	7.25	0.75	22.5	16.3

Comparing groups 1 and 2 in sessions 2 and 3 respectively it can be seen that density is greater for the network in session 2 showing a high connectivity between participants. Session 2 also has higher average communication per person despite having lower number of participants than session 3. However, session 3 still manages to achieve a better average performance for the task. The average activity score given for each group represents the average number of times each participant updates their map summary. It is assumed that map summaries are updated when new or resolved information becomes available. For session 2 the group seems to be dedicating more time to establishing communications and searching for information from other participants. Session 3 in comparison has lower average communication per participant yet higher average activity. The TLX results for these sessions reveal that session 3 experienced marginally lower levels of mental demand, temporal demand, effort and frustration. These results are presented in chart form below:

[8] Average number of communications per participant.

[9] Average performance. This was scaled so fit with the NASA TLX scale.

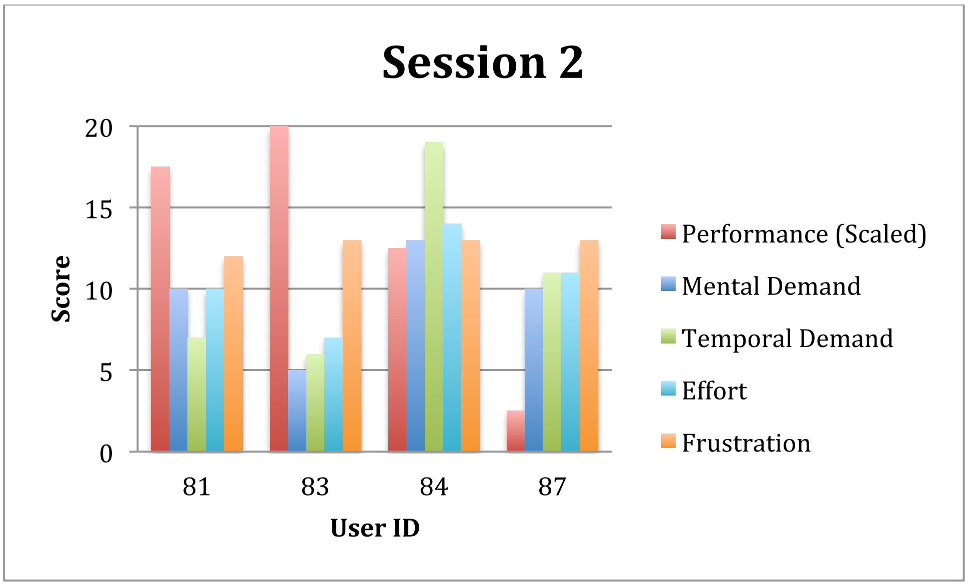

Figure 47 - Session 2 Results – No Broker

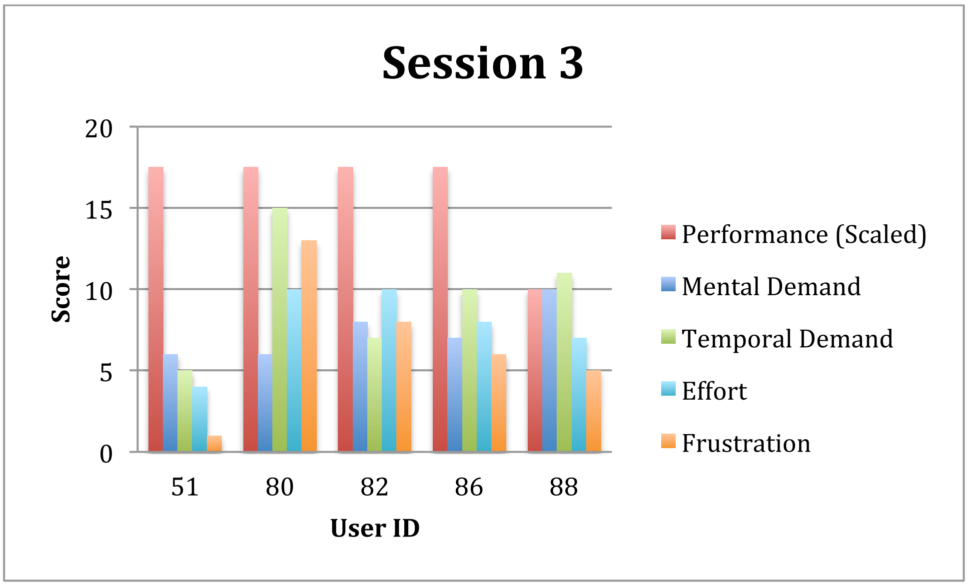

Figure 48 - Session 3 Results – With Broker

Towards the end of session 3 a behaviour developed in group 2 where by one of the participants (User ID; 80) believing he/she had completed the task, copied and pasted his/her solution to the whole group except to the participant with User ID: 88. It is unclear why this participant was excluded. Effectively User ID 80 tried to broadcast the completed solution. However, there was an error in the broadcast solution. The recipients, rather than comparing the broadcast with their own current situational awareness, decided to accept the received solution. As such their scores for the task (Figure 48) are all the same but not completely correct with the exception of the excluded participant. User ID 80 was able to complete the task before the other participants before broadcasting his/her solution.

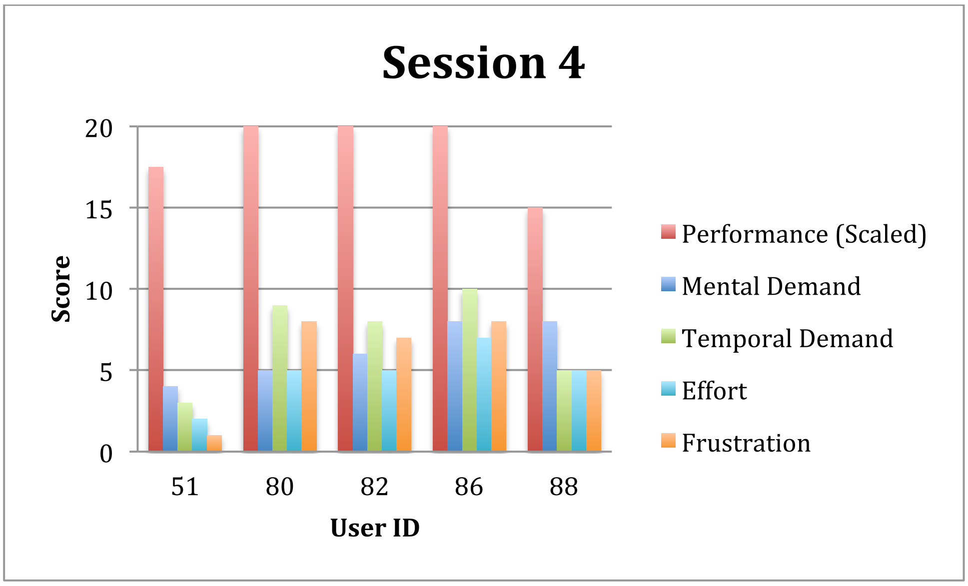

FIGURE 49 SESSION 4 RESULTS - NO BROKER

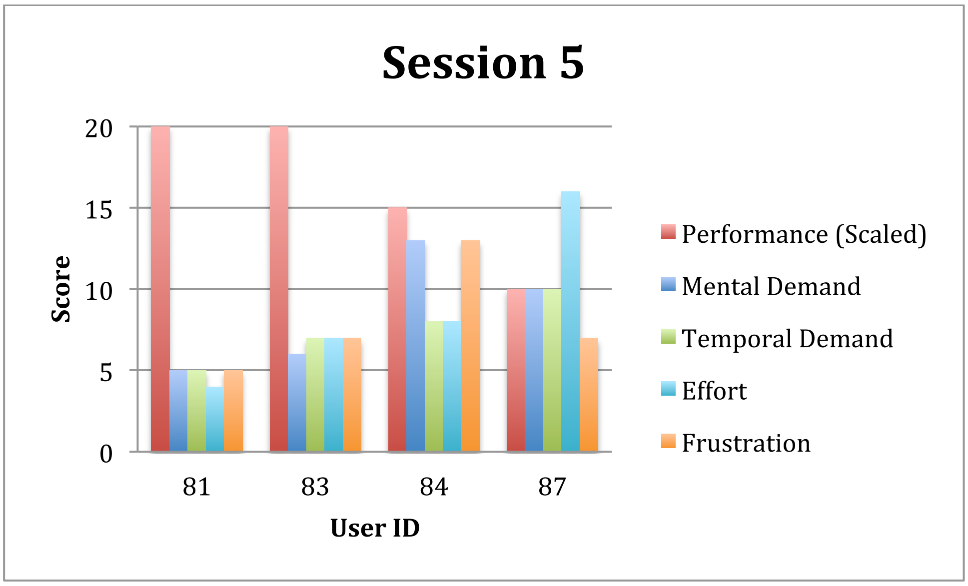

FIGURE 50 - SESSION 5 RESULTS - WITH BROKER

For sessions 4 and 5, all participants had adopted the practice of broadcasting map summaries. However, in these sessions the participants were sending incomplete map summaries and not just the solution. This had the effect of flooding the social network with information, which then became too costly for all users to sort through. This was reflected by the reduction in the average number of communications per participant in conjunction with activity levels remaining similar to the first two sessions. This suggests that the participants were now focusing their efforts on reading through the broadcasted map summaries and sifting for information that was not present on their own maps. This approach minimises the necessity to communicate so frequently but adds overheads in terms of time spent analysing incoming broadcasts. This might explain why performance was less consistent amongst participants in sessions 4 and 5. However, the access to this large broadcast of map summary information from all users did have the effect of increasing average performance scores for each group from their first respective sessions.

5.5 Discussion

The pilot study proved useful in identifying some features that would be necessary in a Communications Broker technology. Firstly, any technology should only be deployed where the task is large enough to necessitate assistance with communications. The networks in this task were small enough that the amount of information being broadcast from each node still was not overloading the capability of each node to process this information completely. Clearly adopting this behaviour in larger networks acting on more challenging tasks, would lead to information overload for most nodes. Unfortunately time and participant numbers were not great enough to test larger groups in this initial study. In chapters 3 and 4 there was discussion about group size and its effect on collaborative sensemaking. Larger groups have more overheads in terms of establishing common ground. The results and behaviour of the group in session 3 emphasises that contrastingly small groups struggle with a common knowledge effect. The unexpected behaviour of broadcasting, what was perceived to be, the completed map summary reaffirmed that misinformation can be taken as valid when the numbers in the crowd aren't large enough to resolve it. To clarify, there was not a person in that group willing to challenge or reassess the situation report that was broadcast to the majority of the group; it was collectively accepted without question.

A debrief with participants after the task, revealed that some participants had made use of the Communications Broker in the run of the experiment (sessions 2 and 3). Comparing sessions 2 and 3 it can be seen that density and average communications are lower for the group in session 3 (despite session 3 having one extra participant) and average activity is higher. These indicators suggest that the group in session 3 was working more efficiently. However, this cannot be established because the performance scores for this group were skewed by the late introduction of the broadcasting behaviour in that session. The participants did suggest that in a task where one is searching for missing information, it is actually more useful to ask participants with whom you have less information in common, since they are more likely to have new information for you. This echoes Granovetter's (1973) work on the strength of weak ties. There were no restrictions on the organisation of the networks in this task, thus allowing each node in the network to search freely for missing formation. It makes sense that any assisting technology must be presented in such a way that these explorative networks have the ability to see where they need to search. As such, the proposal in this instance is that the current Communications Broker design might actually have a detrimental effect by aiding common knowledge effect: Achieving an amount of consensus is important in collaborative sensemaking exercises. Socially plausible explanations of the situation will ultimately guide collective action. Thus, the mechanism by which consensus is achieved is important. In the ELICIT study a certain amount of independence in the way that participants sought and processed information in the environment seems to have been key to good group performance. The Communications Broker in this experiment was designed to try and help the pragmatic sensemaking of the group by suggesting communication links that would direct the participant's search for information. However, if that direction is based on commonalities, the technology might be inadvertently encouraging common knowledge effect by removing the possibility of weak ties.

The theory of weak ties is about a person (the weak tie) moving between existing clusters of people to collect and dispense information. The premise is that people in existing clusters learn more important information from the weak tie than the people already within the cluster. In this scenario, a more useful supporting technology might look for existing or evolving clusters of people within the experiment and then create a weak tie by suggesting the information contained within that cluster. Returning to the emergency response scenario presented at the beginning of this chapter, the potential usefulness of this in maintaining AIRWAVE chat groups or tactical chat, chat rooms can be seen. This will be considered in furthering the design of the Communications Broker in the next chapter.

6 Noisy Map Experiment - Zombie Study

This chapter builds on lessons learnt from the pilot study presented in the previous chapter and also brings together aspects of the ELICIT (chapter 4) and the Article Commentary (chapter 3) studies. The design of an experimental platform is described. The experiment required participants to perform a task in a changeable and unstable environment with differing group organisations and supportive technology. The key question addressed was how technology designed to support collaboration in groups affected group performance (both in terms of communication and in terms of response management).

6.1 Introduction

An interesting aspect of Noisy Map pilot study was that it displayed the problems of common knowledge effect within small groups. Although it did not give a clear insight as to whether the Communications Broker technology was able to influence or alter the behaviour of a networked group of participants, it did highlight some design features that should be added into the technology to be developed in this chapter and shows the benefits of conducting pilot studies; the broadcasting behaviour that developed could have been detrimental to this subsequent experiment otherwise. It was not obvious how the groups had responded to the broker, which was designed to direct participants towards the formation of subgroups on the basis of information in common. The groups of participants were not quite large enough to stop the broadcasting behaviour from becoming completely debilitating to the network. Inadvertently, there is evidence to suggest that one of the reasons why the Communications Broker had no empirical effect on the behaviour of the groups was because group size was already optimal for the task in the pilot study. Hoegl (2005) reviewed 58 cases where teamwork quality and communication were strongest in teams of 3 to 6 people, quoting an average of 4.4 people. The average group size in chapter 5's pilot-study was 4.5 people. In contrast the worst performing teams according to Hoegl(2005) were sized 7 to 9 people (average 7.8). From a pragmatic sensemaking viewpoint, this re-enforces the notion that small groups do not have the communications overheads of larger groups where it is more difficult to decide where information should be sent (section 3.7.1). It also suggests that group sizes do not have to increase drastically for these problems to become evident.

It was unfortunate that access to good numbers of participants was a problem in the pilot study. This highlights a problem with research into teamwork; in order to study the behaviour of teams which are large enough to be analogous to the size of teams encountered in real operations, one needs to assign more than 6 people to a team, but in an average cohort of students, this might only result in a small number of teams, and the need to ensure that there are equal numbers of teams to ensure comparison presents a further challenge. While it would have been beneficial to run studies using multiple teams under each condition, recruitment problems meant that this was not possible. In the experimental work in this chapter group sizes of 7 to 9 people were obtained, which in accordance with Hoegl(2015) should more challenging for the participants when contemplating communication strategies. The study reported in this chapter involved participants with experience of emergency response, drawn from the Emergency Management and Homeland Security programme at Florida State University (FSU). While it might have been possible to recruit students from Undergraduate programmes (as in the previous chapter), it was felt that the benefit of involving experienced personnel outweighed the demands for large numbers of participants that would have permitted statistical analysis of the results. With hindsight, a larger sample of participants *might* have been desirable but it is felt that the pattern of results produced from this exercise concord with previous work on decision-making in teams (Sorensen et al 2011, Salmon et al 2011, Walker et al 2009b) sufficiently to allow some confidence in the conclusions reached.

The pilot study provided lessons on how a Communications Broker should be presented. Presenting strength of shared information in a task that involves searching for useful information is not necessarily desirable since new information tends to come from weak ties (Granovetter 1973). It is possible that simply showing shared information would only serve to exacerbate common knowledge effects (Gigone 1993) and this is one possible explanation for the behaviour observed in the pilot study, i.e., because the broker was emphasising shared information, participants decided to make sure that they were sharing all of their information (by cutting and pasting situation reports). Additionally, it was felt that the pilot study task was not sufficiently challenging for the participants and, as such, did not present them with a sensemaking task so much as simply an information-pushing task. Using this knowledge a new experimental platform was developed; the Zombie study.

6.2 Experimental Design

6.2.1 Group structure

The experimental platform was designed to allow for two types of organisation; a C2 network and an edge network. In the C2 network there was a central (commander) node connecting with a number of other nodes (Figure 51). This configuration was used to represent a classic model of a network for commanding activity. This network is reliant on passing salient information to a commander to process and then distribute situational awareness to connected nodes. The commander might also choose to pass on raw information back to other participants in the group. The edge network can be seen in Figure 52 and shows a network of equally interconnected nodes. The communication links in this network are not constrained by the structure and as such should allow for the free dissemination of information to all edges in this network. In comparison to the C2 network there will be effort expended in establishing where information needs to go. This is discussed in greater detail in chapter 4. It is hypothesised that edge networks have a greater ability to behave as a network of exploration as nodes in the network are all equal and able to search for information. This flat organisation disperses decision-making powers to all nodes in the network. The hierarchical condition by comparison will be reliant on a central decision maker being fed information from the subordinate nodes of the network.

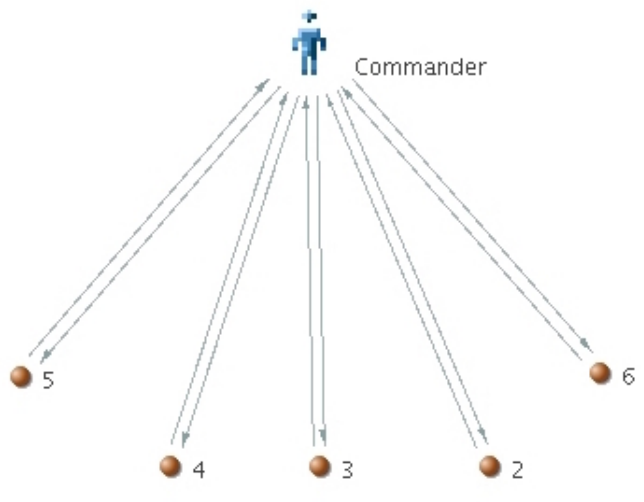

FIGURE 51 - C2 NETWORK EXAMPLE

This study combines elements from the Article Commentary study (chapter 3), the Noisy Map pilot study (chapter 5) and the ELICIT study (chapter 4). Mimicking the latter study, the experimental platform presented in this chapter sought to test and compare the edge and C2 organisations for a different and possibly more challenging task than the ELICIT experiment. The experiments differ in that the Zombie study involves sharing free text (like the Article Commentary), instead of fixed information (factoid) sharing. The removal of information sharing restrictions could cause the C2 networks to perform better in this experiment compared with the ELICIT study: If the central node is able to process incoming information and distribute a complete sit-rep report without becoming overloaded with information, it might be possible for the hierarchical groups to our perform the edge groups. However, good group performance is dependant on a single node in the network doing all of the sensemaking. Participants in the edge group by comparison might be hindered by the ability to share free text with one another. Firstly it is hypothesised that the members of the edge network will lose some amount of independence and be influenced too heavily by other nodes. In the ELICIT study (chapter 4) it was concluded that independent search and analysis in edge networks is a factor in their performance. Secondly there is no restriction on the number of messages that each node can send which may lead to nodes in this configuration becoming overwhelmed with information requests.

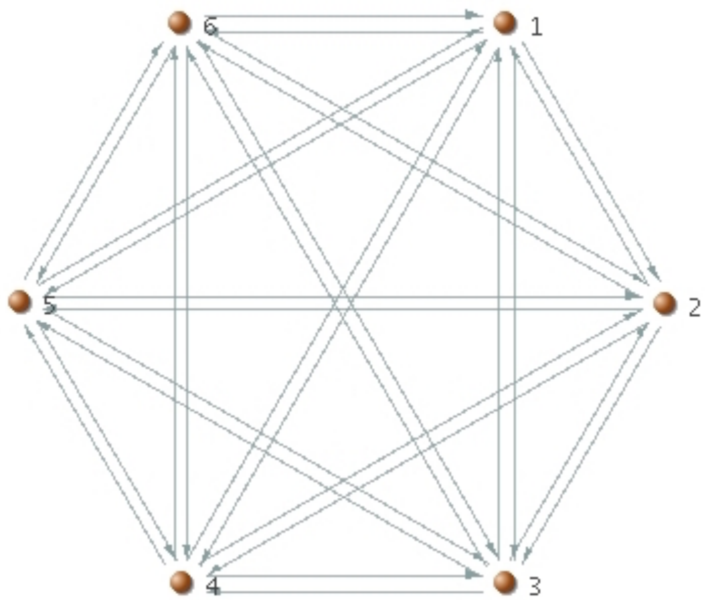

FIGURE 52 - EDGE NETWORK EXAMPLE

In one experimental condition, an edge network would have access to the Communications Broker (detailed in section 6.2.6), which suggests map locations that other users are talking about. The edge group will be used as a control group against which the edge group with the Communication Broker can be compared. The participants in the edge group with the Communications Broker should be directed by the broker to find information about relevant map locations. The broker in this sense encourages independent search and analysis of group knowledge. It was envisaged this would lead this broker group to perform better at the task than the standard edge group.

6.2.2 Task

In the Zombie study the participants are faced with a scenario where a number of zombies have appeared in an area. Zombies are a prominent theme in emergency management and initiatives such as the 'Zombie Preparedness Campaign' from the Center for Disease Control and Prevention (CDC website) are a way of engaging the public on issues of disaster preparedness. The theory is that if you're ready for a zombie apocalypse you are prepared for real disasters such as hurricanes and tornadoes.

In the author's experiment, the zombies posed 4 possible threat levels; high, medium, low and unknown. It was the responsibility of the participants to identify the precise location and threat level of each zombie presented in the scenario. This built on the scenario of the pilot study but increased the difficulty of the original task by altering the location of the zombies over time. Like the pilot study each participant started with a certain amount of information on a map given to them and initially the task was to communicate with other participants in their group to resolve missing information on their map. Each trial was separated into different time periods; t1, t2 and t3 each time period was 7 minutes. This is considerably less time than the 25-30 minutes the participants had to achieve the pilot study. The aim of the platform was to create an unstable environment for the participants where the situation was rapidly changing and they needed to communicate quickly and efficiently to establish the situation. At the end of each time period the experiment was paused and before it could restart, each participant had to answer 2 questions:

- Which building do you think is the MOST safe to be in currently?
- Which building do you think is the LEAST safe to be in currently?

The questions were asked to ascertain if each participant could correctly identify safe and risky areas on the map given their current situational awareness. These responses are used to calculate a collective performance score the group (section 6.3).

6.2.3 Platform

The participant's view of the experimental platform is shown in Figure 53. It was hosted on a webserver that which allowed participants to access it from any personal computing device with an internet connection and a web browser. This platform contains a number of technological enhancements from the platform designed for the pilot study and the constituent parts are details below:

6.2.4 Map

In the pilot study a paper map was distributed to participants, which acted as the noisy initial communication received by responders. The information was different on each distributed map and participants were able to update the map by hand as information became available to them. On this new platform, an interactive software map has superseded the paper version. The map is powered through an application program interface (API) provided by Google (Google Maps API v3) and can be centred anywhere on Earth. This allows for the quick creation of different scenarios i.e. it is possible to shift the map-centre to a different position and plan a new scenario bespoke to this location.

The map is used in this experiment as a replacement for the paper map but also nullifies the need for the map summary box from the pilot study. The processes for editing the representation of individual situational awareness has been streamlined by affording participants the ability to drag and drop icons that represent threats to the map. The map exists embedded in a divider within the HTML of a webpage and API requests to the Google Maps server return map tiles to display within this divider. The API doesn't support the functionality that allows objects external to the map to be dragged onto it. This problem was resolved by using the jQuery JavaScript library to make customised 'draggable' objects that could easily be added and removed from the situation map. This provides the participants with a more ergonomic means of updating their situation awareness, which was important because the scenario is more complex and time restrictions are greater. The creation of this map system eliminated any amount of time that was required in the pilot study to write map summaries and acts as a representation of a participant's individual situational awareness at any given time. The four objects that could be dropped on to the map were high, medium and low threat-level icons and also an 'unknown' pin. This latter pin was used to mark an area of risk where the threat level had not been identified.

6.2.5 Chat System

The chat system on the Zombie platform was also updated from the rudimentary system that was used in the pilot study. The chat system appropriated code written by Garg (2009) and was integrated into the experimental platform. Garg's code embeds a chat system onto a HTML page using jQuery and PHP scripts. Messages sent between users were stored in a table in a MYSQL database, which was used to inform the Communications Broker (section 6.2.6). The chat system can be seen on the right of Figure 53 under the title 'Communications Links'. The condition shown in Figure 53 is for the edge group with Communications Broker. For the normal edge condition a list of buttons exists without the list of locations and participants under the title 'Talking about". When a participant clicks a button it opens a chat window with the participant whose ID was on the button. Chat windows can also be easily minimised and closed. The chat system was updated to resemble systems in use on modern social network sites such as Facebook and Google Mail so they would be more familiar to the participants (students). It is also more user-friendly in general than the system used in the pilot study. The Zombie platform puts the participants in a scenario that is unstable and where task work needs to be completed in a short period of time. As such, providing this familiar system minimises the probability that performance would be hindered by unfamiliar or un-tested technology.

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.

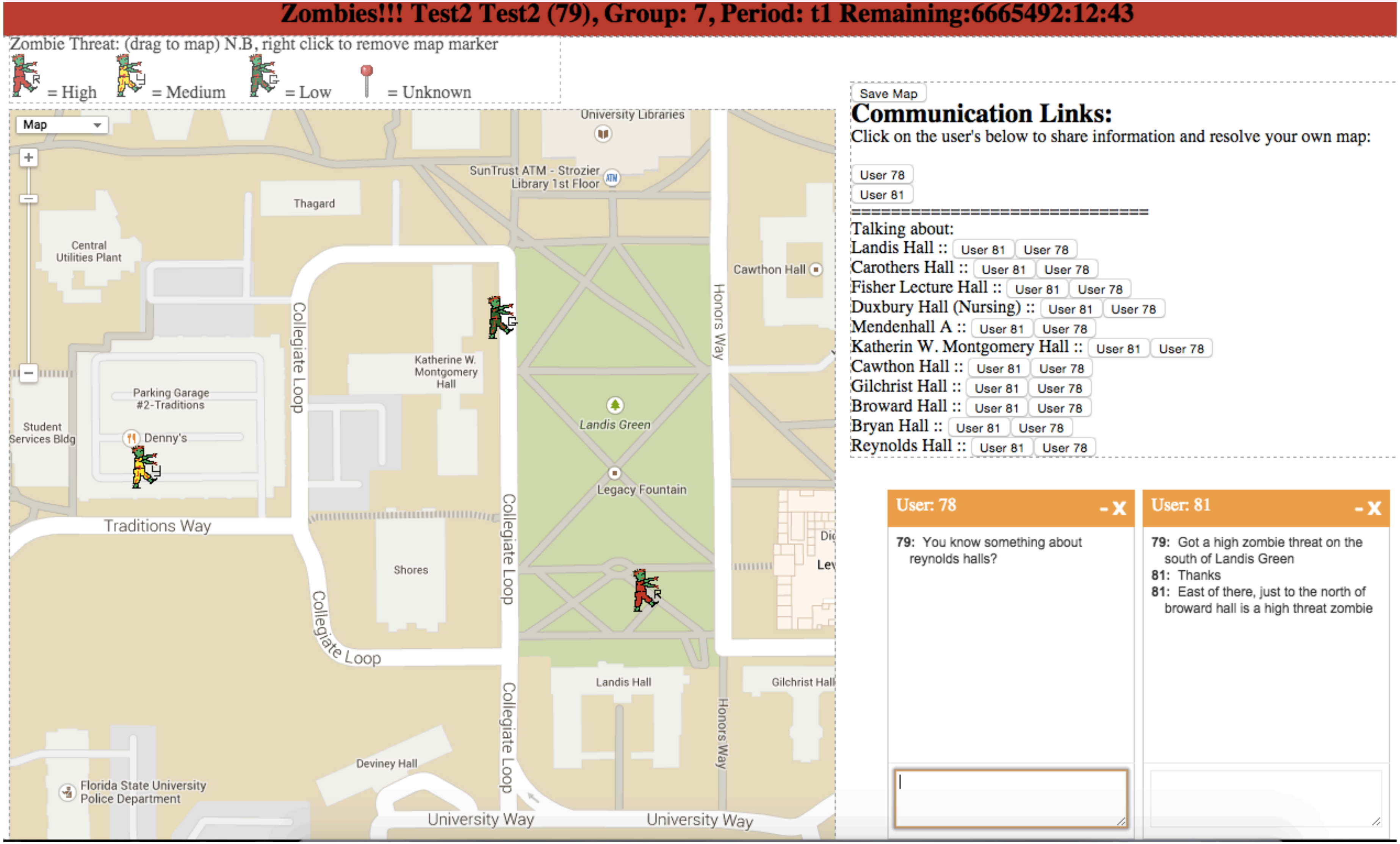

Figure 53 - Zombie experimental platform

6.2.6 Communications Broker

The experiment ran under 3 conditions; C2 network, Edge network and Edge network with the Communications Broker. In the pilot study (chapter 5), the communications broker presented a graph to participants showing the strength of the information they shared with other members of their group. It was realised that this had limited use and encouraged undesirable 'broadcasting' behaviour. Thus, in a task that requires participants to explore an environment, it was felt that it would be more useful to highlight groups of participants talking about certain topics. In this case the topics have been limited to locations within the scenario area (i.e. FSU campus). The original aim in creating the Communications Broker was to help networks of exploration reduce the overheads they have in establishing some amount of consensus on the meaning of the environment they are facing. As discussed in chapter 3, consensus does not have to be unanimous, but it has to be enough that the conversation can move forward. So, the Communications Broker's role is to encourage this by aiding in the pragmatic coordination of the network. As such, the broker was altered from the pilot study to show users who were talking about certain locations.

No changes were made to the underlying text analysis techniques that are detailed in chapter 3. The broker pulls out key terms by continuously monitoring the stream of communications passing within a group of participants using the chat system detailed above. The difference in this platform is that the key terms were used as search terms to query a database table of locations on the FSU campus. The table campus locations was created by attaining a complete list of campus building from the FSU website (FSU Building Directory). The Communications Broker takes any locations retrieved from this query and adds them as an indexed location to the appropriate record in the chat table in the database. An example of this is given below (Figure 54):

id	group_id	session_id	time_period	from	to	message	sent	recd	date	indexed_locations
12	1	6	t1	29	12	Any thing near woodward?	2013-04-09 00:31:39	1	1365463899	a:1:{i:0;s:19:"Woodward Ave Garage";}

FIGURE 54 - EXAMPLE RECORD FROM CHAT TABLE IN DATABASE

In the above example, participant 29 has sent a chat message to participant 12. The message reads; "Any thing near woodward". The Broker has matched 'woodward' as a key term and queried the database table of locations and found 'Woodward Ave Garage'. The broker then adds this to the chat record as an indexed location. On the participant view of the platform, the broker maintains a list of locations and the participants who are talking about them. This can be seen in Figure 53 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 imaged that participants would refer to locations in the scenario by using familiar or shortened versions of the building name. This is seen in the example above where the student simply refers to the 'Woodward Ave Garage' as "woodward". This helps the broker establish a common lexicon for referring to locations that is common by 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. It is thought this would give the participant better awareness of who knew what in the network, effectively maintaining a Transactive Memory Network on behalf of the participant (see chapter 1). This should also help support independent searching of information which was deemed to be important from the ELICIT trials (chapter 4).

Considerations are also made for theme decay (section 2.8.1) in the Broker system. Themes in the network will become less salient as the experiment moves through the different time periods. There are three time periods for the experiment; t1, t2 and t3. At the end of each time period the experiment is paused and while two questions are asked about safe and dangerous locations. The scenario changes at the beginning of time-periods 2 and 3; i.e. several of the zombie threats on the map move location and/or change. As such, information presented to participants in the previous time period becomes redundant. The broker therefore 'forgets' the previous time period's communications as the time period advances and bases its analysis on communications from the current time period only.

6.2.7 The Trial

The trial took place at Florida State University (FSU) using participants from the 'Disaster Systems' course, which is a core module of the Emergency Management and Homeland Security certificate. The author was granted a short period of time to run the Zombie experiment using the classroom base students, of which there are approximately 25 each year, as participants. Students taking this course have good appreciation of risk, situational awareness, disaster communications and the importance of map based planning. This differs from the pilot study where participants were from differing disciplines within a school of engineering and had no relevant experience in either disaster management or map-based planning. As such, it is reasonable to suggest that the participants in the Zombie study might approach this task more like real-world emergency service personnel. The main campus at FSU was used as the area for the scenario in the trial (Figure 55). This area is geographically large enough to stage a scenario with multiple hazards and points of safety. Also it acted as a method to engage the participants in the task due to their familiarity with the campus.

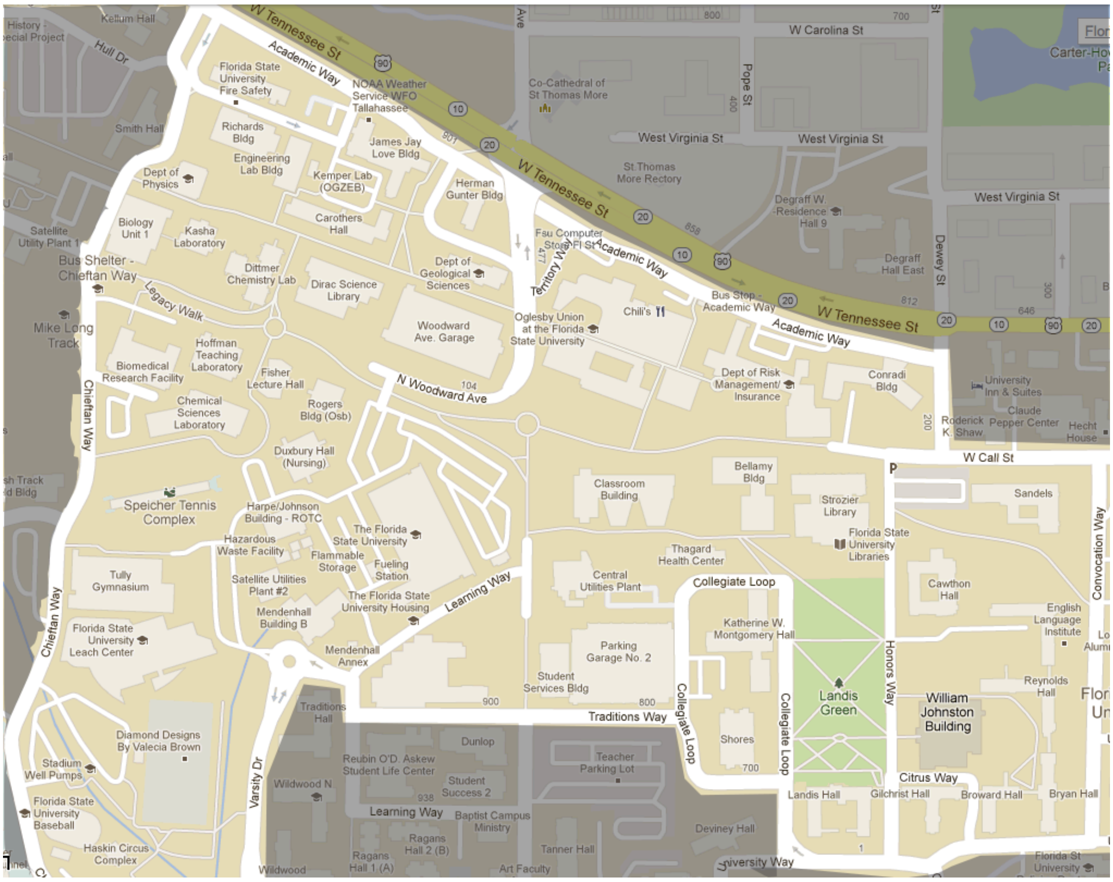

Figure 55 - Area used for Trial Scenario

The participants from the trial were assigned on presentation into one of three groups. Each group operated under a different condition:

Group 1: C2 network.

Group 2: Edge network.

Group 3: Edge network with communications broker.

The participants accessed the system by logging in with an email address. Upon log in, an administrator was able to assign the participants into one of the three groups listed above. Once the groups have been assigned the participants are kept on a holding page until the administrator starts the trial. Upon the trial starting, the information of the threats from the scenario is broken up and dealt amongst the participants in each group.

This is done in the same way that cards are dealt. For example, there were 13 zombie threats in the trial scenario and eight participants in the edge group. The threats were dealt to each member until there were none remaining. This means that five members of the group of eight would have two zombie threats identified in full on their maps at the beginning of the trial and three members would have one each. To ensure fairness, each participant was also dealt the starting location of each threat on the map as an unknown. In this way, no single group might start the experiment with an advantage.

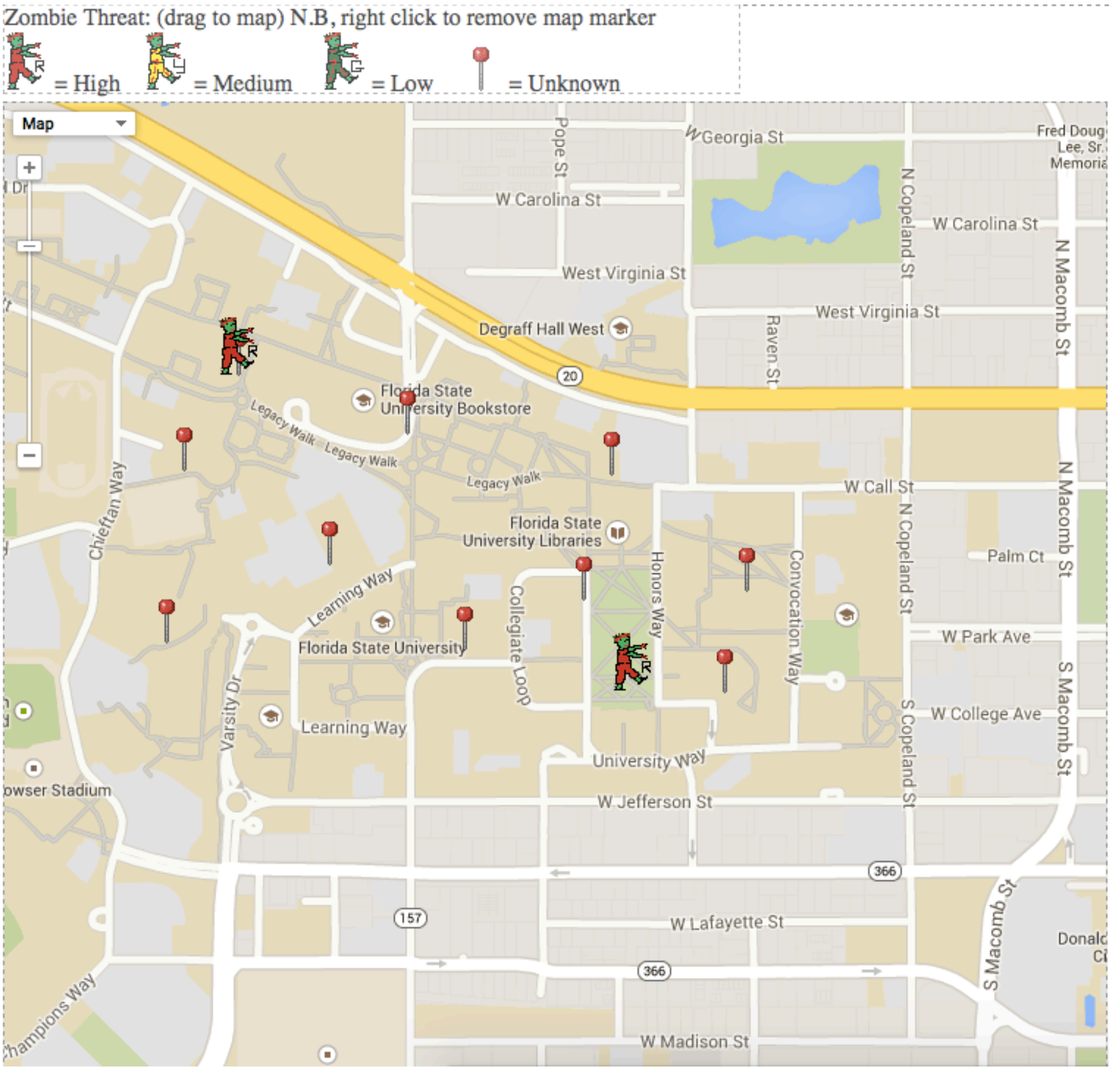

FIGURE 56 - EXAMPLE STARTING MAP OF PARTICIPANT

The scenario was planned in advance of the trial and contains all information of the location and level of zombie threat through all time periods. The trial has three time periods; t1, t2, and t3. At the beginning of time period's t2 and t3, some zombies change location and threat level. A record is kept of the zombie threats that were fully disclosed to individual participants at the beginning of the trial. When the scenario changes at t2 only the zombies that were originally given to a participant change place. The participant is then uncertain if any of the unknown threats that were resolved in the previous time period are still current.

This creates a very challenging environment for the participants. During t1 the participants know at the least the locations of all the threats but, with the exception of the threat(s) that were dealt to them, do not know what threat level they are. Therefore at t1, the task of the participant is to resolve information about the unknown markers. As the experiment moves into t2 and t3 task work becomes more challenging as any information about threats that was previously resolved may now have become old. At the end of each time period the experiment was paused to gauge the participants' perception of current risks. Each participant was individually asked to select from a list of locations from the FSU campus, what building they perceived to be the safest and most dangerous to be in given their current situational awareness (section 6.2.2).

6.3 Performance Criteria

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

1 Which building do you think is the MOST safe to be in currently?
2 Which building do you think is the LEAST safe to be in currently?

The participants selected one answer from a list of every building within the FSU campus area. This method of assessing situational awareness (SA) is adopted from the Situational Awareness Global Assessment Technique (SAGAT)(Endsley, 1988) where participants are polled to ascertain key SA probes periodically during a task.

The answers given by the participants were layered onto a map. 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 a safe area (Figure 57). Locations with zombie threats were represented as a risk heat-map layer shown on the map below as red, yellow and green circles. 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.

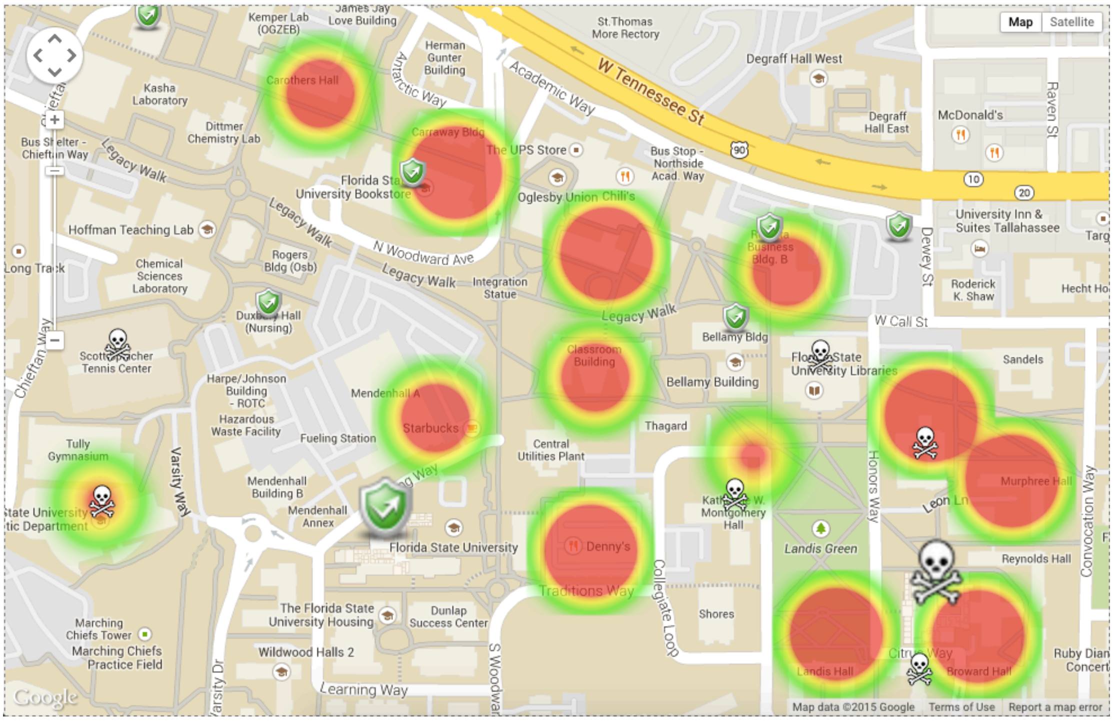

FIGURE 57 - EXAMPLE OF REPORTED RISK LOCATIONS

In performing the task, individual participants are trying to identify areas of risk identified from the trial scenario. This information is distributed amongst a given participant's group but may not be realised in the individual's 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. The total number of responses is equal to the number of people in each group. Signal detection is reviewed in chapter 2 (see section 2.9.5). For the Zombie study the criteria for hits, misses, correct rejections and false alarms is described as follows (summarised in Figure 58):

- Hit = Correctly identifying an area of risk.
- Miss = Not correctly identifying an area of risk.
- Correct Rejection = Correctly identifying an area of safety.
- False Alarm = Not correctly identifying an area of safety.

	RISK AREA	$\overline{RISK\ AREA}$
Skull	HIT	MISS
Shield	FALSE ALARM	CORRECT REJECTION

FIGURE 58 - SIGNAL DETECTION MATRIX FOR ZOMBIE STUDY

6.4 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 most 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. Therefore after the initial period the expectation was that most the groups would perform well assuming they didn't underestimate the possible level of threat at a location. 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.

In the tables presented below. 'N' is the number of participants in that group. 'FA' is the number of false alarms and 'CR'. Hit rate is a function of the number of 'hits' and the number of responses (which is the same as 'N').

6.4.1 After t1

TABLE 10 – T1 PERFORMANCE RESULTS

	N	HITS	MISS	FA	CR	Hit Rate	FA Rate	d'	TPC
Group 1	9	4	5	2	7	0.44	0.22	0.62	0.61
Group 2	8	3	5	3	5	0.38	0.38	0.00	0.50
Group 3	7	4	3	2	5	0.57	0.29	0.75	0.64

TABLE 11 – T1 NETWORK STATISTICS

	Condition	N	Avg Comms	Avg. Avtivity	Density
Group 1	C2	9	3.11	17.8	0.19
Group 2	EDGE	8	3.50	13.9	0.34
Group 3	EDGE w/ CB	7	7.29	10.6	0.50

Table 10 shows that group 3 operating as an edge network with the aid of the communications broker has the highest sensitivity when detecting risk in the scenario. Group 1, performing as a C2 network, has slightly lower sensitivity but its TPC score is only 3% lower than that of group 3. This is a result of strong score in correct rejections coinciding with a high miss-rate. Group 2 has neutral sensitivity in that its false alarm rate is equal to its hit rate. This translates to a TPC score of 0.5%.

Group 3's high performance coincides with average communications rates more than double that of groups' 1 and 2. Density for group 3 is the highest, showing it is utilizing 50% of the possible communication links in its network. Contrastingly its average activity per participant is the lowest of the groups. Group 1 is approaching its' highest possible density score as it quickly utilises all the communications links in the C2 structure. 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:

$$Density_{Max} = \frac{Possible\ C2\ Links}{N(N-1)} = \frac{8 \times 2}{9(9-1)} = 0.222$$

6.4.2 After t3

Results after the third period (Table 12) show group 3's sensitivity had declined to be the worst of the three groups. TPC had fallen also mostly because of a rise in false-alarm rate. Group 3's results also show a fall in **d'** which is a reflection of high miss rate. The TPC is stronger than that of group 1 which is mainly due to a high correct-rejection rate. Group2 shows an improved sensitivity and TPC and is the best performing group.

TABLE 12 - T3 PERFORMANCE RESULTS

	N	HITS	MISS	FA	CR	Hit Rate	FA Rate	d'	TPC
Group 1	9	1	8	3	6	0.11	0.33	-0.79	0.39
Group 2	8	5	3	3	5	0.63	0.38	0.64	0.63
Group 3	7	3	4	6	1	0.43	0.86	-1.25	0.29

TABLE 13 - T3 NETWORK STATISTICS

	Condition	N	Avg Comms	Avg. Avtivity	Density
Group 1	C2	9	8.44	34.4	0.21
Group 2	EDGE	8	9.88	25.4	0.77
Group 3	EDGE w/ CB	7	24.86	19.4	0.71

Table 13 shows the network statistics for each group following the third period. Group 1 has achieved highest possible density by exhausting the possible communication links within the C2 organisation. Groups 2 and 3 have similar density scores showing they have utilised 77% and 71% of network capacity respectively. Average communications per participant in group 3 is more than double that of groups 1 and 2. As with results after t1, the group (group 3) that has the highest communication rate also has the lowest activity rate. Contrariwise, the group with the lowest communication rate (group 1) also has the highest activity rate.

The tables below are samples of the messages sent between participants during the trial The chat logs for relevant time periods can be seen in Appendix 3. Table 14 shows the C2 group commander dictating, what he or she considers to be, the safest building in the scenario area currently.

Table 14 - Sample Communications from Group 1 t3

Time	From	To	Message
12:53:13	12	15	tully good
12:53:20	12	29	tully good
12:53:56	12	35	tullys good
12:54:07	29	12	Tully's good.
12:55:46	12	46	go to tully
12:55:58	12	35	none are at tully

Table 15 shows participants from group 3 (edge group with broker) 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 15 - Communications Sample Group 3 t3

Time	From	To	Message
12:56:46	26	34	Who has a zombie at Dittmer Lab
12:56:47	26	6	Who has a zombie at Dittmer Lab
12:56:49	26	16	Who has a zombie at Dittmer Lab
12:56:51	26	17	Who has a zombie at Dittmer Lab
12:56:53	26	76	Who has a zombie at Dittmer Lab
12:56:56	26	45	Who has a zombie at Dittmer Lab
12:56:56	34	17	threat level of leach zombie?
12:57:03	34	45	threat level of leach zombie?
12:57:06	34	16	threat level of leach zombie?
12:57:08	34	76	threat level of leach zombie?
12:57:11	34	26	threat level of leach zombie?'

TABLE 16 - COMMUNICATIONS SAMPLE GROUP 2 T3

Time	From	To	Message
12:46:15	11	9	medium zombie threat mendenhall a
12:46:29	47	37	HIgh level at leach
12:46:31	9	37	high alert zombie at landis
12:46:36	11	31	medium zombie threat mendenhall A
12:46:36	40	37	Medium threat zombie at Montgomery
12:46:44	40	47	Medium threat zombie at Montgomery
12:46:46	47	37	low level at katherine montgomery hall
12:46:46	40	11	Medium threat zombie at Montgomery
12:46:48	40	9	Medium threat zombie at Montgomery
12:46:49	37	40	med threat at woodward parking
12:46:51	40	31	Medium threat zombie at Montgomery
12:46:55	4	40	I have a med level of walkers in the union
12:46:57	47	40	high at leach and low at montgomery hall

Table 16 shows a sample of messages from group 2 (standard edge group). The group members in the time period t3 appear to be sending zombie threat information arbitrarily to other members in the group but show less of the mass broadcasting behaviour of group 2. Arguably where the same message has been sent several times to multiple users, these messages could be codified and grouped and treated as a single occurrences in the analysis. However, the pragmatic sensemaking aspect of this investigate is concerned with the communication strategy of the participants in each group. As such, the emphasis is less on the content of message but where the message is being sent. Thus, codification of messages was not considered.

6.5 Discussion

It was hypothesised that the C2 group in this experiment could out perform the standard edge group but this was dependant on the capability of the participant acting as the commander to process incoming information and distribute situational awareness. The standard edge group, it was envisaged, may struggle due to a lack of support for independent search and information analysis that was seen to be successful for edge groups in the ELICIT study (chapter 4). It was anticipated that the edge group with the Communication Broker would show a performance score better than the other two groups because of the support for directed searching of threat information.

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, the majority of participants in group 1, can only communicate with the group member who is acting as the commander in the network.

Naturally 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 59 shows the state of group 1's social 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.

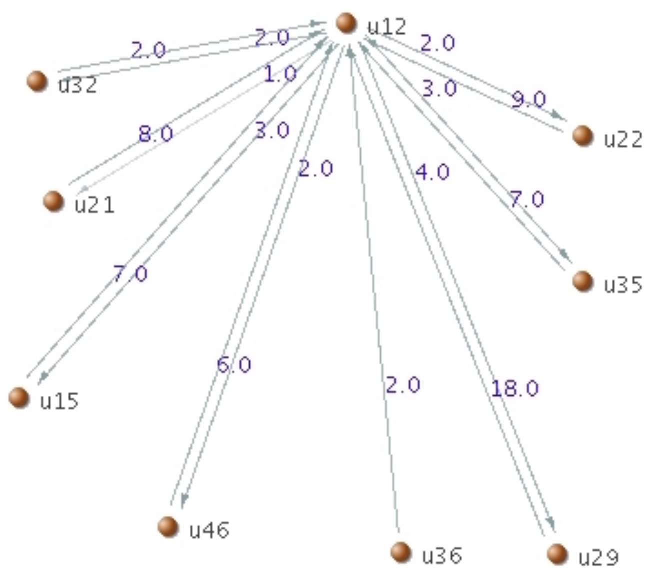

FIGURE 59 - GROUP 1 SOCIAL NETWORK AFTER T3

The bottle-necking of situation information is the probable cause for the poor performance of group 1 by the end of time period t3. Group 1 maintains the highest activity level throughout the trial however there is little evidence to suggest why this is the case. 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 his/her's situation map, uploading it to an image hosting website and then sharing the link to this image with the group commander (Figure 60). Fortunately, the commanding node did not pass on this method of spreading situational awareness to the other participants and as such the rest of the network did not adopt the behaviour.

Towards the end of the third period, the commander in group 1 adopted an approach by which he/she stopped passing on situation information to other participants and instead broadcast what they believed to be the area of lowest risk (Table 14). This approach was hypothesised as a way that group 1 might obtain a strong performance score.


From: 22 => To: 12
Time: 00:51:03
screenshot of my map: http://i.imgur.com/VgCX8lE.png

FIGURE 60 - EXAMPLE OF MAVERICK/INNOVATIVE PARTICIPANT

Group 3 has the highest levels of communications of all groups. The communications broker seems to encourage communication within the group. This coincides with the lowest activity rate. In this group there seems to be an emphasis on teamwork and information sharing at the cost of carrying out the task work. The cost of this is that the group has a lower sensitivity to risk. The communications broker appears to have had a detrimental affect 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. Some evidence to support this theory is shown in Table 15. The multiple enquiry approach seen here might be more expected of group 2 where no communications broker is available. However, group 2 maintain 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. Group 2 participants, on the whole, send small snippets of factual information rather than ask questions about locations (Table 16) i.e. they are pushing information around the network.

In summary, the three conditions in the experiment provoke three different behaviours and strategies from the corresponding groups taking part. Group 1 has a C2 structure, which restricts communications and isolates group members. The result is they attempt to find more meaningful ways to communicate or tinker with their maps. Group 2 operates by sending factoids of information about the situation without a great deal of searching and question asking occurring. The communications broker in group 3 encourages participants to look for information rather than just share the information they have on their map.

7 Conclusions

7.1 Introduction

Collaborative sensemaking is critical to the effective management of disasters. Teams of responders need to be able to understand what situation is facing them and to coordinate their response effectively and efficiently. This thesis has developed the theoretical concept of collaborative sensemaking for this domain. Although the main focus has been on emergency response and disaster management, the concept has much wider applicability and can be applied in any domain where groups of people are making sense of unknowns. It is associated with the multidisciplinary study of cognitive engineering and it is important in the development of modern socio-technical systems. This study set out to ascertain whether collaborative sensemaking could be measured within the dynamics of social networks and if technology could play a role in aiding the process.

The thesis began with a review of the current literature on sensemaking and related concepts. From the review two working definitions of sensemaking were extracted. Firstly, that on one hand, sensemaking is an abstract, meta-cognitive process that is concerned with thinking about what needs to be thought about. This is defined as *semantic sensemaking*. In this respect, 'sense' is analogous to meaning and sensemaking is a matter of interpreting complex, ambiguous and dynamic information to define the meaning of a situation. For this thesis, that situation is primarily applied to disaster management. Secondly, the process concerned with establishing the constraints under which action can take place is defined as *pragmatic sensemaking* (section 2.5.2). In this case, the action concerns not only how best to respond to the situation but also how best to manage the sharing of information between people. This distinction between semantic and pragmatic sensemaking is similar to the way Clark (1988) defines 'common ground' as both the shared meaning that people need in order to have a conversation and the ways in which people indicate that they have understood the meaning sufficiently to allow the conversation to proceed. This notion of common ground is important to the development of collaborative sensemaking because it underpins the arguments that (a) 'sense' between people does not have to be complete for there to be an agreement on an appropriate response and (b) the process through which sense is made can be as important as the outcome of this process.

In its purest form semantic sensemaking is concerned solely with attribution of meaning to data in an environment. In order to define a plausible response to this meaning (in terms of seeking further information, reaching consensus, acting on the situation etc.), pragmatic sensemaking is engaged. Sensemaking, as such, operates as a joint optimisation between the semantic and pragmatic modes where either mode affects the other and both modes are always operating side by side but not always in equilibrium. The benefit of separating sensemaking into these two modes is that it acknowledges that collaborative sensemaking has both process and performance aspects. For example, a group of marketing strategists discussing a new market opportunity will perform sensemaking differently from a group of first responders to a disaster. The constraints on the organisations are different in both scenarios and being able to view these as pragmatic considerations allows for the development of systems that could aid the overall sensemaking process. The following research questions were presented at the beginning of this thesis:

Q2. Can collaborative sensemaking be measured in the behaviour of social networks?
- Q2.1. Which social network organisations favour collaborative sensemaking?
- *Q2.2.* Are there social network analysis (SNA) metrics that reflect this organisation?
- Q2.3. Is it possible to develop a metric for collaborative sensemaking?

Q3. In what ways can technology aid in the process of collaborative sensemaking?
- *Q3.1.* How might the behaviour of a social system be modified with technology?
- *Q3.2.* How might we indicate the sensemaking activity of the network to members of that network?

Analysing the cognitive processes of groups of people is inherently difficult and research tends to rely on monitoring the observable behaviours and actions that are outputs of these processes. As an alternative approach, this research sought to answer the research questions above by developing experimental platforms, utilising network analysis tools and creating performance metrics for groups engaged in sensemaking tasks.

7.2 Empirical Findings

The Article Commentary experiment in chapter 3 looks at the processes by which people attempt to gain a shared meaning of some ground truth (in this case, the content of a news article). From this experiment, correlation statistics showed a tendency for participants to align the language (key terms) in their interactions with the language used in the article as opposed to the language introduced by other participants. Relating back to Clark & Brennan (1991), this shows measurably that the participants are forming common ground by aligning their contributions with the article content. The inference is that key terms can represent themes (Vilhena et al, 2014) and that common ground is then a resolution of key themes between participants. From a semantic sensemaking standpoint, one might logical, one might conclude that consensus on meaning could also be measured through the alignment of key terms. However, given that language is made up from complicated constructs, simply using the same key terms does not necessarily indicate that the same meaning is being produced. There is a suggestion though that the collective filtering of key terms (themes) for use in discussion is a semantic sensemaking process because it implies the agreement that certain terms are important with respect to the content.

This was not a task driven exercise and no requirements existed for the formation of a group, or groups, to achieve a goal. However, ratios of replies to comments (Figure 21) in conjunction with the number of interactions per article suggested that, on some articles, discussions were taking place. The emergence of these discussions correlated with an increase in participant language-use that was divergent from the language use in the article. Since the number of interactions was a function of the number of people commenting on an article, it suggested that group size inhibits the processes of collective sensemaking. In accordance with the concept of Groupthink (Janis 1971), it was concluded that the overhead of introducing new opinion to a network of people increases with the number of people in the network. This potentially explains the increased preference of participants to reply to an existing comment rather than introduce a new comment; the overheads of being novel in the subgroup are less than that of the network as a whole.

In terms of Q1.1, this suggests that large networks will naturally organise into smaller subgroups when confronted with unknowns and when the constraints on the network allow it. Not only does this seem to be related to the issue of semantic sensemaking (in terms of managing meaning with the potential overhead of contributing new information) but also to pragmatic sensemaking (in terms of keeping track of the opinions of multiple group members). The preferable sizes of these subgroups are unclear. The discussion at the beginning of chapter 4 surmises it is less than 40, which agrees with work done by Dunbar (1998). Dunbar makes observations of human social group sizes and presents findings that group sizes have a tendency to cluster around the values {5, 12, 35, 150, 500 and 2000}. Dunbar's (1998) argument is that there is an optimal number, of 150, around which human social groups appear to be saturated, i.e., the Dunbar number. The lower group sizes were defined in terms of kinship or family groups, suggesting that the number of people who form a 'close' group seems to peak between 12 and 35. This might, for instance, be a matter of managing social capital or trust within a group. Although the author does not speculate further in this thesis on how or why groups seem to form at such well-defined sizes, in chapter 4 it is suggested that a group size of between approximately 12 and 40 people seems most appropriate to collaborative sensemaking, which supports the findings of Dunbar (1998). At the lower end of this scale, the Noisy Map pilot study in the chapter 5 considered group sizes of five or less to be too small for the study of collaborative sensemaking behaviours. Groups of this size, in the scenarios tested, were affected less by problems such as information overload but suffered from common knowledge effects. The above has implications on the design of technologies to aid collaborative sensemaking and also shows the benefit of seperating the notion into two modes. Aiding the pragmatic side of sensemaking could lead to the development of technologies, such as the Communications Broker that would allow us to harness the sensemaking potential of larger groups by managing social overheads.

Considering Q1.1 further, the ELICIT study data (chapter 4) was employed to test if different types of organisation favoured collaborative sensemaking. The task in the ELICIT study, from a semantic sensemaking viewpoint, is an individual endeavour. The participants are collating snippets of information ('factoids') to fit into the framework of an evolving hypothesis. At intervals in the experiment, each participant is requested to submit a credible working hypothesis. Organisational constraints dictate the pragmatics of how the information passes through the network and where information is sourced. The analysis of the ELICIT study in this thesis concludes that the collective sensemaking capabilities of the edge (fully connected) networks have better performance at the task than the hierarchical networks. This reaffirms the findings of (Stanton et al, 2012) and supports similar notions presented by Baber et al (2008) on networks of exploration.

It is concluded in Chapter 4 that the performance of the edge networks is better because of the participants' ability to maintain a degree of independence, focus on the task and access information freely. This prevented the participants influencing one another's frames for information collection. Edge network members were able to focus on the intelligence analysis task in the experiment because they were unconstrained in their ability to access the available information. Contrastingly, the hierarchical groups were reliant on other participants within the network to provide information to them as they only have partial access to the information in the experiment. As such, the behaviours of the individuals in the two networks are different. Behaviour in the hierarchical groups is directed at establishing distribution of information without emphasis on generating 'correct' solutions. In other words, the hierarchical networks appear to be driven by pragmatic sensemaking (i.e., applying methods to ensure appropriate sharing of information). The edge network by comparison is more focused on the task and passes factoids, which support a hypothesis. In other words, the edge networks appear to be driven by semantic sensemaking (i.e., looking for meaning in the information).

These findings are surprising because although it was hypothesised that edge networks would perform better at sensemaking tasks, it was thought this would be due to the unrestrained capability of edge members to share information. It was supposed that this would lead to more communication and sharing in the edge groups whereas the opposite was actually observed. The challenge for the edge networks, it was hypothesised, would be to do with pragmatic sensemaking i.e. assessing where to send the appropriate information. However, it appears in this instance this was less of a concern for the edge networks; perhaps because they had other means of obtaining information and were able to work more independently on the task. This independence leads to the edge network, as a system, filtering out noise in the problem space by predominantly passing the factoids that support the correct hypothesis. This filtering property of the networks was suggested by the 'S' shape nature of the median group-performance score over time (chapter 4, Appendix 2). The observation that the 'correct' hypothesis appears to diffuse through the edge network, when there is not any direct method of influencing neighbouring nodes, suggests that the participants are influencing one another by passing evidence that supports the correct hypothesis. This is analogous to concepts such as wisdom of crowds (Surowiecki 2004) and also artefact driven sensemaking (McMaster et al, 2012). This collective filtering (or rejection of noise) process was also seen in chapter 3's article commentary experiment and can be seen as a semantic sensemaking process. It also suggests that both semantic and pragmatic sensemaking are driven and influenced by information searching. From the semantic perspective, the question in relation to information searching is, what information helps me develop a model of my current situation? The question from a pragmatic perspective is, who might have information that would inform my model of the situation?

In chapter 4, a finding relating to Q1.2, concerning the manner in which information diffuses through a network, was the relationship between the coefficients of innovation

(p) and imitation (q) in relation to group performance in the ELICIT analysis. Bass (1969) diffusion curves were fitted onto the performance results of the groups. Plotting 'p' and 'q' against performance reveals good group performance with clusters of 'p' values in the region of 0.38 and 'q' values in the region of 3.0. Values of 'p' and 'q' in studies of market diffusion models are generally around 0.3 and 0.3 - 0.5 respectively (Mahajan et al, 1995). The suggestion is that 'p' could be a measure of the innovativeness of the group as a whole: for the ELICIT groups, the value of 'p' might imply that the ability of a group to process information and produce more 'correct' hypotheses. 'q' would then be the rate at which, given the filtering properties of the network, participants are able to imitate this hypothesis given the supporting factoids. This could be a useful metric to have in establishing the collaborative sensemaking capabilities of the group and is suggested as an area for further research.

The behaviours observed in the ELICIT study were found through the use of a bespoke metric developed for assessing the performance of the groups. In addition, other metrics were used to count sharing and information pulling and network density. Returning to Q1 in this thesis, to what extent do these metrics allow the measurement of collaborative sensemaking behaviour? The performance metric defined for the ELICIT study is not a measure of sensemaking per se, but is a measurement of the output of an individual participant's sensemaking process. Aggregating and averaging all participants' scores gave an average for that group. These measurements give a retrospective view of a group's sensemaking capabilities. Applying these measures after the task is completed, allows for the study of various facets of the social system such as network structure, density and information retrieval, in comparison to performance, and is useful for proposing the strengths and weaknesses of various organisations at performing certain tasks. Calculating the average amount of information sharing and pulling gives an insight into the focus of the groups; are they more concerned with sharing information or searching for new information?

Network density shows the connectedness of a group and indicates distribution of information (Walker et al, 2009). Where it is possible to measure the performance of the task, density also gives an indication of efficiency. Network density was used in this thesis as a means of measuring the connectedness of networks and can be interpreted in various ways. Relating to Q1.2, density by itself shows potential communication links being utilised in a network. In conjunction with performance, density can show the efficiency of a network in performing the task in comparison with other networks. It can also indicate network behavioural tendencies towards task-work or teamwork. In the ELICIT study analysis, density was not found to be a conclusive factor in performance when considering all the groups. This was because there were other factors such as access to information that were affecting group performance. However, when analysing the edge and hierarchical groups separately, a positive correlation is observed between density and performance (chapter 4, section 4.4.3). This suggests that density is only useful when comparing like structures and might give an indication as to why a certain network performed better than another. In conclusion, density is versatile network metric but its interpretation is dependant on the context in which it is used. It can be seen that these metrics offer an insight into the behaviour of various networks performing tasks, but what implications do these findings have for real-world scenarios such as disaster management? The introduction of cross-agency compatible radios, as discussed in chapter 5 (section 5.1), to emergency personnel in the UK implies the technology that would allow the communications network of responders to be monitored at a tactical level is becoming pervasive (Figure 41). While responders are trained in cross-agency procedures there could be a benefit in allowing responders to form more fluid teams (akin to edge networks or networks of exploration section 2.6) to deal with certain aspects (themes) during the initial phases of a large-scale response. Network metrics could be used at the tactical level (control room) to monitor and evaluate the way that response teams are forming in real-time; this might alleviate some of the coordination challenges in the early stages of the response. Low density values in the emerging network, for example, may indicate to tactical command that more effort should be invested in the coordination of communications between teams. These notions contribute to a larger design policy that should be considered when developing technologies to aid pragmatic sensemaking in disasters which is important to second research question (Q2) in this thesis.

This thesis approached Q2 from the viewpoint of the working definitions defined for collaborative sensemaking (section 2.5.2). It was proposed that technology that currently supports sensemaking is focused on aiding the semantic processes. Technologies reviewed were predominately concerned with supporting the creation of shared representations (section 2.8). Comparatively, technologies that support pragmatic sensemaking are rare. The benefits of edge networks for doing sensemaking were noted in the case study in Chapter 5 and it was hypothesised that individuals in loosely defined or decentralised groups, could operate as a network of exploration without the pragmatic overheads normally associated with the group. To achieve this, a supporting technology was envisaged that attempts to broker communication within these groups by suggesting links to other participants in the network who had potentially useful information. The Communication Broker design principle was to work as a technological proxy for a transactive memory network (section 2.7). The Broker would keep track of organisational knowledge and broker communications links between network members dependent on information commonality. The Noisy Map pilot study (chapter 5) presented a Communications Broker as part of a collaborative task. The participants in the study developed an unexpected behaviour of broadcasting large amount of information in their network. The feedback from the trial however indicated design flaws in the presentation of the Broker in that it highlighted ties between network members who had information in common. As such, this encouraged common knowledge effect in the network. This makes sense because it is known from Granovetter (1973) that it is actually weak ties that bring more valuable information. A redesigned Communications Broker was used in chapter 6's Zombie study. In this second prototype (relating to Q2.2), following lessons from chapter 4, it was decided to encourage independent search in the members of the edge network by indicating topics (map locations were used as topics) and the respective network members who were talking about them. The Broker seemed to encourage a large amount of communication between members of the network but hah a detrimental effect on performance.

The metrics developed in this thesis suggest that collaborative sensemaking can be measured through the behaviours of social systems in hindsight. The Zombie study in chapter 6 utilises Signal Detection Theory (SDT) as another means of measuring group performance. It analyses the networked group in the task as a receiver that is trying to detect risk in the task. Performance is measured as a group's sensitivity (**d'**) to detecting risk. The task presented in the Zombie study differs from the ELICIT study in that new information is only accessible to the participants through other nodes in their network. As such, higher density scores than that of the ELICIT study were seen. The design of this experiment emphasised aspects of pragmatic sensemaking; the participants needed to establish where information needed to go. The experiment concluded that the edge network, which performs without the aid of the Communications Broker had the highest sensitivity. This reaffirms existing findings that the edge organisational structure is the best at supporting sensemaking although repeated trials would give more confidence to these findings.

These studies suggest that through experimental work the conditions under which groups perform well at sensemaking tasks can be defined. It would be interesting to determine how well these findings can be reproduced for groups in the real world.

Research in this area is expanding and there is a growing understanding that networks which have less centralized power and better distributed decision rights (Stanton et al, 2012) support collaborative sensemaking. Q1.3 alludes to the possibility of developing a metric for collaborative sensemaking. Sensemaking is an on-going process and developing a single metric that reflects the current sensemaking activity of a group or organisation would have limited use due to the high number of factors that affect sensemaking; what would the metric say about the group doing the sensemaking? From this research the thesis suggests that developing a set of key indicators would be a more useful pursuit. At a minimum, the metrics that would feature on a list of key indicators for collaborative sensemaking would include network density, average activity on task, average communications and group size.

7.3 Contribution of Research

The work presented in this thesis represents a contribution to the field of cognitive engineering where understanding the human element in socio-technical systems in the context of tasks, tools and environment is paramount. The concepts of semantic and pragmatic modes of collaborative sensemaking add to the evolution of sensemaking theory and will aid future discussions and research in the field. It is suggested that these definitions fit in the gaps left by the notions of sensemaking provided by Weick(1995), Klein et al (2007) and Pirolli and Card (2005). Whereas Klein et al and Pirolli and Card present models of sensemaking that describe the individual's cognitive and meta-cognitive processes they omit to elaborate on how this might expand beyond the individual. Alternatively, Weick (1995) gives us the notion that sensemaking is a social activity but does not elaborate on the processes by which this happens. This thesis takes the work of Klein et al and Pirolli and Card and extends it to represent the semantic sensemaking considerations of the collective. Similarly, by utilising the social aspect of sensemaking from Weick (1995) and developing a pragmatic notion of sensemaking, the work in this thesis has explored ways to measure this process in groups. While this work is not conclusive it does pave the way for future research in this field. These two notions of sensemaking are interlinked by the concept of common ground presented by Clark & Brennan (1991).

The pragmatic mode of sensemaking also helps to widen discussion on the design of technologies to support sensemaking. Current technologies such as GATOR (section 2.8) rely too heavily on a 'pins on a map' representation of incidents. While these approaches are useful to an extent, it is felt that more effort could be focused on producing representations that show the relationships between information and people operating in dynamic situations.

7.4 *Recommendation for future work*

7.4.1 Network Comparison

Continuation of the ELICIT experiment to compare different edge networks where the controlling factor is the ability of groups to access information is proposed. This would create a requirement for network members to collaborate in sharing information and would be more useful in analysing real-world disaster management scenarios. It would also allow for insights to be developed into the interdependence of group members in collecting information would affect the performance of the sensemaking task. In disaster management agents are more reliant on other agents for obtaining information about the situation. This proposed research would help the understanding of how edge networks evolve structure in response to the demands of the situation.

7.4.2 Advanced Language Techniques

The text analysis techniques described in chapter 3 which are used throughout this thesis for analysing content and retrieving key terms could be enhanced. Originally the text analysis module was designed as a lightweight tool to help investigate the concept that participants would align their language with either the articles presented or other participants. The text analysis module could be enhanced by the addition of a 'synonym matcher'. The synonym matcher would combine key terms extracted from the content with similar meanings and create a unified score for combined key terms. As an example lets suppose an article contains the terms; 'castle', 'fortification', 'dog', 'cat', 'dog'. The frequency array would be:

[castle] => 1;
[fortification] => 1;
[dog] => 2;
[cat] => 1;

The frequency array would be passed through the synonym matcher, which would match 'castle' and 'fortification' as being synonymous, remove the latter and add the frequency score of the latter to the frequency score to the former. The results frequency array would then read:

[castle] => 2;

[dog] => 2;

[cat] => 1;

Entering 'synonym API' into a search engine reveals multiple possible API's that could be used. This could further refine results for chapter 4 by finding stronger correlations between participant and article content. It could also enhance the Communications Broker system; combining key terms with similar meanings could highlight topics within participant communications that might serve as the basis for chat rooms.

7.4.3 Communications Broker Trials

The zombie study was only able to run one trial due to constraints on time and access to participants. It is not possible therefore to have high confidence in the results obtained although some aspects of the results did support conclusions from other previous research (see chapter 6). The experiment requires a minimum of seven people in a group and three groups for each trial. Preferably group size would be larger so as to get a better impression of whether the Communications Broker can enhance the capabilities of the edge groups to find pertinent information in the network. If time allowed, a trial could consist of a single group of seven participants repeating the experiment three times under different conditions. However, a single run through the experiment takes approximately 25-30 minutes excluding initial setup time, which would require participants to be in situ for approximately 1h30.

The Zombie study was an attempt at a proof-of-concept for displaying the ongoing pragmatic sensemaking of a group, back to the same group by highlighting topics that other group members were talking about in their communications. The study also provided sufficient data to allow development of metrics (in this case using SDT) which is a key focus of the work in the thesis. It was established on the basis of these metrics that technology (i.e., the 'Broker') *did* alter the Edge group's behaviour. This suggests that caution is needed before introducing technology that focuses exclusively on managing social network connectivity and that further research would involve an investigation into the best method by which to present this information to the group.

7.4.4 Group Size and Network Evolution.

In the experiments presented in this thesis, group size and organisational constraints have remained the same throughout the experiments. The classic sensemaking problem is that one does not know which people to involve in the problem until it is evident what is going on but paradoxically, one does not know what is going on until some people are involved. So, the Edge network provides a 'network of exploration', which is a kind of rapid prototyping approach involving several people. It is likely that, once the network of exploration has made sense of the problem, the next stage in response would be to initiate Standard Operating Procedures and these are most appropriately managed through more traditional Hierarchical command networks.

The Author believes that Collaborative sensemaking is still possible in larger groups (N>40) but the mechanism by which it is achieved is slightly different from smaller groups. When group size becomes too big the overheads become too great. There is a suggestion that moving towards a wisdom of crowds approach might work better. This is reliant on mechanisms for aggregating individual sensemaking outcomes. This is different to the ELICIT study where good performance was reliant on the independent searching and processing of information and then influencing other participants by sharing supporting factoids.

7.5 Practical Implications

7.5.1 FSU Virtual Operations Support Team

The Author had the good fortune to take a research position at the Center for Disaster Risk Policy (CDRP) at Florida State University (Tallahassee, Florida). The centre maintains a very close working relationship with the Florida Division of Emergency Management (FDEM) allowing opportunities for real-world emergency management research. The Author was able to apply lessons learned from this thesis in developing and managing a Virtual Operations Support Team (VOST) for FDEM.

The concept of VOSTs originated in the United States around 2011 and was in response to growing trends of social media use during disasters. The use of social media in disaster management and emergency response is now become a wide spread practice. Social media generates tens of thousands of data points per hour from a wide variety sources including citizens, media outlets and official government organizations. With such large quantities of data being provided, it is very challenging and time intensive for people in emergency organizations to monitor and filter relevant data to be parsed into useable information during a large scale event. In essence this is a semantic sensemaking challenge. Several attempts have been made to solve this problem. Some Organizations in the USA like the American Red Cross have built dedicated social media operations centres for humanitarian relief. These centres have proved useful in tests run during tornadoes in the Midwest of America where they were able to collect and analyse data in such a way as to inform Red Cross teams where to position workers on the ground. These centres can provide a high level of awareness during a disaster but despite contributions from donations, volunteers and private business partners, involve high operational costs (Kash, 2012).

As such, many response organisations such as FDEM wanted to utilise social media for data retrieval more effectively during an event but could not justify the fiscal outlay for technologies and staff to do so. As a solution CDRP created a virtual operations support team at FSU (FSU.VOST). The principle behind VOST was that it was to act like an on-demand network of exploration for making sense of the "virtual space emergency" (Schneiderman & Preece, 2007). The VOST was predominately made of student volunteers distributed around the Tallahassee region in Florida. Figure 61 shows the organisational chart of FSU.VOST. The VOST was activated from within FDEM when required for an event. Once activated the Team leader would distribute the mission goals to all volunteers and schedule shifts.

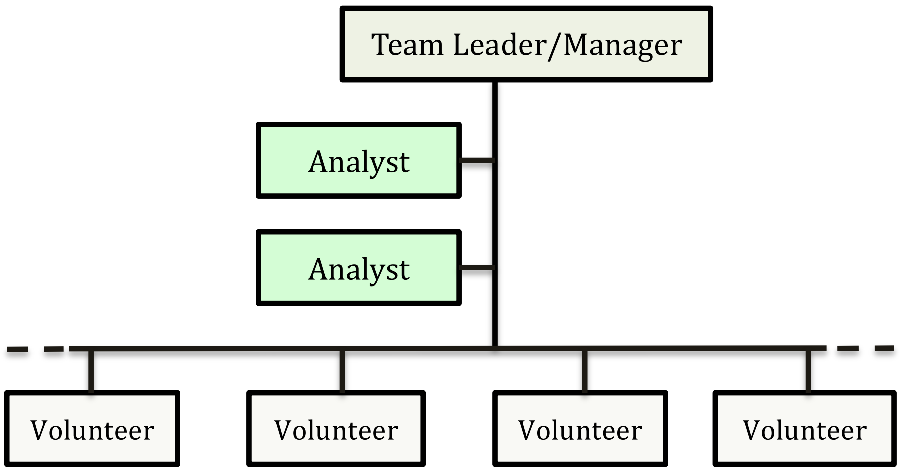

FIGURE 61 - FSU.VOST ORGANISATION

Using lessons learnt from this research, the VOST was organised to encourage a degree of independence for volunteers searching the digital landscape. The volunteers were free to use whatever tools they wanted to search social media. However, as is evident from this thesis, if there is high independence there has to be a means for aggregating sensemaking outputs. As such, an instance of Ushihidi; a crowd reporting software, was set up to allow volunteers a method of forming small reports on what they'd found. The volunteers are fundamentally gathering data and forming stories. Semantic sensemaking was occurring at the individual level and being aggregated through software. Analysts' duties were to verify the findings of the volunteers and approve findings to be entered into situational reports that were sent to FDEM at 8-12 hour intervals.

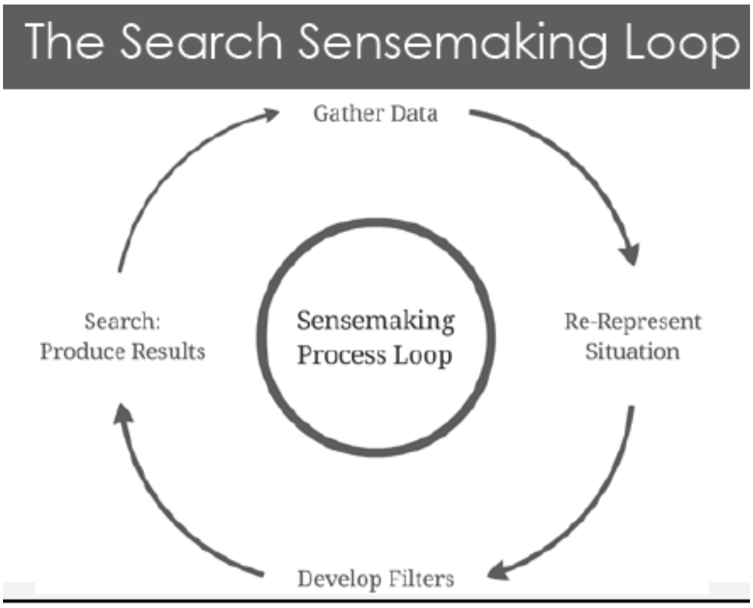

FIGURE 62 - SEARCH SENSEMAKING LOOP

In an effort to improve effectiveness and efficiency, the author provided training for volunteers on a mechanism for the search process (Figure 62). The introduction of a common knowledge book for current and discarded filters echoes the communities of practice approach to searching and helped avoid repetition in searches.

Stage	Process
Gather Data	Collect Posts from social media platforms. Look for data that confirms previous SA, provides new SA and/or verifies sources.
Represent Situation	Transform data into useable information (story creation). Upload report into Ushihidi.
Develop filter	Reassess key terms and hashtags in search filters. Discard redundant filters. Update knowledge book of filters.
Search	Apply filters onto monitoring tools. Gather results.

FSU.VOST was deployed three times during 2014 and has now become an integral part of FDEM's approach to disaster response and has received plaudits for its effectiveness as enhancing situation awareness during events.

The VOST shows how disaster management agencies in the modern era are being forced to respond to disasters in the online digital environment in addition to the real world. The VOST is a semantic sensemaking tool for analysing the data in the online social-media environment. As with all types of sensemaking there is a challenge of collecting and filtering salient data in order to create a sense of the event.

7.6 Final Remarks

This thesis has defined and developed a new notion of collaborative sensemaking, which it has been shown is possible to measure beyond the internal cognitive functions of the individual. This notion differs from the data/frame model of sensemaking presented by Klein as it extends beyond the individual to incorporate groups performing sensemaking tasks. It also differs from the collaborative sensemaking in the work of Weick by elaborating on collaborative processes and offering measurements for performance. Combining the semantic and pragmatic aspects of sensemaking as two sides of the same coin, provides a good explanation for the activities involved in disaster management and emergency response scenarios. The implication of this definition of collaborative sensemaking is that it may not be possible to design a single technology that supports all aspects of sensemaking. Future technologies in the realm of disaster management should look to support both semantic and pragmatic aspects of sensemaking to aid both information collection and strategic planning and response.

8 References

1. 7 July Review Committee (2006) Report of the 7 July Review Committee. London: Greater London Authority. Retrieved June 2007 from: http://legacy.london.gov.uk/assembly/reports/7july/report.pdf.

2. Alberts, D.S., & Hayes, R.E. (2006) Understanding Command and Control. Assistant Secretary Of Defense (C3I/COMMAND CONTROL RESEARCH PROGRAM) Washington DC.

3. Airwave (2010) Review of business. Retrieved December 2011 from: www.airwavesolutions.co.uk.

4. Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. *Handbook of reading research*, *1*, 255-291.

5. Artman, H. & C. Garbis (1998) Situation Awareness as Distributed Cognition, *Proceedings of the European Conference on Cognitive Ergonomics – 9,* Limerick.

6. Baber, C., Cross, J., Smith, P. and Robinson, D. (2007) Supporting implicit coordination between distributed teams in disaster management, in J. Löffler and M. Klann (Eds.): *Mobile Response 2007,* LNCS 4458, pp. 39-50, 2007

7. Baber, C., Smith, P., Cross, J., Hunter, J. & Mc Master, R. (2006). Crime scene investigation as distributed cognition. *Pragmatics and Cognition* **14**(2): 357-385.

8. Baber, C., Stanton, N. A., Atkinson, J., McMaster, R., & Houghton, R. J. (2013). Using social network analysis and agent-based modelling to explore information flow using common operational pictures for maritime search and rescue operations. *Ergonomics*, 56, 889-905.

9. Baber, C., Stanton, N., Houghton, R. & Cassia, M. (2008) Hierarchical command, communities of practice, networks of exploration: using simple models to explore NEC command structures*,* *Proceedings of NECTISE Conference*, Leeds.

10. Babitski, G.; Probst, F.; Hoffmann, J.; Oberle, D. (2009) Ontology Design for Information Integration in Disaster Management. In *Informatik 2009. 4th International Workshop on Applications of Semantic Technologies (AST'09).*

11. Bartlett, F.C., (Sir). (1932) Remembering: A study in experimental and social psychology. Cambridge University Press.

12. Bass, F. (1969). "A new product growth for model consumer durables". *Management Science* **15** (5): p215–227

13. Bharosa, N., Janssen, M., Rao, H. R., & Lee, J. (2008) Adaptive information orchestration: Architectural principles improving information quality. Paper presented at the *5th International Conference on Information Systems for Crisis Response and Management* (ISCRAM 2008), Washington, DC, USA.

14. Bharosa, N., Lee, J., & Janssen, M. (2010). Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers*, *12*(1), 49-65.

15. Burnett, M., Wooding, P., & Prekop, P. (2004). *Sense Making-Underpinning Concepts and*

Relation to Military Decision-making. DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION CANBERRA (AUSTRALIA).

16. Canoe, C. H., & Rosson, M. B. (2012). Articulating collaborative contributions to activity awareness. *Theories of team cognition: Cross-disciplinary perspectives*, 209.

17. Center for Disease Control and Prevention (CDC) website. Zombie Preparedness. Retrieved March 2014: http://www.cdc.gov/phpr/zombies.htm

18. Chan, K., Pressley, R., Rivera, B., & Ruddy, M. (2011). Integration of communication and social network modeling platforms using elicit and the wireless emulation laboratory. ARMY RESEARCH LAB ADELPHI MD.

19. Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, *13*(1991), 127-149.

20. Cross, R., Parker, A., Prusak, L., & Borgatti, S.P. (2001) Knowing what we know: Supporting knowledge creation and sharing in social networks. Organizational Dynamics, 30(2), 100-120.

21. Cross, R., Borgatti, S.P., & Parker, A. (2002) Making invisible work visible: Using network analysis to support strategic collaboration. California Management Review, 44(2).

22. Coroner's Inquests into the London Bombings of 7 July 2005. Hearing Transcripts. Retrieved November 2011: http://7julyinquests.independent.gov.uk/hearing_transcripts/06052011am.htm.

23. Dearstyne, B. (2007). The FDNY on 9/11: Information and decision making in crisis. *Government Information Quarterly*, *24*(1), 29-46.

24. De Jaegher, H., & Di Paolo, E. (2007). Participatory Sense-Making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6(4), 485-507.

25. Dervin, B. (1983). An overview of sense-making research: Concepts, methods and results. Paper presented at the annual meeting of the International Communication Association, Dallas, TX, May.

26. Dervin, B. (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of knowledge management*, *2*(2), 36-46.

27. Duffy, T., McMaster, R., Baber, C., & Houghton, R. (2012). *Towards an Ontology Broker to Improve Cross-agency Sharing in Emergency Response*. Proceedings of the 9th International ISCRAM Conference – Vancouver, April 2012. L. Rothkrantz, J. Ristvej and Z. Franco, eds.

28. Dunbar, R. I. (1998). The social brain hypothesis. *brain*, *9*(10), 178-190.

29. Endsley, M. R. (1988, May). Situation awareness global assessment technique (SAGAT). In *Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National* (pp. 789-795). IEEE.

30. Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors,* 37, 1, 32-64.

31. Endsley, M. R., Hoffman, R., Kaber, D., & Roth, E. (2007). Cognitive engineering and decision making: An overview and future course. *Journal of Cognitive Engineering and Decision Making*, *1*(1), 1-21.

32. Engelbart, D. (1992) Toward high-performance organizations: A strategic role of groupware. In proceedings of the GroupWare '92 Conference, San Jose, CA, August 3-5, 1992, Morgan Kaufmann.

33. Erasmus, A. C., Bishoff, E., & Rousseau, G. G. (2002). The potential of using script theory in consumer behaviour research. *Journal of Family Ecology and Consumer Sciences/Tydskrif vir Gesinsekologie en Verbruikerswetenskappe*, *30*(1).

34. ESF-6 SOG: Emergency Support Function 6 Mass Care Standard Operating Guide (Oct 2013)

35. FDEM: Florida Department of Emergency Management (Nov 21 2013). Emergency Support Functions, http://www.floridadisaster.org/emtools/esf.htm, Accessed: 9/1/2014

36. FEMA: Federal Emergency Management Agency EOC Activation Levels, http://emilms.fema.gov/IS775/assets/EOCActivationLevels.pdf, Accessed 9/1/2014.

37. Froese, T. (2012). Sense-making with a little help from my friends: Introducing Ezequiel Di Paolo and Hanne De Jaegher. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, (2), 143-146.

38. Florida State University (FSU) Website. Building Directory, Retrieved March 2013: http://campus.map.fsu.edu/index.aspx.

39. Garg, A. (2009) Gmail/Facebook Style jQuery Chat. Retrieved Aug 2012 from: http://anantgarg.com/2009/05/13/gmail-facebook-style-jquery-chat/

40. Gersh, J. R., McKneely, J. A., & Remington, R. W. (2005). Cognitive engineering: Understanding human interaction with complex systems. *Johns Hopkins APL technical digest*, *26*(4), 377-382.

41. Gigone, D., & Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology*, *65*(5), 959.

42. Google Maps API v3. https://developers.google.com/maps/web/

43. Granovetter, M. S. (1973). "The strength of weak ties." American Journal of Sociology(78): 1360-1380.

44. Granovetter, M. (1983). "The Strength of Weak Ties: A Network Theory Revisited". *Sociological Theory* 1: 201–233.

45. Granovetter, M. S. (1973). "The Strength of Weak Ties". *The American Journal of Sociology* **78** (6): 1360–1380.

46. Haddow, G.D, Bullock, J.A. and Coppola, D.P. (2011) "Introduction to Emergency Management", 4th edition, Butterword-Heinemann.

47. Head, H., Rivers, W. H. R., & Sherren, J. (1920). *Studies in neurology* (Vol. 2). London: Oxford University Press.

48. Heeger, D. (1998). Signal detection theory.

49. Heyes, R., (2005) PHP 5 implementation of the Porter Stemmer algorithm. Retrieved April 2012: http://www.phpguru.org/

50. Hill, R. A., & Dunbar, R. I. (2003). Social network size in humans. *Human nature*, *14*(1), 53-72

51. Hoegl, M. (2005). Smaller teams–better teamwork: How to keep project teams small. *Business Horizons*, *48*(3), 209-214.

52. Hutchins, E., & Lintern, G. (1995). *Cognition in the Wild* (Vol. 262082314). Cambridge, MA: MIT press.

53. Janis, I.L. (1971) "Groupthink", *Psychology Today* 5, 43-46, 74-76

54. Kåhre, J. (2012). *The mathematical theory of information* (Vol. 684). Springer Science & Business Media.

55. Kampis, G., Guly´as, L., Sz´aszi, Z. & Szakolczi, Z. (2009). Dynamic Social Networks and the Textrend/CIShell Framework. Applied Social Network Analysis Conference. University of Zurich and ETH.

56. Kash, W. (2012) Red Cross Launches Social Media Ops Center For Humanitarian Relief. http://gov.aol.com/2012/03/07/red-cross-launches-social-media-ops-center-for-humanitarian-reli/.

57. Kean, T. H., & Hamilton, L. H. (2004). The 9/11 Report: The national commission on terrorist attacks upon the United States. St. Martin's Paperbacks.

58. Keel, P. E. (2007). EWall: A visual analytics environment for collaborative sense-making. *Information Visualization*, *6*(1), 48-63.

59. Kendra, J. M., & Wachtendorf, T. (2003). Elements of resilience after the world trade center disaster: reconstituting New York City's Emergency Operations Centre. *Disasters*, *27*(1), 37-53.

60. Keuhlen, D.T., Bryant, O.L. and Young, K.K. (2002) The Common Operational Picture in Joint Vision 2020: a less layered cake, Norfolk, *VA: National Defense University Joint Forces Staff College.*

61. Klein, G. (2008) Natural Decision Making, Human Factors, Vol. 50, No. 3, 456-460.

62. Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986, September). Rapid decision making on the fire ground. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (Vol. 30, No. 6, pp. 576-580). SAGE Publications.

63. Klein, G., Moon, B., & Hoffman, R. R. (2006a). Making sense of sensemaking 1: Alternative perspectives. *Intelligent Systems, IEEE*, *21*(4), 70-73.

64. Klein, G., Moon, B. and Hoffman, R.F. (2006b). Making sense of sensemaking 2: a macrocognitive model. *IEEE Intelligent Systems*, *21*, 88-92

65. Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making* (pp. 113-155). Psychology Press.

66. Koffka, K. (1935). Gestalt psychology. *NY: Harcourt, Brace & World*.

67. Köhler, W. (1929). Gestalt psychology.

68. Landgren, J. (2005) Shared use of information technology in emergency response work: results from a field experiment, *Proceedings of the 2nd International ISCRAM Conference*, Brussels, Belgium, 2005

69. Lee, D. (2013) "Boston Bombing: How internet detectives got it very wrong", BBC news website, retrieved 19.4.2013, URL: http://www.bbc.co.uk/news/technology-22214511.

70. London Emergency Services Liaison Panel (LESLP) (2007) Major Incident Procedure Manual, seventh edition. *London: Metropolitan Police Service*. Retrieved June 2007 from: http://www.met.police.uk/leslp.

71. Maclean, N. (1992). *Young men and fire*. University of Chicago Press.

72. Mahajan, V., Muller, E., & Bass, F. M. (1995). Diffusion of new products: Empirical generalizations and managerial uses. Marketing Science, 14(3_supplement), G79-G88.

73. Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*(1), 1-21.

74. McMaster, R., & Baber, C. (2006). Assessing the impact of digital communications technology on existing C2 systems; a distributed cognition perspective. In *Proceedings of the 11th International Command and Control Research and Technology Symposium: Coalition Command and Control in the Networked Era* (pp. 26-28).

75. McMaster, R., Baber, C., & Duffy, T. (2012). The role of artefacts in police emergency response sensemaking. In *Proceedings of the 9th International ISCRAM Conference* (pp. 1-10).

76. McNeese, M. D., Pfaff, M. S., Connors. E. S. and Obieta, J. F. (2006). Multiple vantage points of the common operational picture: supporting international teamwork. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. 467-471.

77. Militello, L., Ormerod, T., Lipshitz, R., & Schraagen, J. M. (Eds.). (2012). *Naturalistic Decision Making and Macrocognition*. Ashgate Publishing, Ltd..

78. Minsky, M. (1975). Minsky's frame system theory. Schank & Nash—Webber, Theoretical Issues in Natu—ral Language Processing, Cambridge.

79. Muhren, W. J., Van Den Eede, G., & Van de Walle, B. (2008, May). Sensemaking as a methodology for ISCRAM research: Information processing in an ongoing crisis. In *Proceedings of the 5th International ISCRAM Conference* (pp. 315-323).

80. Nierhoff, M.H. (Dec 2013). *Research: Short Posts on Facebook, Twitter And Google+ Seem To Get More Interactions*. Social Media Analytics Blog, quintly. Retrieved: Aug 2014 https://www.quintly.com/blog/2013/12/short-posts-on-facebook-twitter-google-more-interactions/

81. Norman, D. A. (1986). Cognitive engineering. User centered system design: New perspectives on human-computer interaction, 3161.

82. Oinas-Kukkonen, H. (2008) *Network analysis and crowds of people as sources of new organizational knowledge*. In: A. Koohang, K. Harman, & J. Britz (Eds): *Knowledge Management: Theoretical Foundation* . Informing Science Press, Santa Rosa, CA, US, pp. 173-189.

83. Omodei, M. M., Wearing, A. J., McLennan, J., Elliott, G. C., & Clancy, J. M. (2005). More is better? Problems of self-regulation in naturalistic decision making settings. How professionals make decisions, 29-42.

84. Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of consulting psychology*,

29(3), 261-265.

85. Perrow, C. (1977). The bureaucratic paradox: The efficient organization centralizes in order to decentralize. *Organizational Dynamics*, *5*(4), 3-14.

86. Piaget, J. (1926). *The Language and Thought of the Child*. Kegan Paul, Trench, Trubner & Co.

87. Pirolli, P. & Russell, D.M. (2011) Introduction to this Special Issue on Sensemaking, *Human-Computer Interaction, 26*, 1-2, 1-8

88. Pirolli, P. & Card, S.K. (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis, In *International Conference on Intelligence Analysis* (McLean, VA, May 2005).

89. Reitter, D., & Lebiere, C. (2012, December). Social Cognition: Memory Decay and Adaptive Information Filtering for Robust Information Maintenance. In *AAAI*.

90. Rothermel, R. C. (1993). Mann Gulch fire: A race that couldn't be won. General Technical Report INT-299, Forest Service, United States Department of Agriculture.

91. Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In; RJ Spiro/BC Bruce/WF Brewer (Eds.): Theoretical Issues in Reading Comprehension.

92. Salas, E., Burke, C. S., & Cannon‐Bowers, J. A. (2000). Teamwork: emerging principles. *International Journal of Management Reviews*, *2*(4), 339-356.

93. Salmon, P., Stanton, N., Jenkins, D., & Walker, G. (2011). Coordination during multi-agency emergency response: issues and solutions. Disaster Prevention and Management: An International Journal, 20(2), 140-158.

94. Scaife, M. & Rogers, Y. (1996). "External cognition: how do graphical representations work?" International Journal of Human-Computer Studies(45): 185-213.

95. Schank, R. C., & Abelson, R. P. (1975). *Scripts, plans, and knowledge* (pp. 151-157). Yale University.

96. Selvin, A., Buckingham Shum, S., Seirhuis, M., Conklin, J., Zimmerman, B., Palus, C., ... & Li, G. (2001). Compendium: Making meetings into knowledge events.

97. Shneiderman, B., & Preece, J. (2007). 911. gov. *SCIENCE-NEW YORK THEN WASHINGTON-*, *315*(5814), 944.

98. Simon, R., & Teperman, S. (2001). The World Trade Center attack: lessons for disaster management. *Critical Care*, *5*(6), 318.

99. SNOWBALL. *The Porter stemming algorithm*, Retrieved 29th Nov 2015 from: http://snowball.tartarus.org/algorithms/porter/stemmer.html

100. Sorensen, L.J., & Stanton, N.A. (2010), Comparison of Situational Awareness in Hierarchical C2 and Edge Network Structures, *15th ICCRTS: Evolution of C2*, Santa Monica, CA.

101. Sorensen, L. J., & Stanton, N. A. (2011). Is SA shared or distributed in team work? An exploratory study in an intelligence analysis task. International Journal of Industrial Ergonomics, 41(6), 677-687.

102. Stanton, N.A., Walker, G.H., & Sorensen, L.J., (2012) It's a small world after all: contrasting hierarchical and edge networks in a simulated intelligence analysis task, *Ergonomics* Vol.55, No. 3, 265-281

103. Stasko, J., Görg, C., & Liu, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. Information visualization, 7(2), 118-132.

104. Surowiecki, J. (2004) *The Wisdom of Crowds*, Anchor Books.

105. Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi*, *28*(1), 23-30.

106. Thorndike, E. L. (1913). *The psychology of learning* (Vol. 2). Teachers College, Columbia University.

107. Travica, B. (1999) *New Organizational Designs: Information Aspects*, Ablex Publishing Corporation

108. Trier, M. & Bobrik, A. (2007). Analyzing the Dynamics of Community Formation using Borkering Activities. In: Proceedings of the Third Communities and Technologies Conference, Michigan, Springer Series.

109. Umapathy, K. (2010). Requirements to support collaborative sensemaking. In Computer Supported Cooperative Work (CSCW) Workshop on Collaborative Information Seeking.

110. Valente, T. W. & Davis, L. D. (1999). "*Accelerating the diffusion of innovation using opinion leaders.*" Annals of AAPSS(566): 55-67.

111. Vilhena, D. A., Foster, J. G., Rosvall, M., West, J. D., Evans, J., & Bergstrom, C. T. (2014). Finding cultural holes: how structure and culture diverge in networks of scholarly communication. *Sociological Science*, *1*, 221-238.

112. Walker, G., Salmon, P., Green, D., Stanton, N. & Baber, C. (2013) Towards A Theory Of Agent-Bases Systemic Situational Awareness. In D.A. Vincenzi, M. Mouloua & P.A. Hancock (Eds.), *Human Performance, Situation Awareness, and Automaion: Current Research and Trends HPSAA II, Volumes I and II,* Psychology Press

113. Walker, G. H., Stanton, N. A., Salmon, P. M., Jenkins, D. P., & Rafferty, L. (2009). Combining social network analysis and the NATO approach space to define agility. In *Proceedings of the 14th International Command and Control Research and Technology Symposia*.

114. Walker, G. H., Stanton, N. A., Stewart, R., Jenkins, D., Wells, L., Salmon, P., & Baber, C. (2009b). Using an integrated methods approach to analyse the emergent properties of military command and control. Applied ergonomics, 40(4), 636-647.

115. Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, Cambridge University Press.

116. Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative science quarterly*, 628-652.

117. Weick, K.E. (1995) *Sensemaking in Organizations*, London: Sage.

118. Weick, K. E. (1988). ENACTED SENSEMAKING IN CRISIS SITUATIONS [1]. *Journal of*

management studies, *25*(4), 305-317.

119. Wegner, D.M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G.R. Goethals (Eds.), *Theories of group behaviour* (pp185-205). New York: Springer-Verlag.

120. Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating Communities of Practice (Hardcover)*. Harvard Business Press; 1 edition.

121. Wolbers, J., & Boersma, K. (2013). The common operational picture as collective sensemaking. *Journal of Contingencies and Crisis Management*, *21*(4), 186-199.

122. Yates, J. F., Veinott, E.S., & Patalano, A. L. (2003). Hard decision, bad decision: On decision quality and decision aiding. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (13-63). New York: Cambridge University Press

9 Appendix 1: Article Commentary Experiment Results

For the reader's benefit some example result sets are shown as network graphs because it was untenable to present them as matrices due to the number of participants involved. The graphs can be used visually to make quick comparisons regarding the amount of alignment between participants in relation to article introduced (AI) and participant introduced (PI) terms. The inclusion of all results sets was deemed unnecessary since beyond visual inspection they do not offer much insight

9.1 Results

9.1.1 Article 14 Results

For this article we observe high levels of alignment between participants for AI content. There is some strong alignment between participants based on PI content. For this article there was higher average of key terms per participant (69.7 key terms per participant) than all of the other articles analysed from the author's experimental platform. Additionally the ratio of interactions to the number of participants (participation ratio) shows that on average each participant posted more than once. This is indicative that more of a discussion was taking place around on this article.

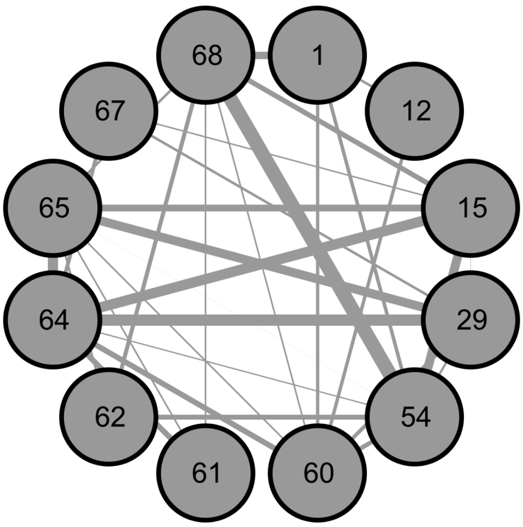

FIGURE 63 - ARTICLE 14: ARTICLE CONTENT ALIGNMENT

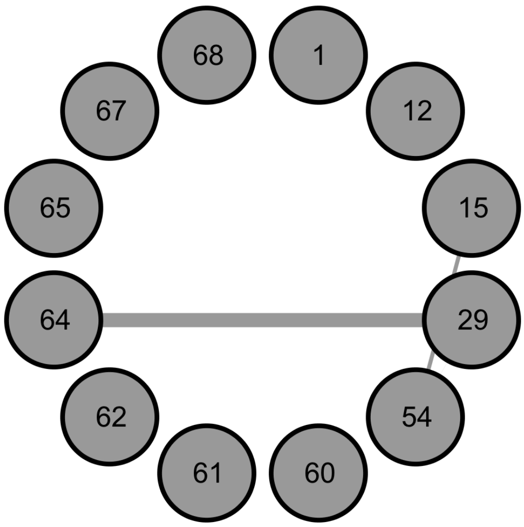

FIGURE 64 - ARTICLE 14: PARTICIPANT CONTENT ALIGNMENT

9.1.2 Article 27 Results

This results set was included to show that for the mined content, the result sets became quite large and the legibility of the network graphs decreased.

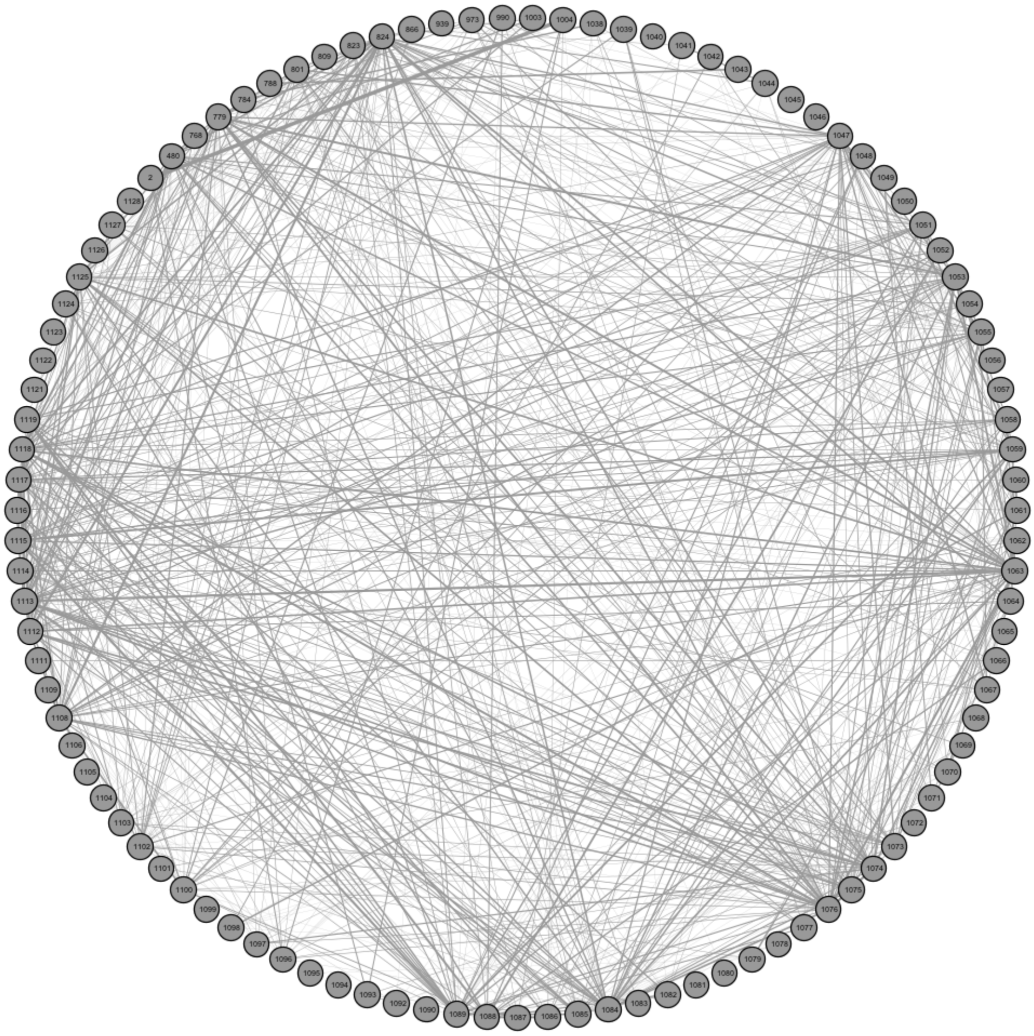

FIGURE 65 - ARTICLE 27: ARTICLE CONTENT ALIGNMENT

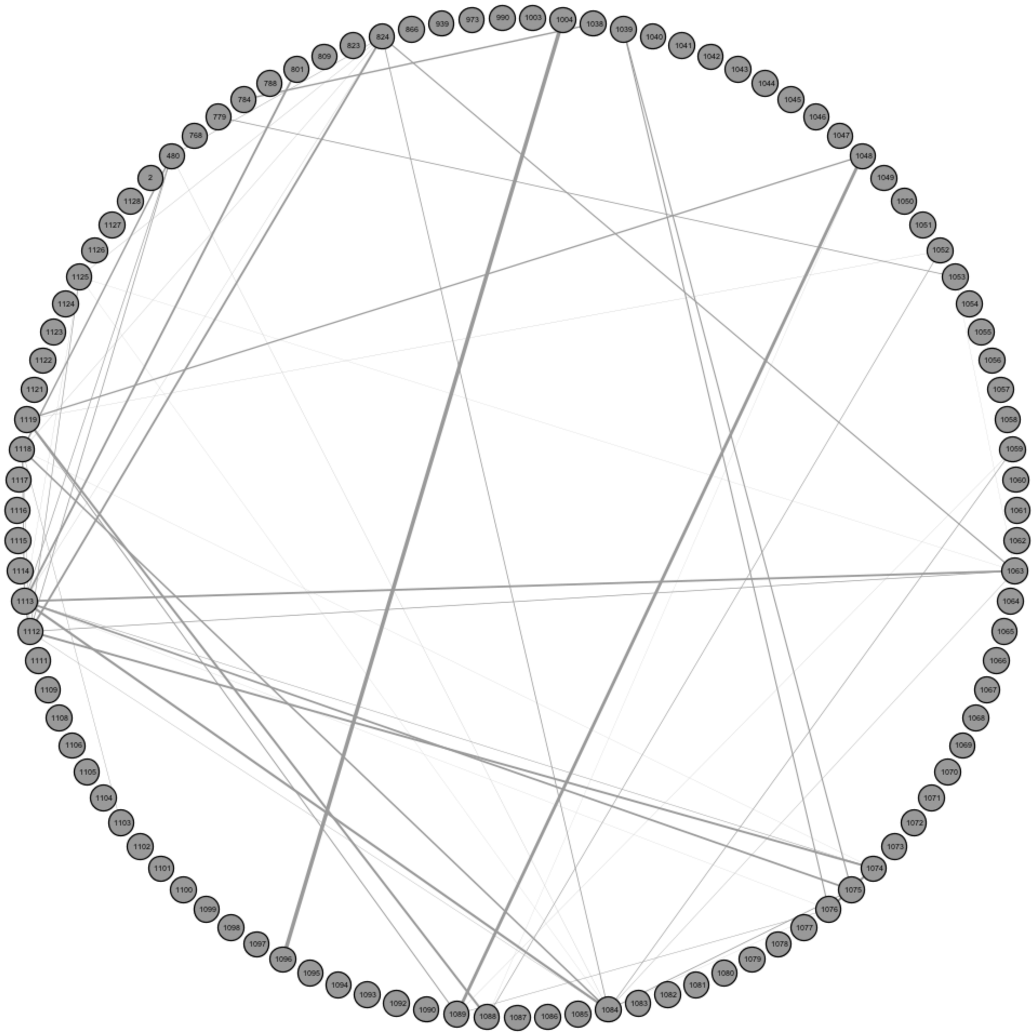

FIGURE 66 - ARTICLE 27: PARTICIPANT CONTENT ALIGNMENT

10 Appendix 2: ELICIT results

10.1 Diffusion Coefficients

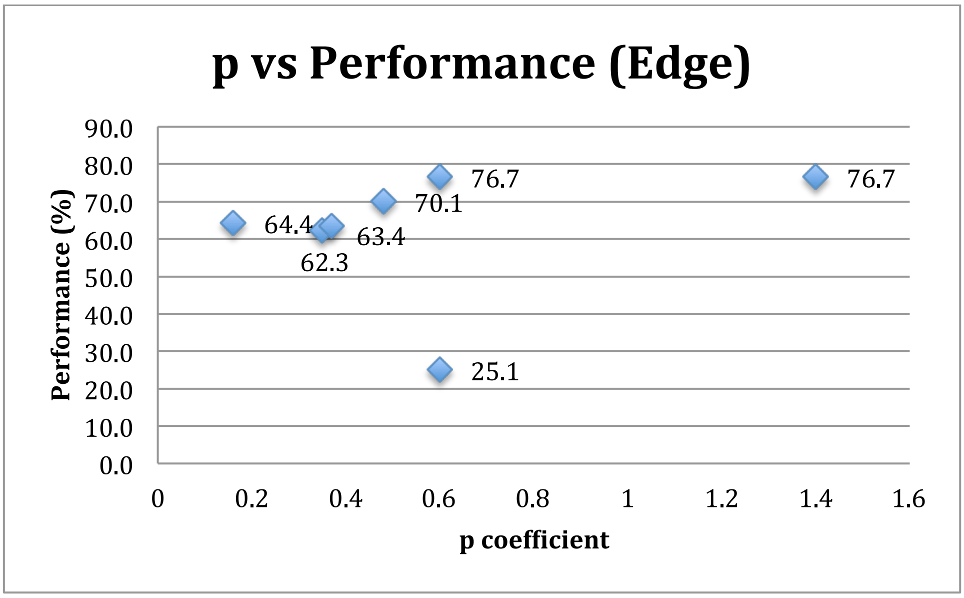

Figure 67 - p Coefficient Vs Performance for Edge Groups

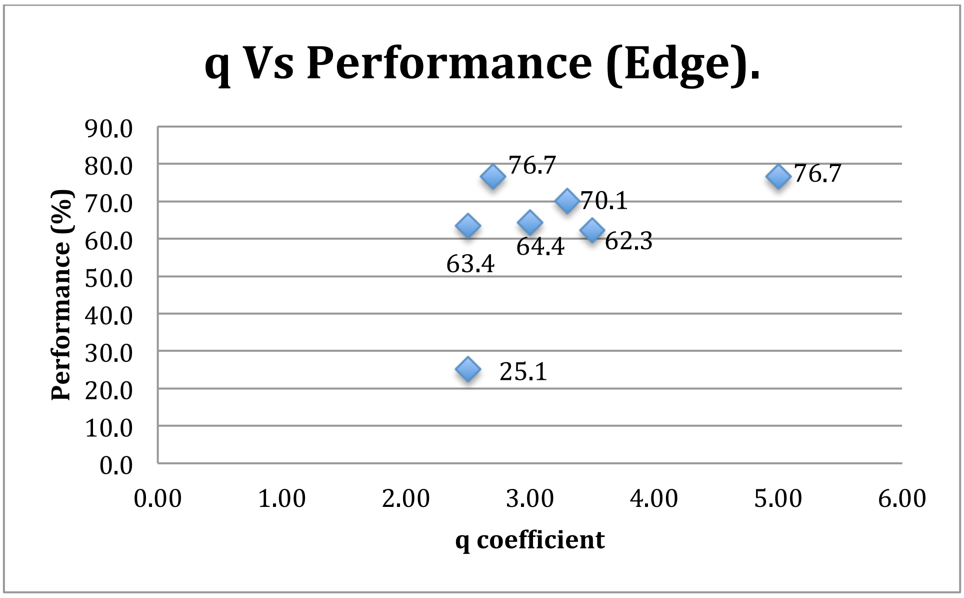

Figure 68 - q Coefficient Vs Performance for Edge Groups

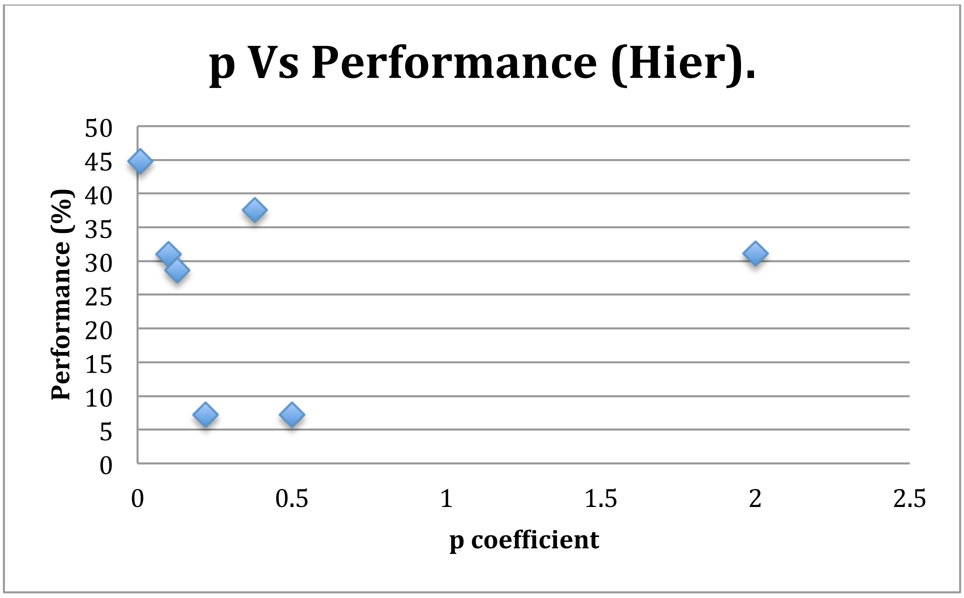

FIGURE 69 - P COEFFICIENT VS PERFORMANCE FOR HIERARCHICAL GROUPS

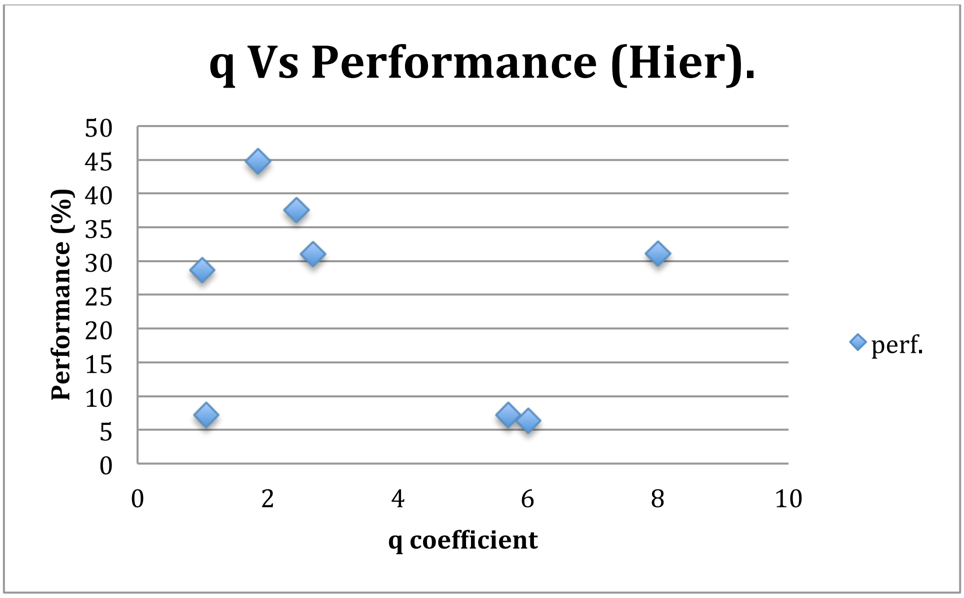

FIGURE 70 - Q COEFFICIENT VS PERFORMANCE FOR HIERARCHICAL GROUPS

10.2 Group Performance Results

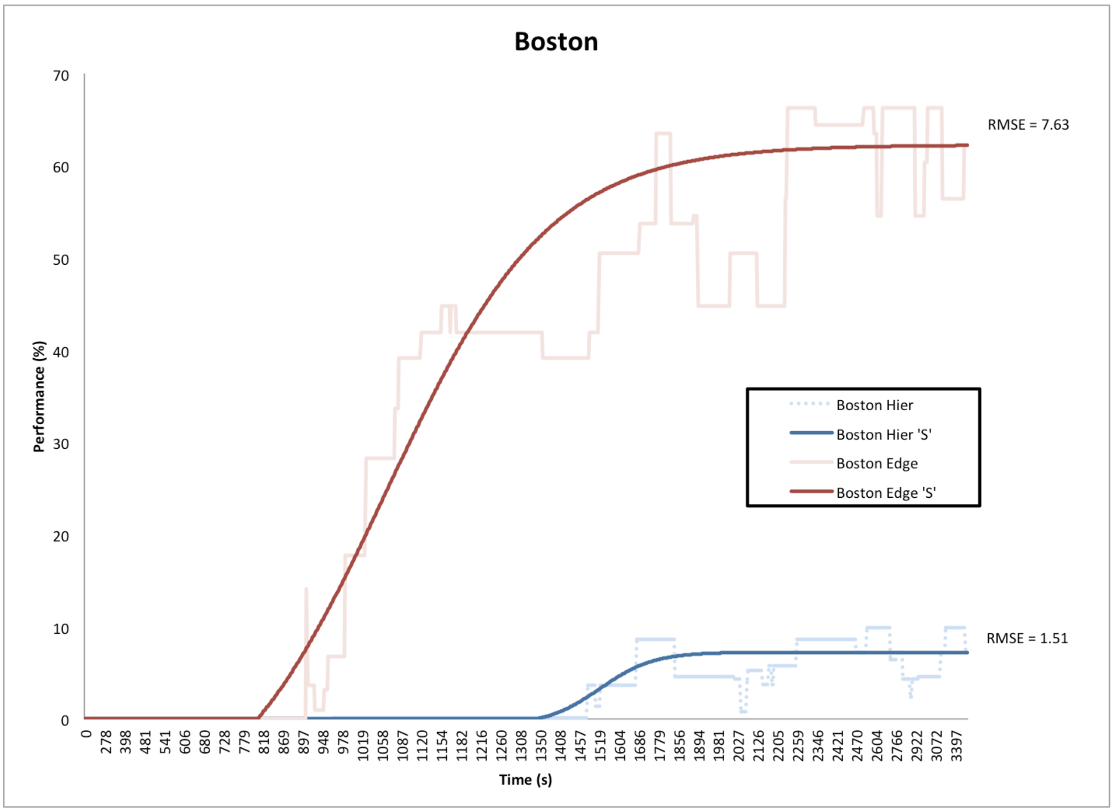

FIGURE 71 - BOSTON RESULTS

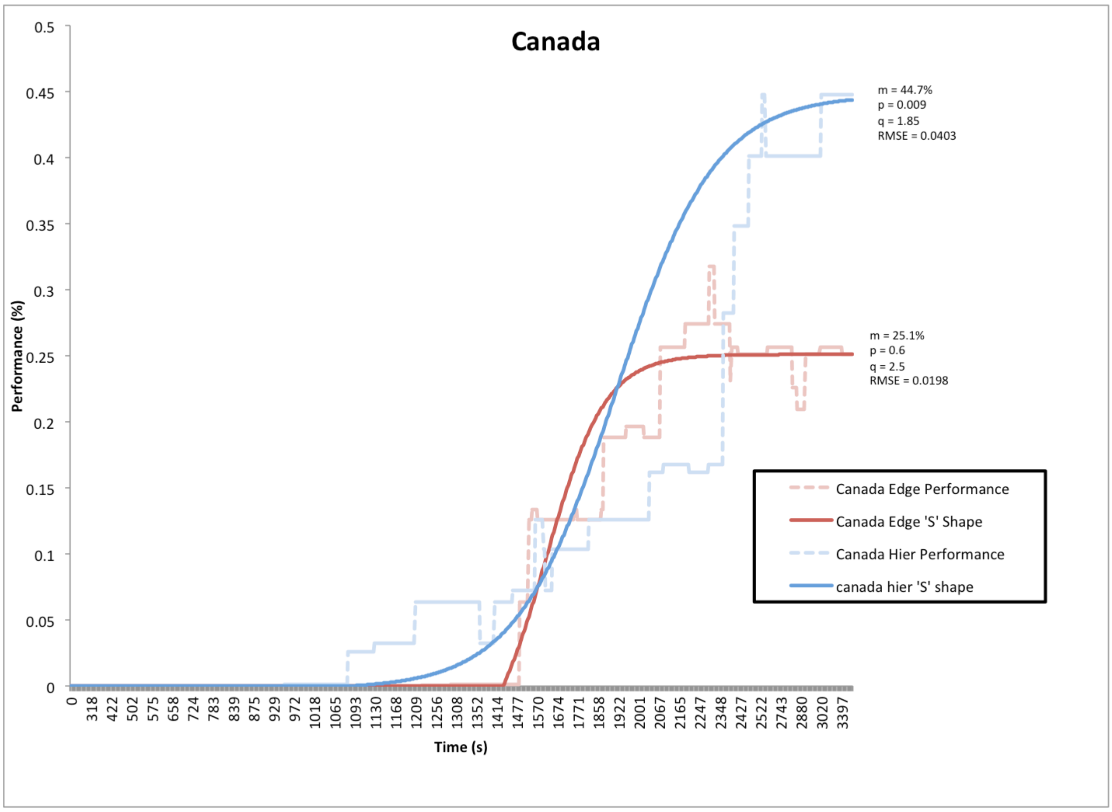

FIGURE 72 - CANADA RESULTS

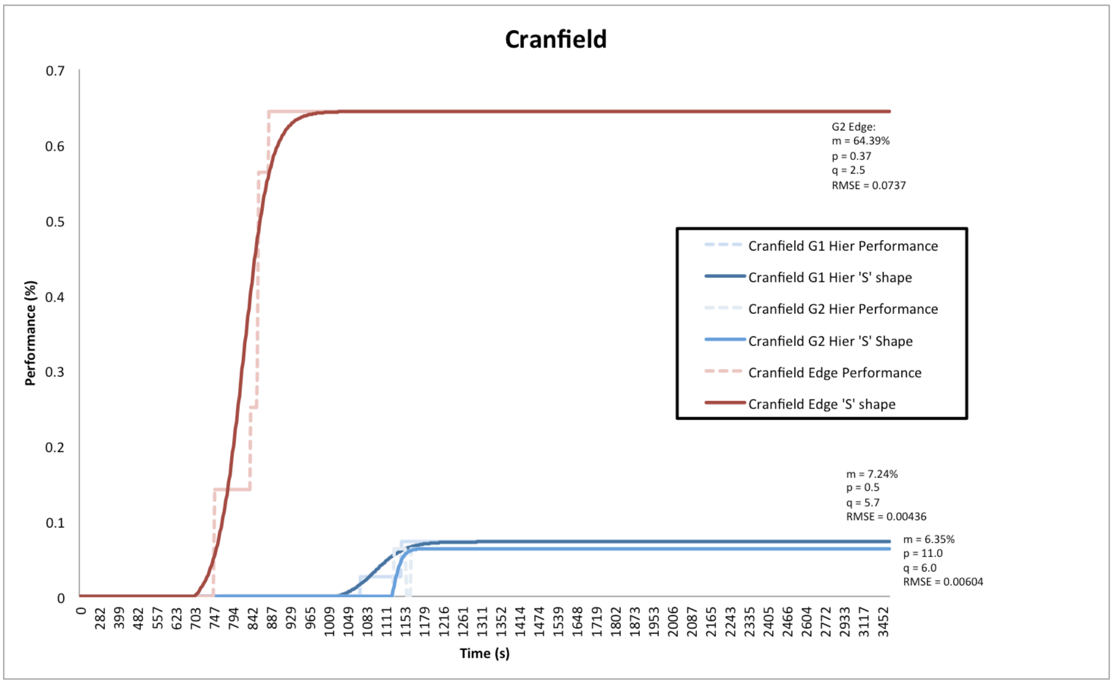

FIGURE 73 - CRANFIELD RESULTS

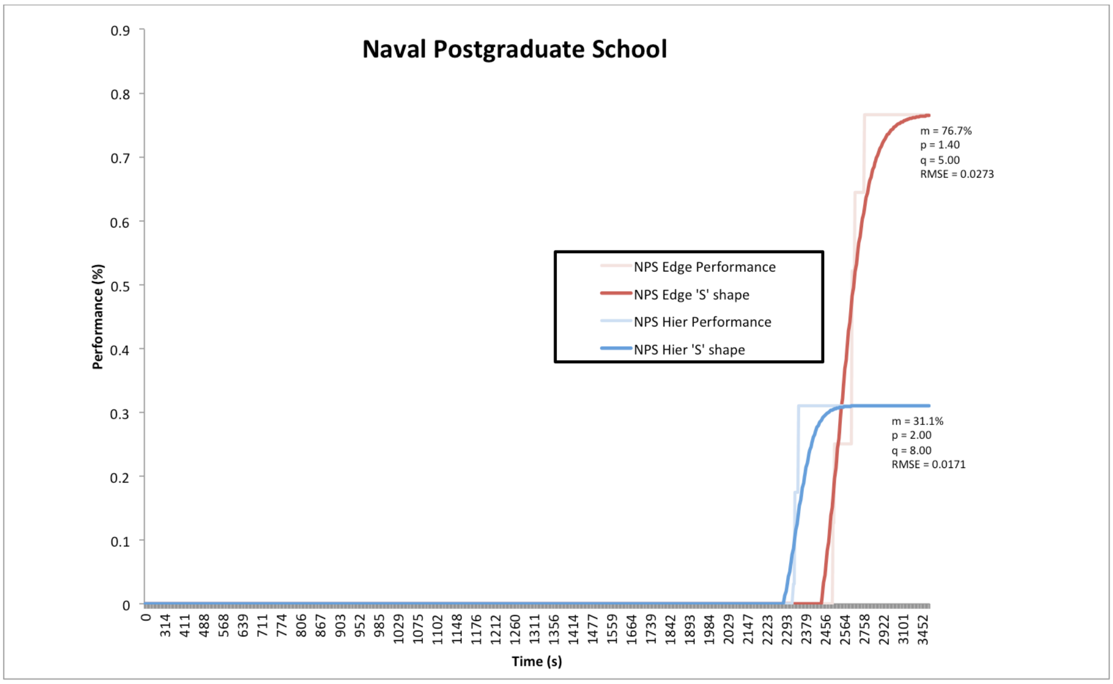

FIGURE 74 - NAVAL POSTGRADUATE SCHOOL RESULTS

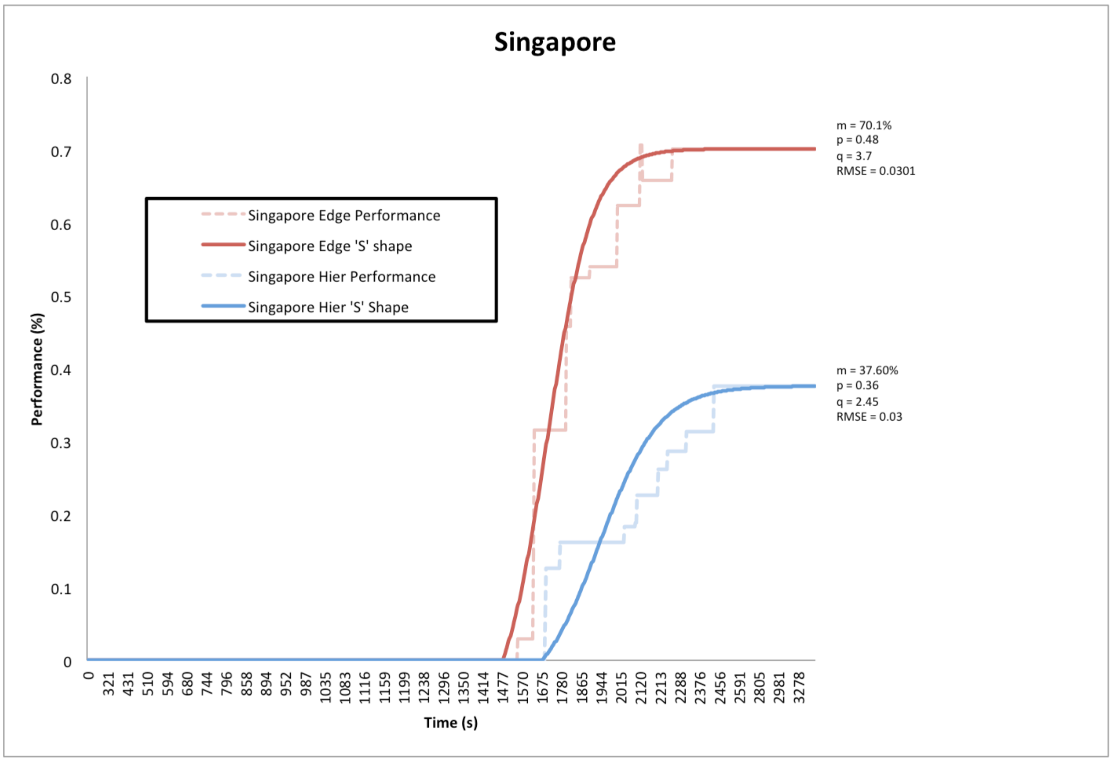

Figure 75 - Singapore Results

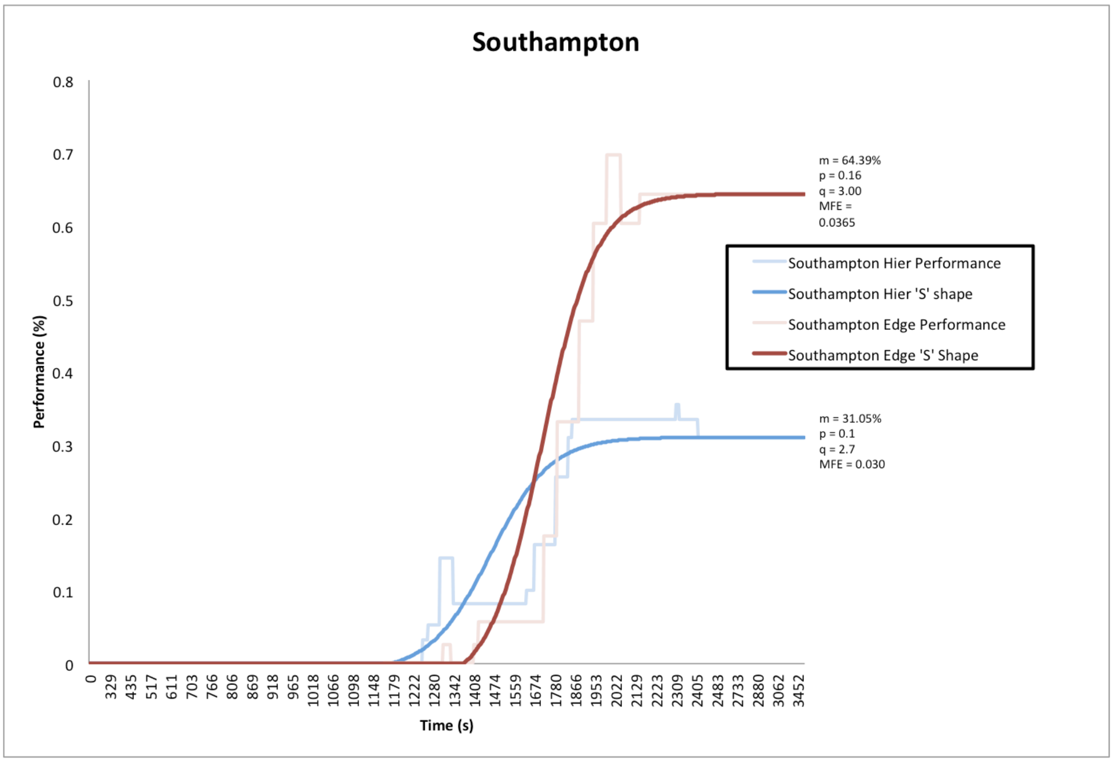

Figure 76 - Southampton Results

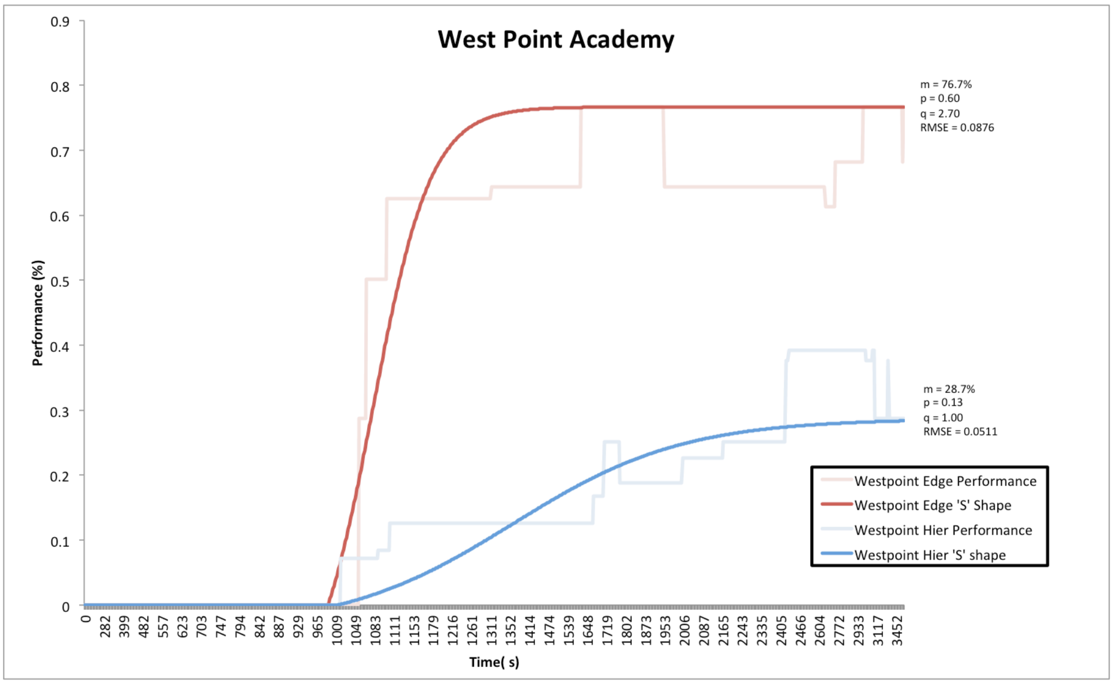

FIGURE 77 - WEST POINT RESULTS

11 Appendix 3: Zombie Experiment Results

11.1 Social Networks

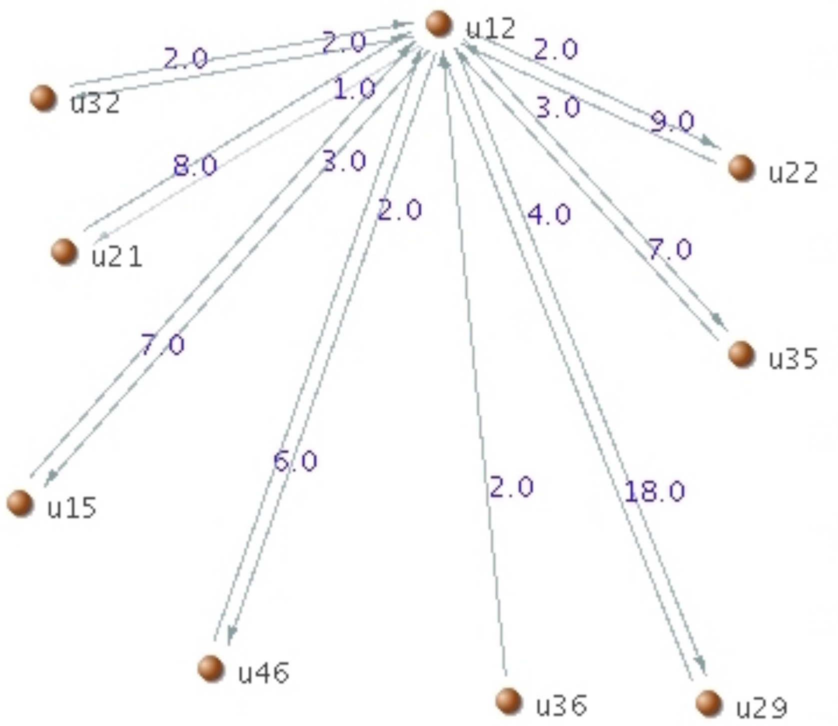

FIGURE 78 - GROUP 1 SOCIAL NETWORK AFTER T3

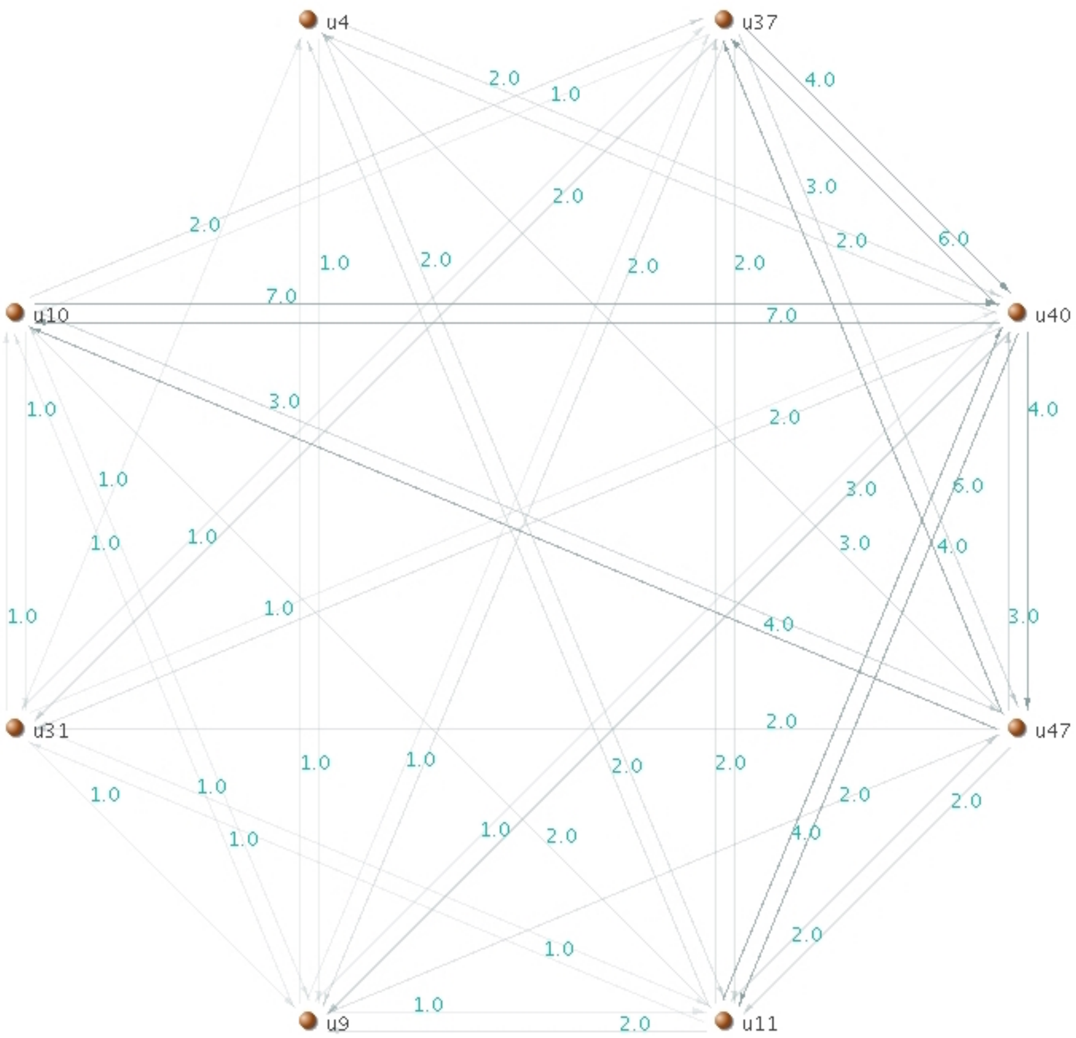

FIGURE 79 - GROUP 2 FINAL SOCIAL NETWORK

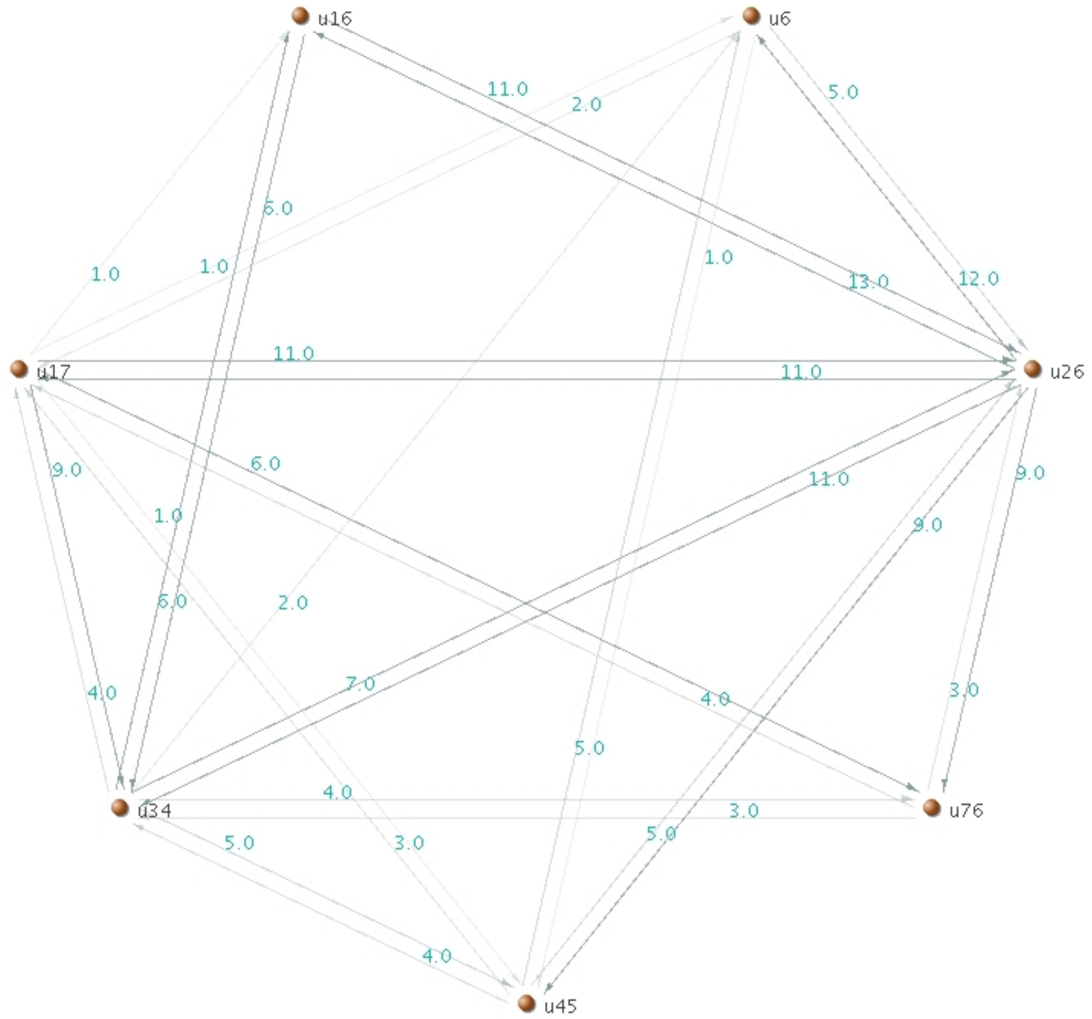

FIGURE 80 - GROUP 3 FINAL SOCIAL NETWORK

11.2 Chat Logs

11.2.1 Group 1: t1

Time	From	To	Message
12:27:50	22	12	w
12:27:51	22	12	1
12:28:15	35	12	ummm...does anyone know what to do?
12:28:39	21	12	traditons medium
12:28:44	15	12	stay away from call st
12:28:46	29	12	Medium level zombie threat at the free standing Starbucks South of the Traditions statue.
12:29:03	46	12	I have no zombies
12:29:03	36	12	hey
12:29:34	35	12	stay away from satellite utility plant 2
12:30:07	46	12	High zombie threat at William Johnson
12:31:10	22	12	I'm going to french town
12:31:39	29	12	Any thing near woodward?
12:31:51	15	12	zombies in Wescott

12:31:56	12	22	D chem lab R
12:32:08	46	12	Are there zombies at the Fueling station??
12:32:18	12	29	D chem lab R
12:32:22	12	35	D chem lab R
12:32:26	12	32	D chem lab R
12:32:37	29	12	What level?
12:32:37	12	21	D chem lab R
12:32:56	12	15	D chem lab R
12:33:02	32	12	woodward ave
12:33:03	22	12	What the hell does "D chem lab R" mean?
12:33:16	21	12	unknowns: landis, english language institute, broward hall, katherine w montgomery hall,oglesby union at fsu, fuelina station, and dept
12:33:18	32	12	medium
12:33:20	21	12	of risk managment
12:33:36	35	12	all my pins are in landis green
12:34:04	29	12	Two zombies on Aacademic way by Mcollum Hall

11.2.2 Group 2: t1

Time	From	To	Message
12:31:57	40	37	I have a high threat zombie by Mike Long track
12:32:07	47	4	I have 1 pin on south landis green
12:32:11	40	37	I have a high threat zombie by Mike Long track
12:32:20	40	47	I have a high threat zombie by Mike Long track
12:32:26	40	11	I have a high threat zombie by Mike Long track
12:32:30	40	9	I have a high threat zombie by Mike Long track
12:32:38	40	31	I have a high threat zombie by Mike Long track
12:32:43	40	10	I have a high threat zombie by Mike Long track
12:32:46	40	4	I have a high threat zombie by Mike Long track
12:32:48	37	40	I show high threat at dittmer
12:33:54	47	40	I have a high threat zombie by the leach center
12:34:05	37	40	low threat outside strozier
12:34:17	40	37	I have low threat zombie outside montgomery hall
12:34:39	47	10	I have a high threat zombie by the leach center
12:34:58	11	40	i have a medium threat by Mendenhall A
12:35:14	47	9	I have a high threat zombi by the leach center
12:35:28	10	47	sweet! i have a medium threat zomb at parking garage 2 just north of traditions way
12:35:36	10	47	before collegiate loop
12:36:00	10	37	7: I have a high threat zombie by the leach center 10: sweet! i have a medium threat zomb at parking garage 2 just north of traditions way
12:36:25	37	47	high threats at mike long, ditmer
12:36:32	40	11	Got it
12:36:58	11	40	thanks!
12:37:09	37	47	low threats at strozier, montgomery
12:37:10	11	37	medium zombie threat mendenhall A
12:37:38	11	47	medium zombie threat mendenhall A
12:37:45	11	47	high zombie threat mike long track
12:37:51	10	40	there is a medium threat zombie at parking garage 2 by traditions way and collegiate loop
12:38:01	10	40	and theres a high threat zomb by the leach

11.2.3 Group 3: t1

Time	From	To	Message
12:35:27	34	6	zombie on south end of landis green
12:35:35	26	6	Ok how does this thing work
12:35:51	34	26	zombie on south end of landis green
12:35:55	34	76	zombie on south end of landis green
12:35:59	34	45	zombie on south end of landis green
12:36:00	6	26	no clue
12:36:02	34	17	zombie on south end of landis green
12:36:04	34	16	zombie on south end of landis green
12:36:43	26	6	Zombie by mike long track, red
12:36:46	17	34	what threat?
12:36:49	26	76	Zombie by mike long track, red
12:36:50	26	34	Zombie by mike long track, red
12:36:53	34	16	Threat level: High
12:37:01	26	16	Zombie by mike long track, red
12:37:01	16	34	what are we supposed to do lol
12:37:02	26	17	Zombie by mike long track, red
12:37:03	34	17	hreat level: High
12:37:04	26	45	Zombie by mike long track, red
12:37:05	34	45	hreat level: High
12:37:07	34	76	hreat level: High
12:37:08	17	34	low threat zombie on collegiate loop
12:37:09	34	26	hreat level: High
12:37:10	6	26	what does that mean?
12:37:11	34	6	hreat level: High
12:37:20	16	26	I don't see any
12:37:21	17	26	low threat zombie on collegiate loop
12:37:33	26	6	Yellow Zombie woodward garage
12:38:29	34	16	Tell everyone where the zombie is that you can see
12:38:30	76	17	unknown near chemistry lab and chemical sciences lab
12:38:41	16	34	I havnt seena ny yet
12:38:46	26	17	Got it
12:38:52	16	34	what do I do once I see one?
12:38:54	17	76	red at college of medicine
12:38:57	17	26	red at college of medicine
12:39:00	17	34	red at college of medicine
12:39:00	26	16	Im telling you where it shows on my map
12:39:11	34	16	Tell every its location
12:39:23	17	76	unknown at degraff
12:39:26	17	26	unknown at degraff
12:39:29	17	34	unknown at degraff
12:39:48	16	34	mike long track has a zombie
12:39:49	26	17	same
12:40:05	26	16	and you tell me where they are on yours
12:40:17	76	17	Al Qaeda is in the Union!!!
12:40:28	76	26	Al Qaeda is in the Union!!!
12:40:31	26	34	Unk at Dittner Chem lab
12:40:38	17	76	red?
12:40:44	26	76	?

12:40:44	76	34	Al Qaeda is on landis!!!
12:41:04	76	26	Osama's zombie body
12:41:16	76	17	Osama zombie

11.2.4 Group 1: t3

Time	From	To	Message
12:50:53	29	12	How is the union?
12:50:53	15	12	green guy at jennie hall
12:51:03	22	12	screenshot of my map: http://i.imgur.com/VgCX8lE.png
12:51:13	21	12	medium on oglesby
12:51:22	21	12	medium at bellamy
12:51:33	21	12	low at reynolds
12:51:40	21	12	high at dither chem lab
12:52:22	29	12	Do not go near the West side of Campus. Anything past the Collegiate loop is dangerous.
12:53:13	12	15	tully good
12:53:20	12	29	tully good
12:53:38	22	12	Screw you guys I'm running off on my own
12:53:39	15	12	green guy? or nothing
12:53:56	12	35	tullys good
12:54:04	12	15	nothing
12:54:07	29	12	Tully's good.
12:54:33	35	12	...teletubbies?
12:55:46	12	46	go to tully
12:55:58	12	35	none are at tully
12:56:27	46	12	On the way
12:56:43	35	12	dept of physics - what level of threat?
12:57:09	35	12	it's off chiefwan

11.2.5 Group 2: t3

Time	From	To	Message
12:46:15	11	9	medium zombie threat mendenhall a
12:46:29	47	37	HIgh level at leach
12:46:31	9	37	high alert zombie at landis
12:46:36	11	31	medium zombie threat mendenhall A
12:46:36	40	37	Medium threat zombie at Montgomery
12:46:44	40	47	Medium threat zombie at Montgomery
12:46:46	47	37	low level at katherine montgomery hall
12:46:46	40	11	Medium threat zombie at Montgomery
12:46:48	40	9	Medium threat zombie at Montgomery
12:46:49	37	40	med threat at woodward parking
12:46:51	40	31	Medium threat zombie at Montgomery
12:46:55	4	40	I have a med level of walkers in the union
12:46:57	47	40	high at leach and low at montgomery hall
12:47:01	9	40	high alert zombie at landis
12:47:09	9	4	high alert zombie at landis
12:47:16	9	10	high alert zombie at landis

12:47:16	47	11	high level at leach and low at katherine montgomery hall
12:47:19	31	37	High threat Zombies to the right of Landis Green--near Wlliam JOhnston and the other near Cawthorn Hall
12:47:21	9	11	high alert zombie at landis
12:47:39	11	40	where is montgomery?
12:47:50	47	9	low at montgomery hall
12:47:57	4	9	Med level of walkers in SSB
12:48:04	47	31	high at leach and low at montgomery hall
12:48:06	4	40	Also in ssb
12:48:16	47	4	high at leach and low at montgomery hall
12:48:25	10	9	high alert zombie at jennie murphy
12:48:25	31	40	High threat Zombies to the right of Landis Green--near Wlliam JOhnston and the other near Cawthorn Hall
12:48:26	40	10	Medium threat zombie at Montgomery
12:48:29	31	11	High threat Zombies to the right of Landis Green--near Wlliam JOhnston and the other near Cawthorn Hall
12:48:34	47	10	high at leach and low at montgomery hall
12:48:53	10	40	high alert at jennie murphy
12:49:04	11	9	thanks!
12:49:14	31	9	High threat Zombies to the right of Landis Green--near Wlliam JOhnston and the other near Cawthorn Hall
12:49:17	31	10	High threat Zombies to the right of Landis Green--near Wlliam JOhnston and the other near Cawthorn Hall
12:49:20	31	4	High threat Zombies to the right of Landis Green--near Wlliam JOhnston and the other near Cawthorn Hall
12:49:30	10	47	high jennie murphy
12:49:36	10	31	high jennie murphy
12:50:02	37	31	low threat at jim moran
12:50:12	37	9	low threat at jim moran
12:50:20	40	11	left of landis
12:50:20	37	11	low threat at jim moran
12:50:21	47	10	lots at english langage institute
12:50:30	40	11	straight line from the pine tree icon on the map
12:50:46	37	47	low threat at jim moran
12:50:56	37	10	low threat at jim moran
12:51:11	47	37	high threat at english language institute
12:51:13	11	40	thanks! high on landis and cawthon
12:51:14	37	40	low threat at jim moran
12:51:40	10	37	high jennie murphy
12:52:16	40	37	where is that
12:52:33	11	37	where is jim moran?

11.2.6 Group 3: t3

Time	From	To	Message
12:56:09	6	45	where r ur zombies
12:56:15	17	34	i have a lot
12:56:46	26	34	Who has a zombie at Dittmer Lab
12:56:47	26	6	Who has a zombie at Dittmer Lab
12:56:49	26	16	Who has a zombie at Dittmer Lab
12:56:51	26	17	Who has a zombie at Dittmer Lab

12:56:53	26	76	Who has a zombie at Dittmer Lab
12:56:56	26	45	Who has a zombie at Dittmer Lab
12:56:56	34	17	threat level of leach zombie?
12:57:02	45	6	green on love buiding
12:57:03	34	45	threat level of leach zombie?
12:57:06	45	6	green on woodward
12:57:06	34	16	threat level of leach zombie?
12:57:08	34	76	threat level of leach zombie?
12:57:11	34	26	threat level of leach zombie?'
12:57:16	17	26	unknown there
12:57:17	45	6	yellow on fuel station
12:57:24	26	34	unk
12:57:31	45	6	red on chemistry lab
12:57:33	17	34	unknown
12:57:43	45	6	red on dept of physics
12:58:00	16	34	there is none there
12:58:09	26	6	Rovetta Building Zombie any one?
12:58:14	26	34	Rovetta Building Zombie any one?
12:58:16	26	16	Rovetta Building Zombie any one?
12:58:17	26	17	Rovetta Building Zombie any one?
12:58:59	26	76	Rovetta Building Zombie any one?
12:59:02	26	45	Rovetta Building Zombie any one?
12:59:17	26	6	Traditions way Zombie?
12:59:23	26	34	Traditions way Zombie?
12:59:24	26	16	Traditions way Zombie?
12:59:25	26	17	Traditions way Zombie?
12:59:27	26	76	Traditions way Zombie?
12:59:28	26	45	Traditions way Zombie?
12:59:51	17	26	unknown at traditions
1:00:06	26	17	Jenne Murphy?
1:00:09	16	26	nope
1:00:16	26	16	Jenne Murphy?
1:00:18	26	34	Jenne Murphy?
1:00:19	26	6	Jenne Murphy?
1:00:22	26	76	Jenne Murphy?
1:00:22	26	45	Jenne Murphy?
1:00:43	26	6	William Johnson Building?
1:00:48	26	34	William Johnson Building?
1:00:50	26	16	William Johnson Building?
1:00:51	26	17	William Johnson Building?
1:00:53	26	76	William Johnson Building?
1:00:54	26	45	William Johnson Building?
1:00:57	17	26	unknown
1:01:05	17	26	red at WJB
1:01:08	16	26	I only have 3
1:01:14	16	26	landis
1:01:15	26	16	Ok
1:01:18	16	26	moore auditorium
1:01:22	45	26	i have a unknown by william johnston
1:01:24	16	26	and woodward garage
1:02:34	45	34	i have unknown for leach

12 Appendix 4: Porter Stemmer Algorithm

Porter's(1980) stemming algorithm is broken into five steps. Within each step are a series of rules. Some rules do an initial manipulation of a stem that is later finished by another rule. The algorithm makes use of a *measure (denoted 'm')* of a word or part of a word. Any word can be broken into a sequence of vowels and consonants. 'C' denotes a sequence of one or more consonants. 'V' denotes a sequence of one or more vowels. As such, any word can be represented by the form:

$$[C](VC)^m[V]$$

Where, [C] represents an optional sequence of consonants; $(VC)^m$ is 'm' repetitions of VC; [V] represents an optional sequence of vowels.

As an example the measure of the word *crepescule* is 4. (example taken from SNOWBALL, no date)

$$cr|\ ep\ |\ usc\ |\ ul\ |\ ar|$$
$$[c]|VC|\ V\ C\ \ |VC|VC|$$
$$1 \quad 2 \quad 3 \quad 4$$

12.1 Notation:

c – consonant.

v – vowel.

m – measure of word.

(*condition*) – a condition exists for the rule to be enacted.

*S – the stem ends with an S (and could be another letter).

v - the stem contains a vowel.

*d – the stem end with a double consonant (e.g. 'TT', 'SS').

*o – the stem ends with cvc, where the second c is not W,X or Y.

L – a generic letter (consonant of vowel).

Note: Where a letter is used that is not part of notation, it represents the literal character.

12.2 The Algorithm

12.2.1 Step 1a

This step removes Plurals.

#	Rule	Example
1.	SSES → SS	Caresses → caress
2.	IES → I	Ponies → poni
3.	SS → SS	Caress → caress
4.	S →	Cats → cat

12.2.2 Step 1b

This step removes past participles.

#	Rule	Example
1.	(m>0) EED → EE	feed → feed ; agreed → agree
2.	(*v*) ED →	plastered → plaster
		bled → bled
3.	(*v*) ING →	motoring → motor
		sing → sing

If rule 1b.2 or 1b.3 is successful then the following rules are tried:

#	Rule	Example
4.	AT → ATE	Conflat(ed) → conflate
5.	BL → BLE	Troubl(ing) → trouble
6.	IZ → IZE	Siz(ed) → size
7.	(*d and not (*L or *S or *Z)) → *single letter*	Hopp(ing) → hop
		Tann(ed) → tan
		Fall(ing) → fall
		Hiss(ing) → hiss
		Fizz(ed) → fizz
		Fail(ing) → fail
		Fil(ing) → file

12.2.3 Step 1c

#	Rule	Example
1.	(*v*) Y → I	Happy → happi
		Sky → sky

12.2.4 Step 2

#	Rule	Example
1.	(m > 0) ATIONAL → ATE	relational → relate
2.	(m > 0) TIONAL → TION	conditional → condition
3.		rational → rational
4.	(m > 0) ENCI → ENCE	valenci → valence
5.	(m > 0) ANCI → ANCE	hesitanci → hesitance
6.	(m > 0) IZER → IZE	digitizer → digitize
7.	(m > 0) ABLI → ABLE	conformabli →conformable
8.	(m > 0) ALLI → AL	radically → radical
9.	(m > 0) ENTLI → ENT	differently → different
10.	(m > 0) ELI → E	vileli → vile
11.	(m > 0) OUSLI → OUS	analogously → analogous
12.	(m > 0) IZATION → IZE	victimization → victimize
13.	(m > 0) ATION → ATE	predication → predicate
14.	(m > 0) ATOR → ATE	operator → operate
15.	(m > 0) ALISM → AL	feudalism → feudal
16.	(m > 0) IVENESS → IVE	decisiveness →decisive
17.	(m > 0) FULNESS → FUL	hopefulness → hopeful
18.	(m > 0) OUSNESS → OUS	callousness → callous
19.	(m > 0) ALITI → AL	formaliti → formal
20.	(m > 0) IVITI → IVE	sensitivity → sensitive
21.	(m > 0) BILITI → BLE	sensibility → sensible

12.2.5 Step 3

#	Rule	Example
1.	(m > 0) ICATE → IC	Triplicate → triplic
2.	(m > 0) ATIVE →	Formative → form
3.	(m > 0) ALIZE → AL	Formalize → formal
4.	(m > 0) ICITI → IC	Elcriciti → clectric
5.	(m > 0) ICAL →	Electrical → electric
6.	(m > 0) FUL →	Hopeful → hope
7.	(m > 0) NESS →	Goodness → good

12.2.6 Step 4

#	Rule	Example
1.	(m > 1) AL →	revival → reviv
2.	(m > 1) ANCE →	allowance → allow
3.	(m > 1) ENCE →	inference → infer
4.	(m > 1) ER →	airliner → airlin

5.	(m > 1) IC →	Electrical → electric
6.	(m > 1) ABLE →	gyroscopic → gyroscop
7.	(m > 1) IBLE →	defensible → defens
8.	(m > 1) ANT →	irritant → irrit
9.	(m > 1) EMENT →	replacement → replac
10.	(m > 1) MENT →	adjustment → adjust
11.	(m > 1) ENT →	dependent → depend
12.	(m > 1) and (*S or *T)) ION →	adoption → adopt
13.	(m > 1) OU →	homologou → homolog
14.	(m > 1) ISM →	communism → commun
15.	(m > 1) ATE →	activate → active
16.	(m > 1) ITI →	angulariti → angular
17.	(m > 1) OUS →	homologous → homolog
18.	(m > 1) IVE →	effective → effect
19.	(m > 1) IZE →	bowdlerize → bowdler

12.2.7 Step 5a

#	Rule	Example
1.	(m > 1) E →	probate → probat rate → rate
2.	(m > 1 and not *o) E →	cease → ceas

12.2.8 Step 5b

#	Rule	Example
1.	(m > 1 and *d and *L) → single letter	controll → control roll → roll