Australian Government
Department of Defence
Defence Science and
Technology Organisation

Effects of Instant Messaging on Recall During Video-Mediated Briefings

Derek Weber

Command, Control, Communications and Intelligence Division
Defence Science and Technology Organisation

DSTO-RR-0380

ABSTRACT

This report explores the effect of interruption from instant messaging (IM) on memory of a video-teleconference brief. A 4x2 factorial pilot study was conducted with 32 participants (gender balanced), using four levels of interruption and gender as the independent variables. Two videos were presented to participants, one of a single person speaking and one of a pair in conversation. Memory was tested with five minutes free recall after each video and recognition questionnaires 24 hours later. Analysis revealed that women performed better than men in free recall, that requiring participants to respond to questions degraded their free recall, and that interruption caused a significant downward trend in free recall performance and confidence. Recognition performance results were inconclusive, however. We recommend repeating this study with more participants. Results obtained could help advise Defence on the development of usage policies for collaboration technologies, specifically regarding limiting the types and number of sources of interaction and sensory input where possible.

Published by

Command, Control, Communications and Intelligence Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: (08) 7389 5555
Fax: (08) 7389 6567

AR-015-275
April 2012

APPROVED FOR PUBLIC RELEASE

Effects of Instant Messaging on Recall During Video-Mediated Briefings

Executive Summary

Effective collaboration is a vital component of the success of any team, especially in a military context where lives may depend on the outcome of any collaborative effort. As part of its research in support of collaboration, the Defence Science and Technology Organisation (DSTO) has run a research activity to develop and support 'smart' meeting rooms called Livespaces. When installed in operational military environments and equipped with domain-specific tools, they are known as Command and Control Developmental BattleLabs (Evdokiou et al., 2004, Vernik et al., 2003).

The Livespaces software (Phillips, 2008) that underpins these spaces provides not only control of devices and information services but also forms a foundation for collaboration software, such as tools for instant messaging (IM), collaborative document editing, and screen sharing. The primary mechanism for group communication between BattleLabs is software-based video teleconferencing (VTC). There are many other technologies available in these environments and it is imperative to investigate how distractions from these technologies affect productivity. The literature on attention, interruption, and dual-task performance strongly suggests that performance will degrade with increases in distraction.

This report investigates how increasing levels of interruption from IM, as a secondary task, may affect memory of short video-mediated briefs, such as a Commander's brief delivered via VTC in a Livespace. It extends upon a Master of Sciences project. We hypothesised that:

1. performance in recall of verbally delivered content would decrease with increasing interruption,
2. confidence in recall of verbally delivered content would decrease with increasing interruption,
3. women would have better recall performance than men, and
4. experience with technology would ameliorate the decrease in recall performance and confidence.

A 4x2 factorial design experiment was conducted using four levels of interruption from IM and two levels of gender. Thirty two participants (16 female) were assigned to gender-balanced groups, and each group received a different degree of interruption. Participants watched two 5-minute videos as a primary task, while receiving various degrees of IM as a secondary task, and subsequently performed free recall and recognition tests on the content of the video. They were also asked for confidence ratings on their recollections. One video was of a single person telling stories, and the other video was of a couple

discussing travel plans, and the IM content consisted of old (irrelevant) news headlines and questions regarding the video content.

Analyses revealed that interrupting participants with questions significantly degraded their recall, that their confidence in recall significantly diminished over time, and also that women had statistically better free recall than men,. Statistically significant downward trends in free recall performance and confidence with increases in interruption were also found. Recognition data was inconclusive, even after rescoring discarding 'unsure' responses, and neither experience with technologies nor age had significant effect on recall performance or confidence. Although not statistically significant, the increase in false positives with increasing interruption may be of particular interest for future studies: participants receiving questions were worse at spotting false statements regarding the content of the videos.

Even with such a small sample size, these results suggest that attempting to engage in IM while listening to someone speak will mean that a significant portion of what is said may be missed, and thus this practice should be avoided. At the very least it may affect confidence in what is recalled. IM may be used as a secondary task for maintaining awareness via broadcast messages that require no response even if the first task is verbal in nature, but requiring a response diverts too much attention to maintain performance on the primary task, when it is a verbal one.

As a pilot study, the results obtained suggest that further investigation is warranted. If interruption from IM, and other technologies, affects memory and confidence in that memory, it is important that Defence develop usage policies that minimise those effects. These policies may be trialled in Livespaces in particular, at least initially, but their use may extend beyond to other collaboration environments used by Defence. Literature on the effects of animation and colour on attention also may suggest revision of policies on the use of combined tickertape and video displays (e.g. CNN news feeds) in work environments.

Author

Derek Weber
Command, Control, Communications and Intelligence Division

Derek Weber is a Defence Scientist in the Human Interaction Capabilities Discipline and has a Master of Sciences (Defence) from the University of Adelaide and a Bachelor of Sciences (Honours) in Computer Science from Flinders University. He has interests in cognitive science, neuroscience, language and evolutionary biology and has worked at DSTO for over 10 years.

Derek's primary recent work has concerned the Livespaces framework, an implementation of a tuple space and model-driven service architecture to support collaboration through meeting room device control and provision of a communications layer for groupware. His current work includes research and development into information retreival supported by visual analytics. Previously Derek worked on a component-based information visualisation framework called InVision, and before that on a distributed imagery and geospatial information exploitation system.

This page is intentionally blank

Contents

1. Introduction

This report details two experiments conducted concurrently in a pilot study. The study investigated the effect of using instant messaging while attempting to pay attention to two videos. The first video was of a single person telling stories, simulating a military brief, and the second was of a couple discussing travel plans, simulating a decision-making interaction. The purpose of the experiments was to explore and investigate the effects of verbally-based distractions and interruptions on individuals attending to a primary verbal task. The report outlines the background of the work, the methodology used, the results obtained, and their analysis, and what the results mean for improving the work practices of knowledge workers in disruptive environments.

1.1 Background

Effective collaboration is a vital component of the success of any team, especially in a military context where lives may depend on the progress and outcome of any collaborative effort. DSTO has been conducting research to develop and support 'smart' meeting rooms called Livespaces (Phillips, 2008), also known as Command and Control Developmental BattleLabs once installed in operational military environments. These environments, although non-operational in nature, provide opportunities for DSTO's military stakeholders to experiment, evaluate, and innovate with tools and methodologies not available in their standard operating environments. This work is related to Interactive Intelligent Workspaces (Vernik et al., 2003) and was performed in the field of Computer Supported Cooperative Work (CSCW). Its original domain application was in support of distributed synchronous collaborative activities of operational planning staff (Evdokiou et al., 2004). Since then, the technology has matured and the research has broadened to investigate distributed intense collaboration in general and mixed presence collaboration through the HxI Initiative[1], a collaboration between Australian government-funded research organisations (Bezerianos and McEwan, 2008, Müller-Tomfelde et al., 2007, Schremmer et al., 2007).

The Livespaces software (Phillips, 2008) provides not only control of devices and information services but also a foundation for collaborative software, such as instant messaging (IM), collaborative document editing, and screen sharing. The primary mechanism for group communication between BattleLabs is software-based video teleconferencing systems (VTC) provided by third party tools such as AccessGrid[2] and Microsoft's ConferenceXP[3]. As technology costs diminish and the push to reduce travel increases, it is expected that more information (e.g. instructions and reports) will be disseminated via multi-cast and personalised video links.

A major limitation of the typical BattleLab VTC systems is that there is only one convenient audio channel between remote sites, and therefore only one conversation can take place at a

[1] http://www.hxi.org.au, last accessed 12th April 2010.
[2] http://www.accessgrid.org, last access 12th April 2010.
[3] http://research.microsoft.com/en-us/projects/conferencexp/, last accessed 12th April 2010.

time using this medium (Mark et al., 2003). Side-bar conversations via the shared audio channel are simply not possible as they immediately interfere with the audio of the original conversation and interrupt its flow. In contrast, for example, if all participants are collocated for a meeting, one can nudge the person next to them and whisper a question, such as "What was that acronym?" without significantly disturbing others. Sticker[4] is an IM tool available in Livespaces, which appears highly suitable for such 'whispering'. However, we hypothesise that task interference will occur when attending to IM as well as the conversation in the meeting.

The issue of overlapping spoken and written communication tasks relates to the broader issue of multi-tasking and is relevant to many military environments. Very often, especially while acting in monitoring and watchkeeping roles, military personnel must respond to stimuli transmitted over many different media. These media include spoken directions, background announcements, sirens, visual alerts such as flashing lights, in addition to electronic stimuli from computers such as IM and e-mail. They must respond within strict time-limits and perform to high levels of accuracy for extended periods of time. Some roles require this attention to detail while others require attention to gist only. Both need varying degrees of precision, but distraction, which causes false positives and false negatives, is a concern for both. More research is required to determine how to best manage distractions and interruptions while maintaining acceptable levels of effectiveness.

Much of the content of this report was originally prepared as part of the author's Master of Sciences (Defence) thesis[5] (Weber, 2008). Further examination and analysis of the data has occurred since the initial report was submitted, and this extra work is detailed in this report.

1.2 Summary Review of the Literature

Collaborating in distributed situations is notably different from, and less effective than, being collocated (Mark et al., 2003). The research area of awareness in CSCW is predominantly focused on supporting people in such situations (see Rittenbruch and McEwan, 2007, for a thorough history). Within the CSCW research community IM has often been considered for implementing notification and communication systems because of its convenience and prevalence (Fuchs et al., 1995, Lövstrand, 1991, Smale and Greenberg, 2005). Some implementations have been especially suitable for experimentation due to the disconnected nature of their underlying publish/subscribe architectures, in particular Elvin (Fitzpatrick et al., 1999, Parsowith et al., 1998).

1.2.1 Instant messaging is distracting

Almost as soon as IM began to be touted as a mechanism to maintain awareness of others, human factors researchers quickly noticed and began to analyse its distracting effect on users. The cost of distraction due to IM can vary depending on the type of the task. Also the strength of the distraction can vary depending on the point within a task at which the distraction

[4] http://www.tickertape.org:8080/projects/sticker, last accessed 12th April 2010.
[5] This Masters was executed through the School of Psychology of the University of Adelaide.

occurs (Adamczyk and Bailey, 2004, Bailey and Konstan, 2006, Czerwinski et al., 2000a, Czerwinski et al., 2000b). Overall, many researchers report significant performance degradation in terms of speed and accuracy with distraction (Cole et al., 2006, Czerwinski et al., 2000a, Medina, 2008, Ramsey et al., 2004). There is the danger that IM will draw too much attention as a secondary task (Cummings and Guerlain, 2004). Also a great deal of criticism has been levelled at the modern notion of constant availability via electronic means like IM (Allen, 2002, Fried, 2006, Rennecker and Godwin, 2003, Song et al., 2007). However this is not a new phenomenon: Gaver (1991) laments, "it is difficult to get work done when constantly in meetings about work" referring not just to electronic distractions but also in-person distractions. Amongst the empirical evidence for the distracting nature of IM, only a few studies (examples including Cummings and Guerlain, 2004, Knott et al., 2006a, Knott et al., 2006b) have examined the use of IM in a specific domain — most are focused on general tasks (Bailey and Konstan, 2006, Czerwinski et al., 2000a, Czerwinski et al., 2000b, Luo and Wu, 2007).

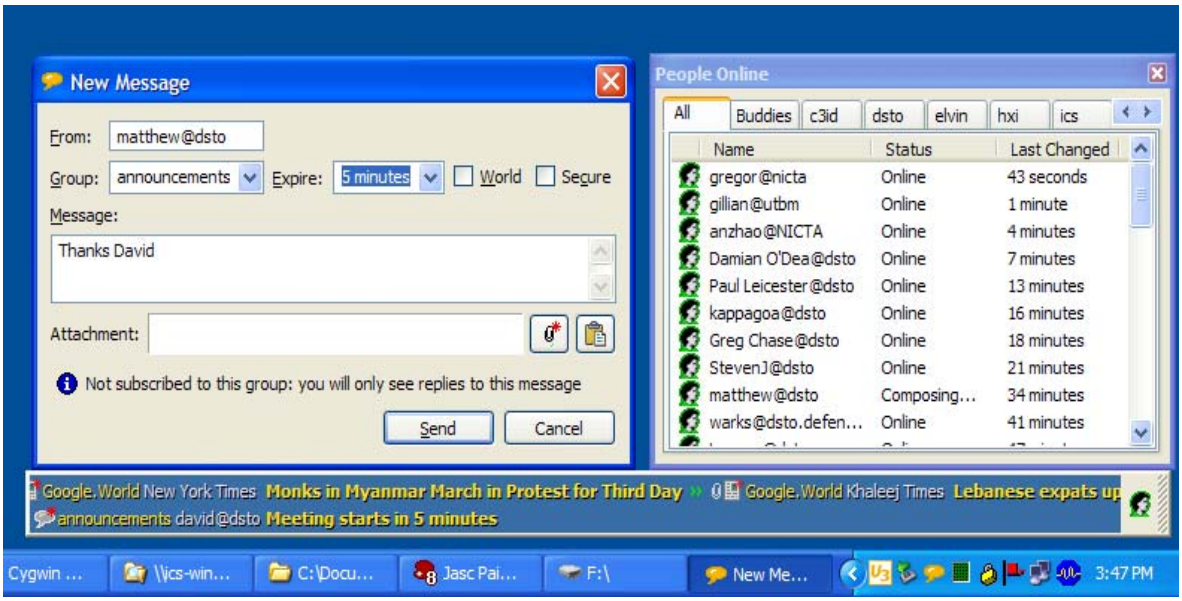


Figure 1 The Sticker IM tool in use. The tickertape display is shown hovering just above the Windows XP Task Bar, with a new message window open on the left and the Presence window (showing information about which users are currently online) on the right above the tickertape display.

To display incoming instant messages and other notifications Sticker uses a tickertape display (Figure 1), also referred to in the literature as "rapid, serial, visual presentation" or RSVP (Potter, 1984). This is similar to notification displays on some websites and news channels (e.g. CNN). The effect on recall when using tickertape displays in combination with video footage seems not to have been considered in the literature. However, in terms of animation, research has found flashing advertisements and animation do not improve, and may even degrade, recall of website content in some circumstances (Day et al., 2006, Hong et al., 2004). Similarly, there is a body of research on the use of peripheral animation and icons for transmitting information (Bartram, 2001, Bartram et al., 2003, Kang and Muter, 1989, Lee et al., 2002, McCrickard et al., 2001).

1.2.2 Attention and dual-task research

The fundamental resource in question with respect to distraction is attention, which has been empirically known to be limited for decades. The cognitive mechanisms underlying human attention have been of interest to psychology since William James' research in the late 19th Century. There has been particular interest in tasks that we have difficulty executing simultaneously, such as those examined in the Stroop colour test (Stroop, 1935). Welford (1952) postulated the *central bottleneck theory* to explain the *psychological refractory period* (PRP) first observed by Telford (1931). The PRP is the brief but measurable period during which a person cannot initiate a second task while occupied with a first. Since then research has all but conclusively confirmed that humans can only perform one cognitive task at a time and that there is a cost to switching and resuming tasks (Altmann and Trafton, 2007, Dux et al., 2006, Monk et al., 2008, Pashler, 1992, Pashler, 1994, Ramsey et al., 2004).

Under very specific circumstances, such as with well-practised and ideomotor-compatible tasks, the bottleneck appears to be reduced (Greenwald and Shulman, 1973, Hazeltine et al., 2002, Schumacher et al., 2001). Alternative cognitive models have been proposed to explain these observations (Meyer and Kieras, 1997a, Meyer and Kieras, 1997b, Meyer et al., 1995). The phenomenon is rare, however, and there is not yet conclusive evidence indicating that the bottleneck does not, in fact, exist. It could still exist but some tasks appear to be scheduled better by the brain than others, giving the appearance that they are being processed concurrently (Lien et al., 2006). This notion is called the *latent bottleneck hypothesis*. The implications of this effect are particularly evident when considering its potentially hazardous effect while simultaneously driving and talking on a mobile phone (Strayer et al., 2006, Strayer and Johnston, 2001) and when coupled with the phenomenon of *attentional blink* (Motluk, 2007). These findings contradict the modern conventional wisdom that younger, technology-savvy people are able to multi-task (Cole et al., 2006, Ophir et al., 2009).

Furthermore, the time taken to resume the primary task can vary depending on the nature and degree of the secondary task, including importantly when it is an externally-driven distraction instead of an internally-driven deliberate change in task (Altmann and Trafton, 2007, Monk et al., 2008). For example, the practice of attempting to read a map while driving might affect the eye's focal length as well as the focus of spatial cognition while orienting oneself to a map or to the road. Reorienting those skills back to driving may take longer when compared with other distractions, such as reacting to loud noises (e.g. truck air brakes) which do not require changing the direction of gaze or redirecting spatial cognition.

1.2.3 Mechanisms for delivery of information

This work also contributes to a continuation of the work of Dobson-Keeffe (2007) who tested participants' recall against the condition of whether the information was delivered by a co-located human messenger or an animated virtual advisor (Taplin et al., 2001). Dobson-Keeffe found that participants' confidence in what they were told by a human was overestimated. In fact, their recall performance was the same regardless of the mode of information delivery. He also found that there was a difference in recall between the genders, but only in the first of two treatments (with women performing better in free recall, as supported by Kimura, 2000).

By examining the situation of information being delivered via a (simulated) video link, we aim to bridge the gap between the human messenger and the virtual advisor.

1.2.4 Memory retrieval

Memory research is very broad and mature, providing a rich understanding of memory systems within humans and other animals. It is known to have several temporal components (sensory, short-term, and long-term) and a large number of types of memory have been identified by cognitive, neuropsychological and neuroimaging evidence (Gazzaniga et al., 2002, Medina, 2008, Yonelinas, 2002, Zimbardo, 1992). Furthermore, the brain cell structures supporting memory (along with the rest of brain function) are known to be common across a wide variety of distantly related species from the *Aplysia* sea slug to humans (Kandel, 2006). However, in this study, the main target of our interest is the mechanisms underlying memory performance and retrieval, to be used as a dependent measure for our experiments. The *dual-process interpretation* of the observed data is the most widely accepted; it states that retrieval consists of two dissociable functions: *recollection* and *familiarity*. Many *dual-process* models predict two separate neural substrates for the two functions and neuropsychological and neuroimaging data support this to a significant degree (Rugg and Yonelinas, 2003, Yonelinas, 2002). However, the data do not rule out alternative simpler models, such as the *signal detection theory* that suggests that the different functions are in fact the same function but with two thresholds (essentially providing a measure of confidence), requiring the existence of only one substrate (Dunn, 2004).

As an approximation recollection can be tested as unprompted free recall and familiarity can be tested as recognition (Yonelinas, 2002), although because they are different functions it makes little sense to compare recognition and recall directly. Together they provide a measure of memory retrieval performance.

Of further interest is that the literature suggests that women have better recall than men, in particular in verbal memory (Kimura, 2000). It is not clear if this should affect familiarity as much as recollection. This finding was replicated in the first of two experimental treatments by Dobson-Keeffe (2007).

1.3 Hypotheses

On the basis of the review of the literature, a number of hypotheses were proposed and are stated here.

- Recall and recognition performance of verbally delivered content would deteriorate with increasing levels of interruption from IM, and this effect would be greater for messages requiring a response.
- Confidence in recall and recognition of that content would decrease in a similar fashion.
- Women would perform better than men at recall and recognition, but men would have higher confidence in their recall and recognition in all circumstances.

- Experience with technologies such as IM, mobile phones' Short Messaging Service (SMS), VTC, and tickertape displays, as well as simply having an affinity for technological gadgets, would result in better recall and recognition performance.

2. Method

2.1 Design

The experiment followed a 4x2 factorial design with the factors of IM LEVEL and GENDER. The four levels of IM LEVEL were NONE, NOISE, QUESTIONS, and QUESTIONS + NOISE (explained in the next paragraph) and the two levels for GENDER were MALE, and FEMALE. Two experiments were run simultaneously, each with a different video (order balanced). We partially controlled for experience with a variety of technologies through the use of the pre-experiment questionnaire (Appendix F). Using the questionnaire answers, participants were ranked in order of experience with technologies, and were then assigned to different IM groups on a round robin basis. Four IM groups were formed to represent the IM LEVELs, each with four male and four female participants.

Participants in IM group NONE received no interruption in the form of IM and acted as a control group. Each participant in IM group NOISE received between 23 and 26 news messages during each video, while participants in group QUESTIONS received 9 or 10 questions during each video, all of which required responses. Participants in IM group NOISE + QUESTIONS received news messages and questions messages in the same manner as participants in IM groups NOISE and QUESTIONS. The researchers estimated that the degree of increase in interruption between each pair of adjacent IM LEVELs was similar and therefore the IM LEVEL factor could be regarded as ordinal[6].

2.2 Participants

Of the 37 people who volunteered and participated in the experiment, the results of five were discarded due to technical difficulty or participant error. All participants were DSTO staff with roles in engineering, science, and administration and their ages ranged from 24 to 56 (M = 34.5). Gender was balanced across the 32 participants whose data were analysed. All participants provided their informed consent and were offered a chocolate bar, a packet of crisps or a soft drink as a token of gratitude for their help.

Ethics approval was sought and obtained from the DSTO's Science and Technology Activity Review team before seeking participants. The project information sheet, consent forms, and guidelines for volunteers can be found in Appendices K through to O.

[6] The suggested order for increasing levels of interruption is NONE, NOISE, QUESTIONS, QUESTIONS + NOISE.

2.3 Materials

On the day of the experiment each participant was provided with a personal computer with individual headphones[7] at which to work, and pens and paper for free recall. Prior to the experiment they received questionnaires for gathering experience data, and after the experiment they received questionnaires for gathering recognition data. Experimental sessions were conducted in a Livespace 'smart' meeting room, which had seven personal computers available for participants (Figure 2). This "real life" collaboration environment, with common distractions from colleagues such as coughing, sniffling, and shuffling of papers, affords increased ecological validity to the experiment. These aspects would have been excluded if participants were tested in isolation.

Two 5-minute videos were used to represent VTC. The "Stories" video consisted of a mature man telling stories 16 and 17 from Dixon, Hultsch and Hertzog's collection of memory testing resources (1989)[8]. The "Conversation" video involved the same man along with a mature woman discussing plans for an overseas holiday (see Appendix C for the transcript).

Dixon et al. (1989) includes a marking guide for scoring free recall of their stories, as well as recognition questionnaires. A similar marking scheme was developed by the researchers for scoring the free recall of the "Conversation" video (Appendix E), and a similar recognition questionnaire was devised according to descriptions in Dixon et al. (1989), but with the same modifications made by Dobson-Keeffe (2007). These included a 7-point Likert scale for each item rather than simple binary (yes/no) options in order to gather information on confidence (Appendix I). Although there was a 24-item recognition questionnaire for each of the two stories (Appendices G and H) contained in the "Stories" video, only one 24-item questionnaire was developed for the "Conversation" video. This was done because the propositional density of the "Conversation" video was not as great as that of the "Stories" video, and therefore developing a 48-point questionnaire would have been very difficult.

[7] The experiment environment was the Intense Collaboration Space Livespace, used because of the convenience of simultaneous access to seven workstations.
[8] Stories 16 and 17 of Dixon et al. (1989) are included in Appendices A and B of this report.

Figure 2 A Livespace 'smart' meeting room (the Intense Collaboration Space)

Each 24-item questionnaire consisted of eight true gist statements, eight false gist statements, and eight elaboration statements[9]. The statements for the "Conversation" questionnaire were formulated by examining each 30-second period of the 5-minute video (resulting in 10 periods). True gist, false gist, and an elaboration statement were developed based on the proposition content of each period. Of the 30 resulting statements, eight of each category were selected on the basis of similarity in difficulty to the statements in the "Stories" recognition questionnaires.

Two IM groups (QUESTIONS and QUESTIONS + NOISE) were sent questions via the Sticker IM application during their videos. These questions were developed in a similar manner to the "Conversation" recognition statements (see Appendix J for these questions). Both videos were divided into 30-second periods, and questions relevant to each period were developed (these were explicitly different to the recognition statements), resulting in 10 questions for each video. News messages sent as a distraction to participants in IM groups NOISE and QUESTIONS + NOISE were genuine news headlines in Australia collected in August 2008. The intention was to make them uninteresting and not related to the video, hence the label 'noise'. Participants were not required to respond to them nor remember their content, so they simply acted as an on-screen distraction.

Participants' typing speed and accuracy were tested with a demonstration copy of the TypingMaster typing application, version 6.30.

The pre-experiment questionnaire consisted of nine items, each with a 6-point Likert scale (0 = Never used, 6 = Use all the time) for determining experience with a variety of technologies and an affinity for technological gadgets (Appendix F).

[9] Story 17 had nine false gist statements and seven elaboration statements, sourced without modification from Dixon et al., 1989.

A modified version of Sticker (version 3.1.0rc2) was used in the experiment for IM. Timing information emitted by the modified Sticker applications was collected using the TeamScope (Evdokiou, in preparation) tool. The modifications made to Sticker by the researcher allowed TeamScope users to record when messages arrived, when messages were clicked on by a user wanting to respond to them, and when a response was sent. This provided the ability to estimate how much time was spent attending to answering a message rather than on the primary task of watching the video.

2.4 Experimental Tasks

The main experimental task required participants to watch a video as a primary task and respond to instant messages on Sticker (IM) as a secondary task. Participants were asked to respond to question messages as soon as they saw them. After each of the two videos, each individual performed the distraction and then the free recall tasks. Participant performance was determined by testing their free recall immediately after each video (after a two minute arithmetic-based distraction task) and via recognition questionnaires the following day.

2.5 Procedure

The procedure consisted of three periods, those being prior to the experiment day, the day of the experiment, and the day after the experiment. Prior to the day of the experiment, participants were asked to fill out the experience questionnaire. This was done some days before the experiment day to use the results to group individuals.

As part of the volunteering process, participants were assigned a unique two-digit identifier (ID) to maintain their anonymity. These IDs were applied to all data collected and only the primary researcher held the mapping between participants' names and their IDs. Participants were assigned to a group, controlling for experience with IM and SMS, as it was hypothesised these would have the most impact on performance with the Sticker application. Participants were asked to attend a particular experiment time-slot. Effort was made to accommodate their preferences, and most experimental sessions included participants from different groups.

At the allotted time of their experiment, participants attended the Intense Collaboration Space room, where they were given a small card with their ID on it and asked to log into one of the seven personal computers. Participants were provided with information regarding the experiment and its context within the project and then provided with an opportunity to ask questions and provide their consent to participate. Each participant carried out a two minute typing test with the default text (Aesop's fables) in the TypingMaster tool. Participants were then trained in how to respond to messages in the Sticker application. The researcher sent a Sticker message to the 'Chat' channel, to which each participate was already subscribed. Participants were then asked to respond to this message, and then to another as practice. In order to avoid participants in IM groups NONE and NOISE worrying that others were typing while they were not during the experiment, they were told that some participants would receive news messages, which required no response, some would receive questions, which would require an immediate response, some would receive both, and some would receive

none. Although a common chat group was used in training participants with Sticker, messages during the experiment were sent to participants privately so there was no opportunity for participants to see each others' messages. As a final preparation for the main experimental task, participants donned their headphones and adjusted their computers' volume levels, being warned to increase the volume as the audio tracks of the videos were quiet.

Each participant launched the main experiment application from a desktop shortcut. This asked for their ID and then showed them the first 5-minute video. For those participants receiving questions, question messages arrived every 30 seconds, starting 30 seconds into the video. For those participants receiving news messages, news messages began 15 seconds into the video and were sent semi-regularly at pre-randomised times with gaps between messages of 5 to 20 seconds. At the conclusion of the video, participants undertook the distractor task consisting of simple arithmetic problems, consisting of single-digit multiplication and addition for two minutes. Immediately following the distractor task, each participant wrote as much as they could remember from what was said in the video within a 5-minute time limit. Participants were then asked to provide a Likert-style confidence rating for each statement or bullet point they had written (1 = I'm not sure this is true, 2 = This is possibly true, 3 = This is probably true, 4 = This is definitely true). This was required for comparison with the confidence ratings in the post-experiment questionnaires. Participants then carried out the same steps for the second video.

At the conclusion of the videos and recall, participants were informed of the purpose of the arithmetic as a distraction task (as required for ethics approval) and given an envelope with the recognition questionnaires to be filled in the following day 24 hours from then.

2.6 Measurement

The dependent measures collected during the experiment were the free recall and recognition scores, the confidence ratings for the free recall and recognition statements, and the time spent by participants attending to question messages (estimated using the time between when they clicked a question to respond to it and when they pressed the 'Send' button). The other measures were the individual's subjective experience ratings gathered from the first questionnaire, the recognition scores and confidence ratings from the post-experiment questionnaires (calculated from the 7-point Likert answer scale for each question) , and participants' typing speed and accuracy.

3. Results

It is important to note again that this report describes the results from Weber (2008), along with the results from rescoring of the data. The rescoring pertains to the recognition task, discounting the 'unsure' option from any responses[10], which yields new recognition scores and counts of valid answers for each participant. These contribute to our understanding of recall performance and confidence. Before explaining the results, however, an overview of the analysis procedures are provided.

3.1 Analysis Procedures

Our hypotheses focused primarily on various effects on recall performance and confidence across the IM LEVEL and GENDER factors. The 4x2 factorial design of our two experiments provided us the opportunity to use repeated measure analysis of variance (ANOVA). Combining IM LEVEL groups, the factors QUESTION LEVEL (those who received questions and those who did not) and NOISE LEVEL (those who received news messages and those who did not) were also analysed using 2x2 repeated measures ANOVAs. Furthermore confidence ratings were considered by the factor of RECALL DAY (ratings from the day of the experimental session based on the free recall data, and those from the next day, based on the recognition data). Further information on confidence was found after rescoring recognition data when 'unsure' answers were regarded as invalid and discounted: the proportion of valid answers was taken to reflect confidence. Variables of VIDEO and recognition task GIST were also used as factors in the analyses performed: VIDEO refers to which of the two videos were viewed as part of the treatment, "Stories" or "Conversation", and GIST refers to the gist of particular statements in the recognition task, roughly equal numbers of true gist, false gist, and elaboration statements.

As mentioned, the post-experiment questionnaires consisted of 7-point Likert scales for testing recognition. A consequence of using 7-point Likert scales was the ability to gauge participants' confidence in their answers, resulting in a 4-point confidence scale with minor arithmetic (1 = I'm not sure this is true, 2 = This is possibly true, 3 = This is probably true, 4 = This is definitely true). The 7-point answers ranged from 1, or "Definitely true", to 7, or "Definitely false", with 4 in the centre being "Unsure". Thus selections of 1 and 7 were regarded as highly confident (4 on the 4-point scale) and a score of 4 was not at all confident (1 on the 4-point scale). Participants were also asked to apply the same 4-point confidence rating to the statements they wrote in their free recall, providing a mechanism to compare their confidence across recall and recognition tests. This scoring procedure was used to ensure that our initial results would be comparable with those of Dobson-Keeffe (2007) and are shown in Table 1.

For reasons described in Section 3.2.1 the recognition data was rescored, discounting all 'unsure' answers (option 4 on the questionnaire). Although the recognition procedure became more correct, removing the ambiguity of being 'unsure' if recognition occurred or not, it

[10] The reasons for discounting the 'unsure' option are explained in Section 3.2.1. In essence, an 'unsure' option makes it unclear whether recognition has taken place (Strauss et al., 2006)..

affected the calculation of confidence scores, leaving only three levels of response (2-4 in the scheme described above). These confidence scores could have been recalculated to fit within the 1-4 scheme, but the effort was not deemed worthwhile because the scores would not have been significantly different. The confidence scores would have been slightly lower than what they are now (at most by a third of a point), and given the variation in number of valid answers (i.e. ones that were not 'unsure') any mean values calculated could have large standard distributions rendering their precision too low to draw conclusions from. Instead, measures of confidence were drawn from the proportion of valid answers. Although not as fine-grained as the previous measure, it still provided some insight into the confidence of participants. Another effect of removing 'unsure' answers is that it rendered elaboration and false gist statements in the questionnaires to be scored the same, so questionnaires effectively consisted of two thirds false gist statements and one third true gist statements. This may affect participants' scores if there is a bias towards scoring better with either type of gist.

Repeated measures ANOVA was used because the repeated component was the participant, and the measures were comparable (the recall scoring procedure applied to recall from watching a 5-minute video). Differences between IM LEVEL stimuli were designed to be ordinal, in the order NONE, NOISE, QUESTIONS, QUESTIONS + NOISE.

An alpha level of 0.05 was used for all analyses, although where a significance of less than 0.001 was observed, this is reported explicitly.

Table 1 Scoring procedure for recognition task. Responses on the 7-point Likert scale can be used for scoring and measuring participant confidence simultaneously. Note that after rescoring the centre option of 4 was discounted and so elaboration and false gist statements were scored identically.

Point for True Gist					⌒	⌒	⌒
Point for False Gist	⌒	⌒	⌒				
Point for Elaboration	⌒	⌒	⌒	⌒			
Response	1	2	3	*4*	5	6	7
Meaning	Definitely false	Probably false	Possibly false	*Unsure*	Possibly true	Probably true	Definitely true
Confidence	High (4)	Moderate (3)	Low (2)	*Very low (1)*	Low (2)	Moderate (3)	High (4)

3.2 Effects on Recall Performance

This section describes the analyses performed to test the first and third hypotheses, namely that increasing IM Level would degrade recall performance, and that women would have better recall performance than men, and is partially sourced from Weber (2008). This includes analyses of the recall and recognition data, specifically dealing with the rescored recognition data, which is a new contribution. Also dealt with are the effects of logical groupings of

participants by QUESTION LEVEL and NOISE LEVEL, and the investigation into the effect of the content of questions on the recall of those participants who received questions.

Two 4(IM LEVEL) x 2(GENDER) x 2(VIDEO) ANOVAs were conducted initially, one using the free recall scores as the measure and one using the recognition scores as the measure. The effect of GENDER on free recall was significant, $F(1,24) = 5.692$, $p < 0.05$, with women performing better than men (Figure 3a and Figure 3b). Also, although a trend can be seen in free recall performance by IM LEVEL (decreasing with increased levels of IM), a separate 4(IM LEVEL) x 2(VIDEO) ANOVA was required to find that it was significant (linear, $p < 0.05$, Figure 4). The only significant effects on recognition score were across the factor of VIDEO, $F(3,24) = 1117.538$, $p < 0.001$[11], and an interaction between GENDER and IM LEVEL, $F(3,24) = 3.635$, $p < 0.05$, though this disappeared with rescoring and reanalysis, the detail of which is discussed below. No other factors were found to be significant. The graph of the initial (pre-rescoring) recognition data against IM LEVEL, with GENDER plotted as separate columns (similar to Figure 5 but not shown) drew our attention to the IM group QUESTIONS. We believed it caused the significant interaction effect because the performance of GENDER appeared strongly reversed in comparison to the other IM groups. After rescoring it still appears reversed, but with males and females scoring ~73% and ~65% respectively instead of males and females scoring ~65% and ~50% respectively (halving the difference between the genders). No particular difference could be found in the members of the IM group QUESTIONS compared to the others, except in recognition score, and, given rescoring removed the significant interaction effect, it is highly likely the significance was due to the scoring procedure.

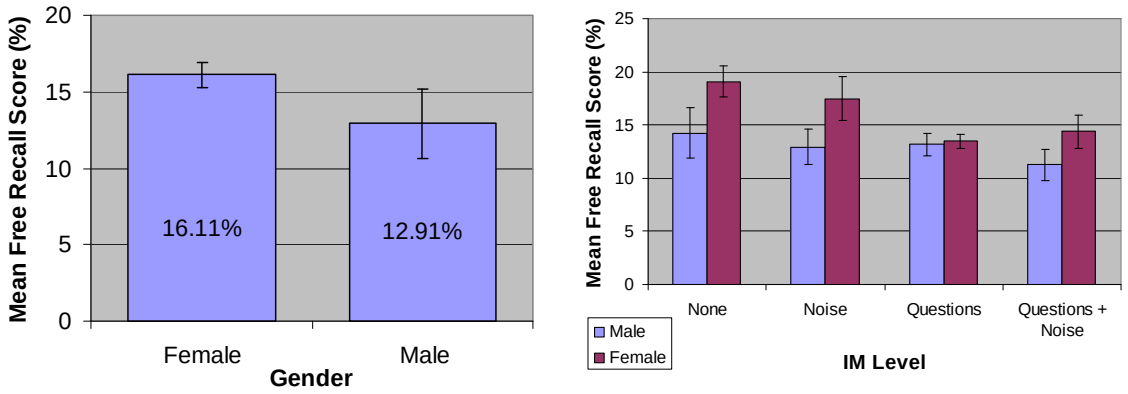


Figure 3 Analyses of mean free recall scores by GENDER and by IM LEVEL with GENDER plotted in separate colours shows (a) women scored better than men overall and (b) women scored equally well and better than men consistently across groups. (Error bars are +/- one standard error)

[11] This effect was not unexpected. The treatments were as similar as could be made with the resources available, but they are not as similar as, e.g., the different stories in Dixon et al. (1989).

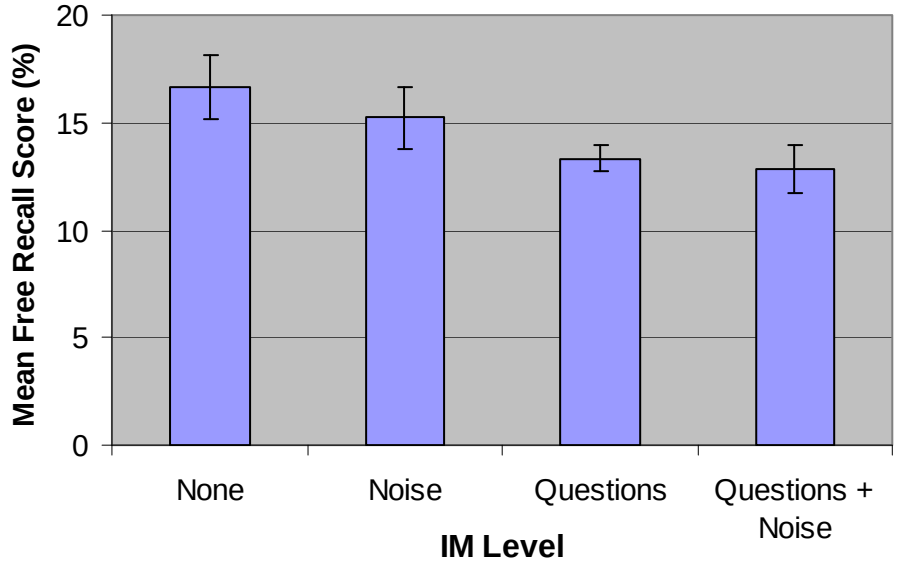


Figure 4 *Analysis of mean free recall scores by IM LEVEL shows the significant trend found between IM groups is downward, indicating performance decreases with an increase in interruption. No significant effect of IM LEVEL was found, however. (Error bars are +/- one standard error)*

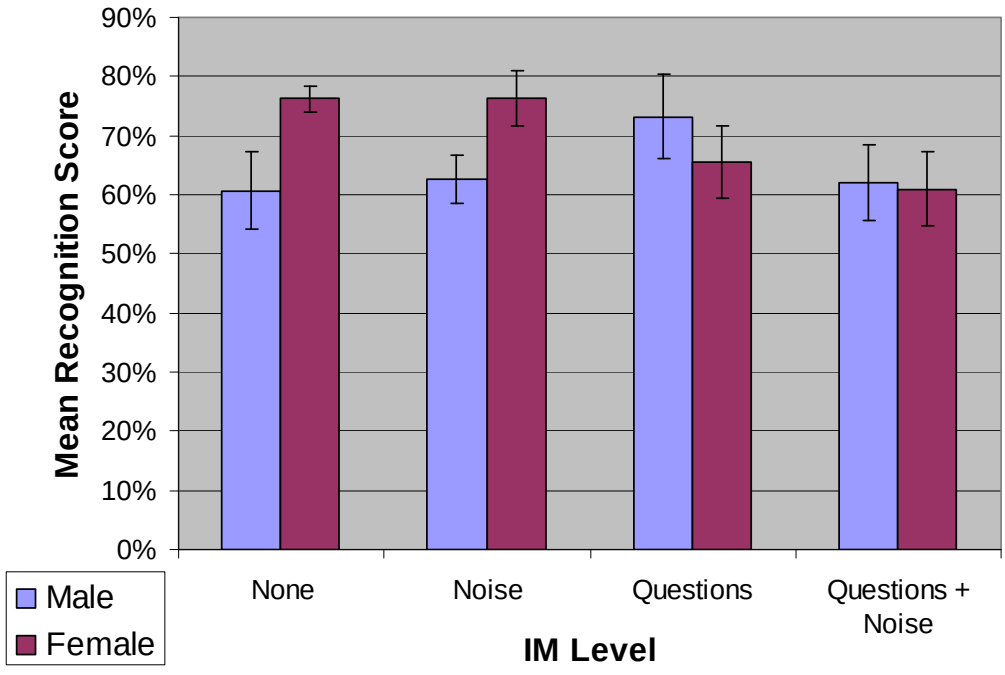


Figure 5 *A 4x2 repeated measures ANOVA initially revealed a significant interaction between IM LEVEL and GENDER across mean recognition scores. This result disappeared after rescoring and reanalysis to discount 'unsure' answers. The revised data are plotted here. (Error bars are +/- one standard error)*

3.2.1 Reanalysis of recognition data

While presenting preliminary results to staff at the University of Adelaide's School of Psychology, the result of the interaction between GENDER and IM LEVEL in the recognition data (rescored data shown in Figure 5) was discussed in some detail and a number of suggestions were made for further analysis. It was noted that recognition performance drops off considerably over periods of days (Medina, 2008) and given it was known that many participants (up to half) did not complete the recognition task on time, this may have affected the initial results. No information was captured regarding exactly when the recognition task was carried out by participants, so this expected degradation in performance could not be accounted for.

Another suggestion was that the recognition tasks should not have included an 'unsure' option. Recognition tests rely on participants providing a definite answer of recognition (Strauss et al., 2006). The confidence in an answer can be gauged by the rest of the Likert score (responses 1-3 and 5-7), but the inclusion of 4 as an 'unsure' response makes it unclear whether recognition has occurred or not. The data was rescored and reanalysed, discarding all 'unsure' answers. Given participants answered 'unsure' to different numbers of questions, each of their mean recognition scores became based on a different number of answers, varying from 22 to 71 answers out of a possible 72 (48 for the "Stories" video and 24 for "Conversation" video). The mean was 43.9 valid answers (just under 61%) per participant with a standard deviation of 10.8. Even when considered by gender, between group differences were small compared with within-group variation resulting in no statistically significant effects; women had a mean of 42.4 valid answers ($\sigma = 12.5$) compared to men's 45.1 ($\sigma = 8.7$).

It was noted that IM groups NONE and NOISE had higher mean recognition valid answer counts (45.5 and 46.25 respectively) than IM groups QUESTIONS and QUESTIONS + NOISE (41.25 and 42.0 respectively), and that males had a higher mean recognition valid answer count of 45.1 compared with females' 42.4 out of 72 questions. However, there was no statistically significant difference across IM LEVEL or GENDER or QUESTION LEVEL as might be suspected. This is discussed in more detail in Section 3.3. Analysis of the rescored recognition data revealed the interaction effect between GENDER and IM LEVEL had gone, $F(3,24) = 2.543$, $p > 0.05$, and scores improved evenly across all participants.

Although the interaction disappeared, it was interesting to note the relative improvements of males and females when videos were considered individually. Although GENDER did not have a statistically significant effect (according to a 2(GENDER) x 2(VIDEO) ANOVA), the fact that females improved approximately 20% compared with males 10% in the "Conversation" video (Figure 6b), yet both sexes' scores improved by only approximately 5% each in the "Stories" video (Figure 6a) is interesting. It appears that although females were less certain of their answers (having selected 'unsure' more frequently than males, males $\mu = 67.4\%$, females $\mu = 58.9\%$) for the "Conversation" video, those answers they were sure of were correct more often than males'.

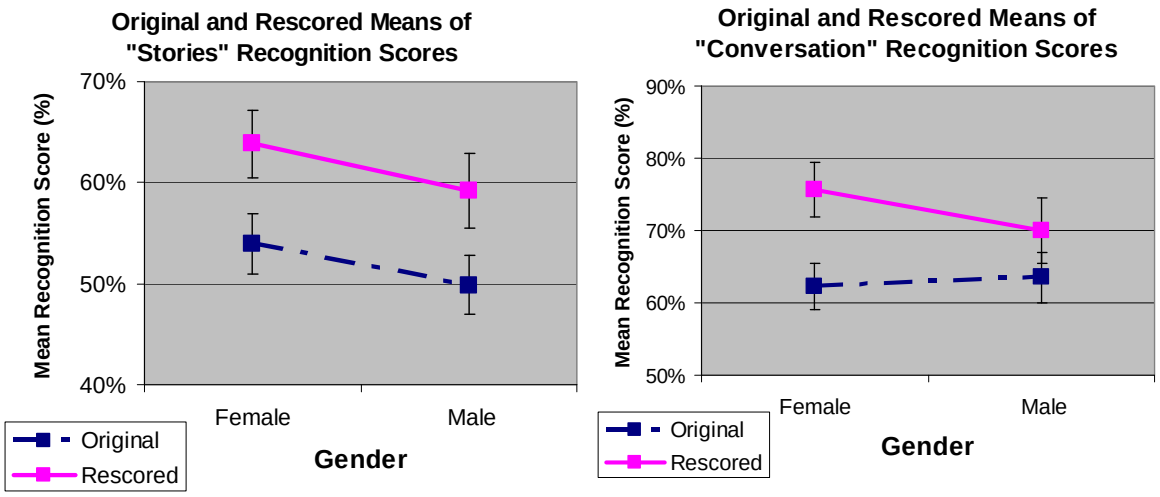


Figure 6 Original and rescored recognition means arranged by GENDER and plotted separately by VIDEO. Both genders improved in the "Stories" video after rescoring, but females improved a great deal more than males in the "Conversation" video (though not statistically significantly). (Error bars are +/- one standard error)

3.2.2 QUESTION LEVEL and NOISE LEVEL factors

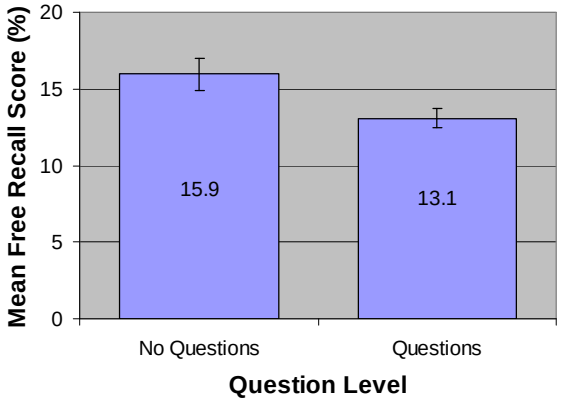


Figure 7 A comparison of mean free recall scores by QUESTION LEVEL clearly shows the significant decrease in performance when participants were asked to answer questions during the videos. (Error bars are +/- one standard error)

We considered that the participants who received questions may have had significantly different recall performance to the other participants, because the extra stimulus of the questions may have helped consolidate their memories of topics or details from the videos. Considering participants by the factor of QUESTION LEVEL (two levels: those who received questions and those who did not) we conducted a 2(QUESTION LEVEL) x 2(GENDER) x 2(VIDEO) ANOVA. Examining free recall scores, the effects of GENDER, $F(1,28) = 6.332$, $p < 0.05$, and QUESTION LEVEL, $F(1,28) = 5.104$, $p < 0.05$, were found to be significant (Figure 3a and

Figure 7). Once again, women's recall performance was better than men's in both groups, but importantly it is clear that requiring participants to answer questions impacted on their ability to recall significantly. This effect may have been greater again if the questions posed to participants had been entirely irrelevant to the videos. Examining the recalculated recognition scores, the only significant effect was an interaction between GENDER and QUESTION LEVEL, $F(1,28) = 7.45$, $p < 0.05$, plotted in Figure 8.

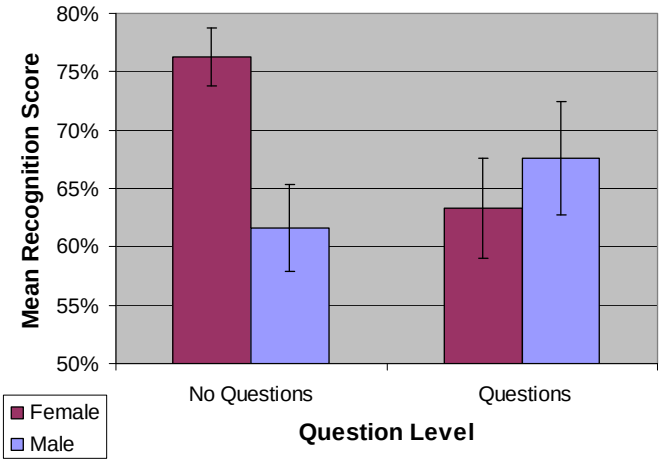


Figure 8 A significant interaction effect was found between GENDER and QUESTION LEVEL though it is suspected that variation due to small sample size and participants not adhering to the requested recognition task timing contributed strongly to this. (Error bars are +/- one standard error)

For consistency we also collapsed IM groups NONE and QUESTIONS to compare against IM groups NOISE and QUESTIONS + NOISE together as NOISE LEVEL groups and then conducted a 2(NOISE LEVEL) x 2(GENDER) x 2(VIDEO) ANOVA. We found that the effect of GENDER was significant in free recall scores, $F(1,28) = 5.263$, $p < 0.05$, once again with women scoring better than men by approximately 20% (as shown previously in Figure 3a). The news messages were a statistically insignificant distraction and no significant effects were noted in the recognition data.

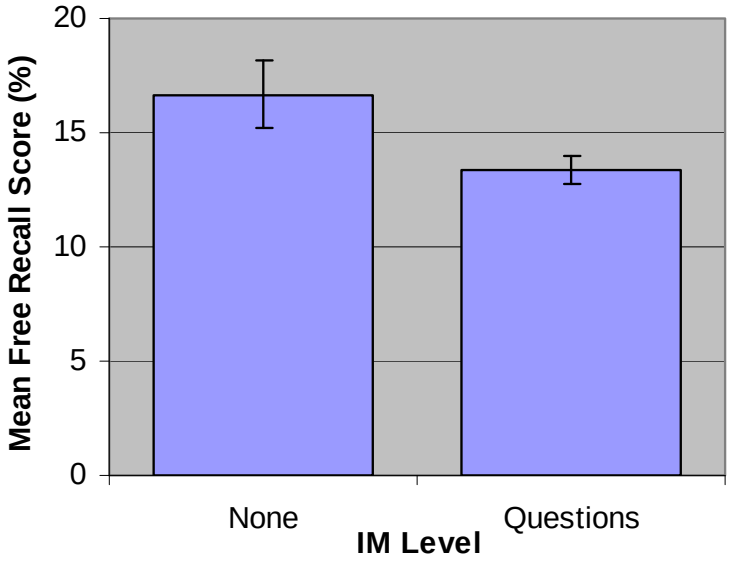


Figure 9 Comparison of mean free recall scores for IM groups NONE and QUESTIONS alone did not reveal IM LEVEL as a significant factor. This figure indicates a drop in performance from over 16.5% down to less than 13.5%, which suggests that variability in the data and the sample sizes may have contributed to why a significant difference was not found. (Error bars are +/- one standard error)

To exclude the effect of news messages, at the cost of halving the sample size, a 2(IM LEVEL) x 2(GENDER) x 2(VIDEO) ANOVA was conducted using only IM groups NONE and QUESTIONS. No significant effects were uncovered ($p > 0.05$ for both GENDER and IM LEVEL), however this may have been only due to the small sample size as the effect of IM LEVEL appears evident in Figure 9, which shows an approximate 20% drop in performance. This result is consistent with those of the QUESTION LEVEL ANOVA described previously.

3.2.3 Effect of distraction questions on recall

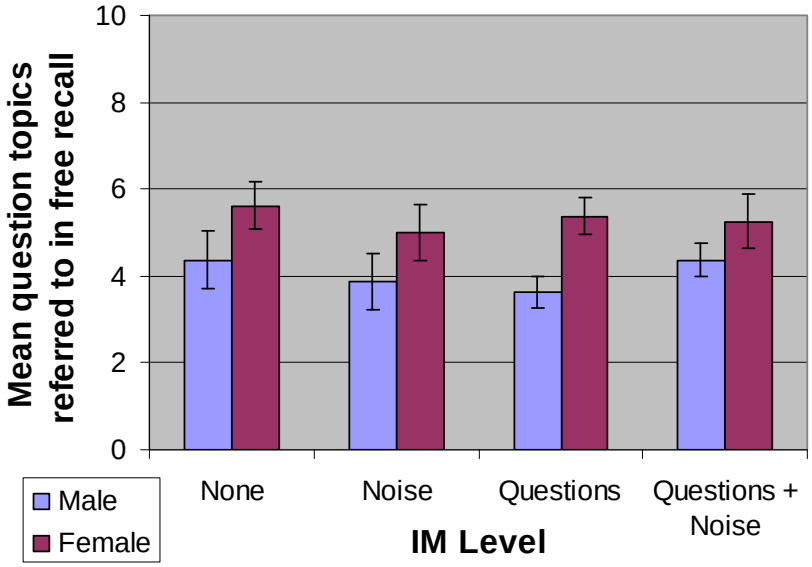


Figure 10 Analysing the effects of IM LEVEL and GENDER on the mean number of question-related topics referred to in free recall revealed no significant effect from IM LEVEL, but women consistently performed better than men. (Error bars are +/- one standard error)

As mentioned in Section 3.2.2, we considered that since the questions put to participants in IM groups QUESTIONS and QUESTIONS + NOISE related to the videos, they might enhance memory of the facts they were asked about. We re-examined the data to determine how many question-related facts were noted in the free recall[12]. We expected a random distribution for IM groups NONE and NOISE (as they did not receive questions) and an above average number (i.e. more than five, as there were ten questions for each video) in IM groups QUESTIONS and QUESTIONS + NOISE. Examining all IM groups independently with a 4(IM LEVEL) x 2(GENDER) x 2(VIDEO) ANOVA, with the dependent variable being the number of facts in the recall relating to the questions asked, the only significant effect, again, was in GENDER, $F(1,24) = 6.332$, $p < 0.05$. In this case women recalled approximately 20% more question-related facts than men (males $\mu = 4.1$, $\sigma = 1.5$, females $\mu = 5.3$, $\sigma = 1.5$, shown in Figure 10). It is interesting to note that even women only just scored a mean of more than five out of ten and there appears to be no drop in performance with an increase in IM LEVEL. As such IM LEVEL was not a significant factor.

Examining the question-related recall scores with a 2(QUESTION LEVEL) x 2(GENDER) x 2(VIDEO) ANOVA, we once again found only a significant difference in GENDER, $F(1,28) = 7.026$, $p < 0.05$, and no significant difference between QUESTION LEVEL. Finally, we attempted to remove the potential complication factor of noise and compared only IM group

[12] The recognition data was not rescored in this manner because the questions asked during the treatments were designed to be independent of the questions/statements presented in the recognition task, to avoid any such effects.

NONE, the control group, and IM group QUESTIONS, which received questions only and no news messages. However, once again, the only significant difference was in GENDER, $F(1,12) = 6.448$, $p < 0.05$, and there was no significant difference in scores across IM groups.

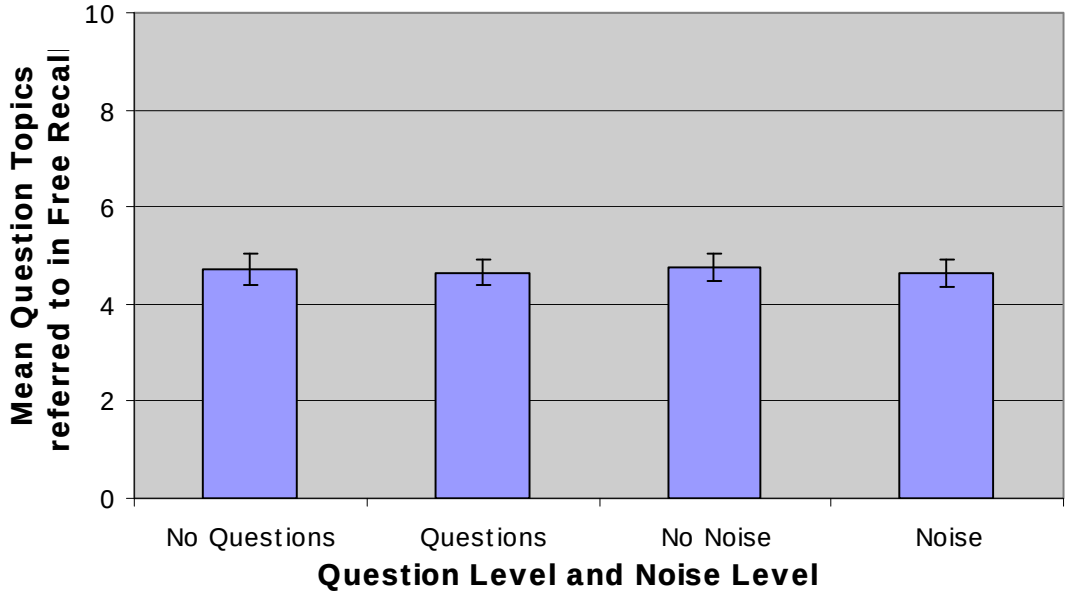


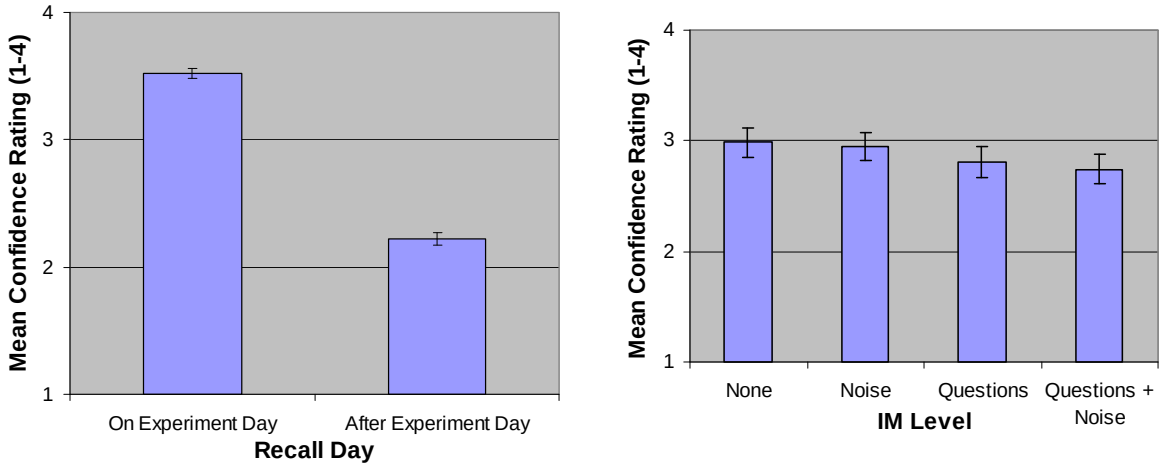
Figure 11 Number of in-video question topics referred to in free recall (out of a possible ten) by the factors QUESTION LEVEL (NO QUESTIONS and QUESTIONS, each with 16 members) and NOISE LEVEL (NO NOISE and NOISE, each with 16 members). There is remarkably little variation. (Error bars are +/- one standard error)

Examining the number of in-video question-related statements by NOISE LEVEL also showed no significant effect. In fact, it is perhaps remarkable how little difference there was across QUESTION LEVEL and NOISE LEVEL (and the extent of the variation within the groups) given the group size of 16 (Figure 11).

3.3 Effects on Confidence in Recall

This section describes the analyses performed on the confidence ratings drawn from the recall data and the Likert scales of the recognition data (previously reported in Weber, 2008), and also confidence information drawn from the valid answer counts arising from the rescoring of the recognition data.

Using the recognition and recall data prior to rescoring, a matched pair t-test over RECALL DAY found a significant difference in confidence in recall from the experiment day to the next, $t = 19.298$, $df = 31$, $p < 0.001$, with a higher mean confidence rating on the day of the experiment (Figure 12a). A 4(IM LEVEL) x 2(GENDER) by 2(RECALL DAY) ANOVA found a significant linear relationship across the IM LEVEL factor, $p < 0.05$, trending as expected with lower confidence with more IM interruptions (Figure 12b).



(a) *(b)*

Figure 12 Analysis of confidence combined across both free recall and (original) recognition shows a significant effect of (a) time, t = 19.298, df = 31, p < 0.001, and (b) a significant downward trend by IM LEVEL , p < 0.05 (though no significant effect across IM LEVEL was found, F(3,28) = 2.661, p = 0.067). (Error bars are +/- one standard error)

After rescoring the recognition data, however, examining male and female valid answer counts specific to each video indicated males were 'unsure' of fewer recognition statements than females in both videos. The difference was only slight for the "Stories" video, where the mean male valid answer count was 1.2% greater than females', but more so for the "Conversation" video, where the mean male valid answer count was 8.5% greater than females'. A 2(GENDER) x 2(VIDEO) ANOVA revealed this was not statistically significant. As mentioned at the end of Section 3.2.1, although women had a lower valid answer count (implying lower confidence in their answers) than men, they had proportionately better scores for the "Conversation" video (though not statistically significantly better). Improvements in recognition scores for the "Stories" video were, as mentioned above, even across the sexes.

Using the original recognition data ANOVAs of QUESTION LEVEL and NOISE LEVEL against free recall and recognition confidence scores also revealed no statistically significant results, although trends for lower confidence with more questions and noise could be seen in graphs (Figure 13). Looking at the number of valid answers by IM group, it was noticed that IM groups QUESTIONS and QUESTIONS + NOISE (i.e. those receiving questions) appeared to have much lower means (Figure 13a), so a 2(QUESTION LEVEL) x 2(valid answer counts per video) repeated measures ANOVA was conducted. No significant difference was found between the valid answer counts of those who received questions and those who did not (Figure 13b). If there had been a significant difference it would have offered evidence that asking questions of a participant affected their confidence in their recognition recall performance, but this is not supported by the data.

No significant difference between genders was found in confidence ratings, counter to the hypothesis, once again perhaps due to variation within our small sample size.

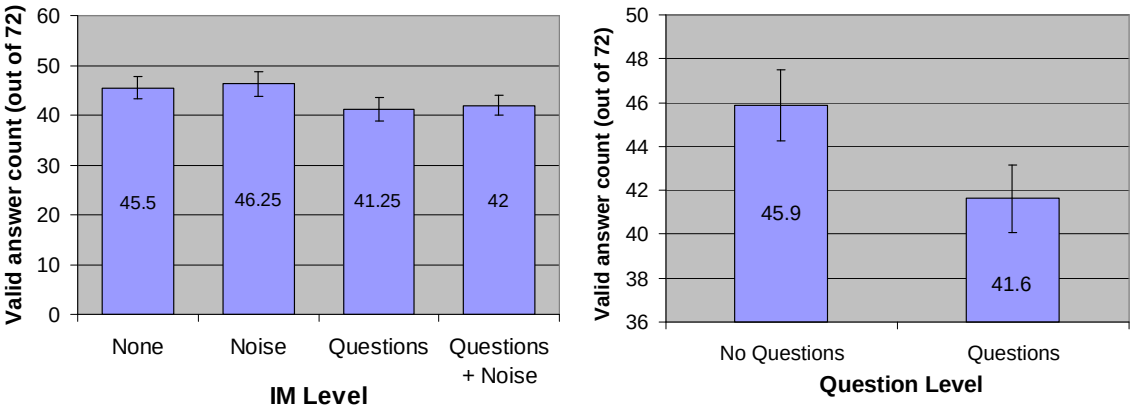


Figure 13 Mean valid answer (i.e. not 'unsure') counts organised by (a) IM LEVEL and by (b) QUESTION LEVEL. No significant difference in valid answer counts (contributing to evidence of confidence) was found. (Error bars are +/- one standard error)

3.4 Effects of Experience with Various Technologies

No significant effects were found for any individual experience factor, nor was any significant interaction effect found for sensible combinations of experience factors, and nor was video order found to be a significant factor. Grouping scores by ranges in participant age (31 years old and younger, 32 to 39, 40 and older, giving nearly balanced groups of 11, 11, and 10 participants respectively) and reanalysing the results revealed no significant difference between the age brackets. Neither the time spent answering questions, nor the number of characters typed, had any predictive effect on recall or recognition performance or confidence.

3.5 Exploratory Findings

Although not statistically significant, some trends in the data appeared interesting enough to report on as potential areas of further investigation in future studies. While analysing the recognition scores (including the 'unsure' responses), an analysis was conducted on the scores broken down by true gist, false gist, and elaboration statements (the GIST factor), across the factors of GENDER and IM LEVEL. In a 2(GENDER) x 3(GIST) ANOVA, GIST was found to be a significant factor, $F(1,30) = 22.738$, $p < 0.001$, and by using the order (false, elaboration, true) a linear relationship was found, $p < 0.001$, with the lowest performance on false gist statements (Figure 14). It is worth reminding the reader here that responses to elaboration statements were regarded as correct when 'unsure' was selected as well as any 'false' rating. Delving into the false gist answers, it appears that VIDEO may have an interaction (though not statistically significant) with IM LEVEL examining the plot lines in Figure 15. Participants scored well for QUESTION LEVEL NO QUESTIONS and poorly for QUESTION LEVEL QUESTIONS for the "Conversation" video, but comparatively evenly for the "Stories" video.

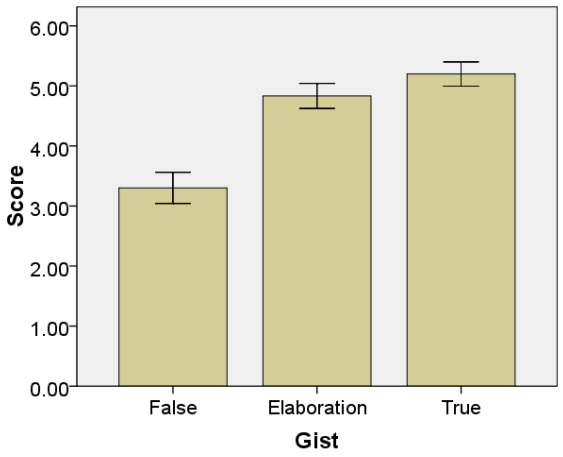


Figure 14 A gist-based analysis of the original recognition scores (out of eight) revealed a significant difference between false gist, elaboration, and true gist statement scores, with participants scoring lowest on false gist statements. (Error bars are +/- one standard error)

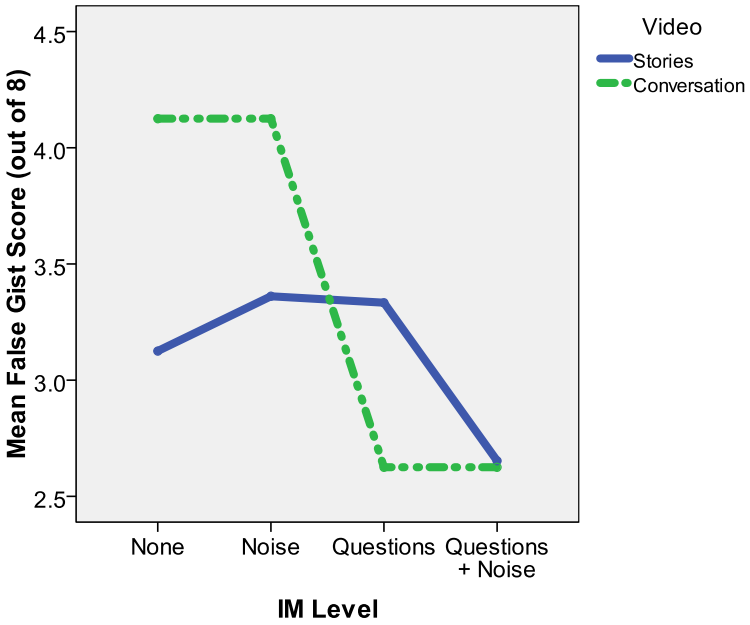


Figure 15 This plot of false gist mean scores by IM LEVEL shows a marked decline in scores with interruption from questions for the "Conversation" video, compared with more even scores for the "Stories" video. N.B. these scores include 'unsure' responses.

4. Discussion

4.1 Support for the Hypotheses

4.1.1 Interruption from instant messaging

Our first hypothesis proposed a significant degradation of recall performance and confidence with an increase in interruption from instant messaging. Although IM LEVEL was not sufficiently strong as a factor, QUESTION LEVEL was, so requiring a response to IM significantly affected the recall performance of our participants. In fact, a significant downward trend was also found in free recall against IM LEVEL and across all confidence ratings combined. The reasons for the weakness of this finding may be associated with between-subjects variation due to the small sample size, visible in the size of the error bars in Figure 4.

By requiring the participants respond to questions, especially when they needed to find questions among news messages, as in the IM group QUESTIONS + NOISE, their recall performance was significantly impacted. This is presumably because the need to process the (mental) verbal input of the instant messages interfered with the consolidation of the memory of the verbal input of the video. This seems to relate to the findings of Stroop (1935) and Cherry (1953) inasmuch as it appears that people have difficulty performing two verbal tasks simultaneously. It would be interesting to carry out a study where attending to the video was the secondary task rather than the primary task. It is quite possible (predicting from Cherry's observations) that participants receiving lots of instant messages that they must attend to would have difficulty remembering anything more than vague details from the videos. A more subtle experiment is required to tease apart the distraction effect of 'noise' messages from answering questions. With the least modification to the existing experiment, the participants' memory of the news messages could be also tested, or participants could be told to attend to them as they might be tested on their content.

Analysis of recognition scores indicated no significant effects from the factors of GENDER or IM LEVEL, even after rescoring. This observation is predicted in the literature, as Yonelinas suggests that recognition performance is not as easily affected by certain manipulated variables (e.g. divided attention) as recollection performance, as it is less specific and more gist-based (2002). In retrospect other factors may have influenced the validity of the recognition data including variation in the time between watching the video and the execution of the recognition task and the inclusion of the 'unsure' option, as mentioned earlier.

As can be seen in Table 2, rescoring of the recognition data resulted in an increase in the control group (IM group NONE) mean score from 14.3 to 16.4 out of 24 and over all participants' mean scores from 13.5 to 16.1. Only after this rescoring do this study's scores compare favourably with those of Dobson-Keeffe's, which is interesting to note, given he did not rescore his data. The obvious difference with this study is that it exposed participants to two stories consecutively, rather than just one, in which case the initial difference in performance could be reasonably expected. Doubling the amount that a participant needs to

remember might be regarded as a non-trivial challenge when each individual story is quite dense with facts, containing between 150 and 180 propositions.

Recognition mean scores from Dixon et al. (1989), Dobson-Keeffe (2007), and this study (both overall mean initial and rescored scores and for IM group NONE, the control group, specifically) are shown in Table 2. Dixon et al.'s mean is approximately 19 out of 24, Dobson-Keeffe's approximately 15 and ours a little higher than Dobson-Keeffe's at a 16.4 for IM group NONE and 16.1 overall. Dobson-Keeffe also did not find a significant difference in recognition scores between his two treatments (though none included distraction similar to our IM interruption), so it is arguable that his mean scores are typical. Our scores are more similar to his than to Dixon et al.'s, which are approximately three points or 12.5% greater. Given that Dobson-Keeffe's and our studies used Australian participants across a range of working ages, while Dixon et al.'s study used participants of retirement age and in their late teens in Canada, age bracket and residential geographical location may affect recognition performance with the tests used. Dixon et al.'s scores, although from a smaller sample, are drawn from tests of ten stories, rather than Dobson-Keeffe's and our two stories each, and those that Dixon et al. used do not include the ones that Dobson-Keeffe and we used. A potential likely influence, however, is the delay in recognition task execution experienced in our study as our initial recognition scores were lower again by one to two points than Dobson-Keeffe's (which included the 'unsure' responses), which would indicate that population variation caused the difference between the studies' mean recognition scores.

Table 2 Summaries of recognition scores for stories from Dixon et al. (1989) from this study (both the original scores and the recalculated scores) and two other authors, Dixon et al. and Dobson-Keeffe (2007). The total scores are out of 24. The stories were designed to be interchangeable so scores should be similar but Dixon et al.'s are noticeably greater. (μ = mean, σ = standard deviation)

Study	Category	Total	
Dixon et al. (1989)	Young	μ=20.1	σ=0.94
	Old	μ=18.4	σ=0.81
Dobson- Keeffe (2007)	Story 1	μ=15	σ=2.8
	Story 2	μ=16	σ=2.6
Current study	Overall	μ=13.5	σ=2.6
	IM group NONE	μ=14.3	σ=3.5
Current study (rescored)	Overall	μ=16.1	σ=3.9
	IM group NONE	μ=16.4	σ=3.7

Examination of recall and recognition results from Dixon et al. (1989) and Dobson-Keeffe (2007) indicated that our scores might have suffered floor effects (Table 2 and Table 3), i.e. the task was so difficult most participants scored poorly and so the variation of scores was limited to the lower portion of the score range. The smaller score range made it hard to draw distinctions between groups of scores due to overlap between the groups. The potential floor effects limit the recall performance scores. Also many participants were still writing at the end of the free recall time, so variation between participants with good memory was not observable because they did not have a chance to perform as well as they could. It is

important to note that our recognition results are similar to Dobson-Keeffe's despite this. With regard to free recall scores in Table 3, however, Dixon et al.'s scores (around 40%) are all significantly higher than those of Dobson-Keeffe (20-25%) and Dobson-Keeffe's are greater than ours (13-16% overall, but 16.7% for the control group, IM group NONE). The standard deviation of our free recall scores (4.5–6.5%) was much lower than those of Dobson-Keeffe (7-9%) and Dixon et al. (11-17%), which implies our population had much less variability than theirs, so perhaps the effects of locality and greater number of story treatments do mostly explain the differences in recognition score.

Table 3 Summaries of free recall scores for stories from Dixon et al. (1989) from this study and two other studies, in Dixon et al. (1989) and Dobson-Keeffe (2007). The stories were designed to be interchangeable so scores should be similar, although the results for IM group NONE, in particular the "Stories" video results, alone should be directly comparable to those of Dixon et al. and Dobson-Keeffe as there is no interference from IM interruption.

Study	Story/Video	Sample size	Mean score (%)	Standard deviation	Minimum	Maximum
Dixon et al. (1989)	Story 16	19	45	11	26	63
	Story 17	18	35	17	6	63
Dobson-Keeffe (2007)	Story A	26	21	9.1	7	41
	Story B	26	26	7.2	13	39
Current study (all IM groups)	Conversation	32	15.9	4.8	7.7	25.6
	Stories	32	13.1	4.75	5.2	26.5
IM group NONE	Conversation	8	17.7	5.2	9.9	21.2
	Stories	8	15.6	6.7	5.2	20.0

It is necessary to elaborate on the issue of floor effects. Any such effects observed may be due to a number of causes. The first is the fact that Dixon et al.'s stories are intended for use in isolation, rather than in combination as we used them, and it is possible this overloaded participants. Secondly, as mentioned above, it was noted that a significant number of participants were still writing at the end of the five minutes provided during the free recall component of the experiments. Thirdly, the stories were certainly American-centric, which may have had an effect; however the two stories chosen out of the 25 available were among the least American-centric specifically to cater for the predominantly Australian participants. Thus it would be difficult to believe this corresponded to such a significant difference in mean recall score (35% and 45% compared with our 16% and 18% in the control group). Dixon et al. used at least five times the number of stories in their tests, and it may be that our story choices happened by chance to include the two most proposition-dense stories. Finally, Dixon et al.'s sample sizes were considerably smaller than those of this study, so there is a chance their small population was not representative, especially as Dobson-Keeffe's scores are much closer to ours than theirs[13]. In any case floor effects could explain the lack of significance between the groups as discussed previously.

[13] Another reason for Dobson-Keeffe's scores being similar to ours is that all of our test subjects were sourced from DSTO, and were likely very different to Dixon et al.'s subjects, in age, education level, personality variables, and possibly competitiveness (given they were work colleagues and many were scientists and engineers).

When we investigated whether the video-related questions asked of IM groups QUESTIONS and QUESTIONS + NOISE enhanced their recall of question-related facts, we found no difference across the factors IM LEVEL and QUESTION LEVEL (though, once again, women performed better than men). This may simply be due to the small sample size, as the dependent measure was number of question-related topics[14] referred to in the free recall, providing a score between zero and ten. The plot of IM LEVEL against number of question-related topics referred to shows that there was very little variation in mean between IM groups but large variation within each IM group (apparent in Figure 10 and Figure 11). Future work should compare the recall performance of people asked irrelevant questions with people asked relevant questions. This would answer the question of whether interrupting someone with a question is acceptable if the question relates to what they are doing as their primary task; one could send through a question like "What was that acronym the Commander just said?" but not "What score did New Zealand make in the cricket yesterday?" This would be particularly valuable for guiding the development of workplace policy. As mentioned, this experiment contained only video-related questions to simulate the first of these example questions.

A final note on recall relates to the gist scores derived from the recognition tests. The recognition questionnaires had three types of statements for participants to attempt to recognise: true gist statements, false gist statements, and elaborative statements (which embellish a true gist statement). Dobson-Keeffe's results indicated little difference in gist scores (2007). In contrast, the results observed in this study indicated a significant difference between gist statement scores (Figure 14). Participants were noticeably worse at recognising statements as false and such results should be expected of familiarity memory retrieval (Yonelinas, 2002). True gist statements may simply have scored highly because it is easier to assume something is true if one cannot find a particular detail in the statement to declare it false, whereas a false statement requires one element of the statement to trigger a specific memory of detail with which to compare. Elaboration statements may have scored highly also because of the scoring procedure, which gave a point if the statement was regarded as false or the participant was unsure about it (a rating between 1 and 4, inclusive). This finding seems to indicate that the participants in this study may have been more suggestible than Dobson-Keeffe's, or that his results were unusual compared with those in the literature.

Some investigation was performed on how the gist scores appeared after rescoring, however there are a number of factors which indicated that the results were not worth reporting. Because 'unsure' results were discounted, many gist-based scores were based on very few samples (there were only eight per gist to begin with), and some participants answered 'unsure' for so many statements that they have no scores at all for some gists. Furthermore, the rescoring resulted in twice as many false-gist equivalent statements as true gist, which may even reduce the usefulness of aggregations across all the gist statement for recognition scores.

When considering the original false gist results more closely (Figure 15), it appears as if participants who received questions were much worse than those receiving no questions at

[14] This refers to topics relating to the subjects of the questions asked of participants in IM groups QUESTIONS and QUESTIONS + NOISE.

judging the truth of statements based on the "Conversation" video, though only those receiving questions and noise performed particularly poorly in the "Stories" video. It is arguable that the "Conversation" video was a more realistic representation of people communicating (given it was genuine dialogue, as opposed to a highly structured artificial monologue), and as such, might yield more realistic recall performance results. This result implies that people distracted with questions may be significantly worse at picking false facts and, as such, may act on such false positive conclusions. Performance in this regard may be improved by training in scepticism but cause trouble in environments such as the military where a great deal of trust is put in what senior officers communicate. If the transmission of such messages is affected by distraction (and misheard) and acted upon incorrectly, the strength of that trust may be diminished over time. Although this is a trend based on a small data set and not a significant result, it may be worth investigating as part of future studies.

Although the true gist and elaboration results were examined more closely, no significant effects could be found, nor any trends of note.

4.1.2 Gender differences

There was significant support for our second hypothesis, regarding a difference in recall performance between the genders, with women consistently scoring higher in free recall, as reported by Kimura (2000). This occurred in nearly every analysis that was conducted against free recall and confidence measures, independent of the arrangement of participants (i.e. by IM LEVEL, QUESTION LEVEL, or NOISE LEVEL). For comparison, Dobson-Keeffe's (2007) found a significant difference in GENDER (females scoring higher in free recall), but it was limited to the first of his two treatments and so it was reported as a learning effect. It is therefore interesting that the effect disappeared in his study while it was maintained in almost every comparison in ours (at least for free recall and confidence). The lack of a significant GENDER effect in the recognition data may be explained by the variability in participant execution of the task, as discussed above, but the data were inconclusive.

4.1.3 Confidence

The finding that participants' confidence dropped significantly during the recognition testing from the free recall testing is not surprising. When performing the free recall, participants were asked to rate their confidence on statements directly after having written them. In these circumstances it is highly likely that they wrote what they remembered most clearly first and left what they were unsure about to last. As many participants were still writing at the end of their allotted five minutes, it is quite possible that those memories that were not clear were simply not written, thus only leaving statements with a high confidence rating as the ones written down. In comparison, the recognition test provided a fact for participants to match to their memory, something that would (and did) result in higher scores (as even partial recognition of a concept might bring forth relevant memories) but perhaps with lower confidence, as it is well established that humans remember stories by gist before detail (Kimura, 2000, Medina, 2008).

4.1.4 Experience and age

There was no significant effect of experience on recall performance or confidence, contrary to expectation. This hypothesis was primarily based on the modern conventional wisdom that younger, technology savvy people are better at task switching, whereas the literature suggests otherwise (Cole et al., 2006, Ophir et al., 2009) and is supported by our results. Furthermore, the youngest participant was 24 and even with a median age of 34.5 years perhaps this group of participants are not sufficiently young to be the young, technology-savvy people referred to by authors such as Cole et al. (2006). However most of the participants would qualify as technology-savvy as they are scientists and engineers.

4.2 Implications and Comments

Even with such a small sample size, these results suggest that a person may miss a significant portion of a conversation or brief if they attempt to use IM at the same time, and thus this practice should be avoided. At the very least it may affect confidence in what is recalled. IM may be used as a secondary task for maintaining awareness via broadcast messages that require no response even if the first task is verbal in nature. Requiring a response, however, diverts too much attention to maintain performance on a primary verbal task. Although increasing interruption yielded only a trend of decreasing recall and confidence and the recognition data were inconclusive, we believe these results may be confirmed and strengthened by further investigation. For example, further study examining how many messages requiring response affect performance on the primary task may help identify a threshold below which performance on the primary task is not degraded or is acceptably degraded. It is also not clear what the effect would be of testing people's recall of broadcast messages in conjunction with the videos, but testing for this would tease apart the effect of distraction from reading instant messages as compared to the effect of reading them and responding to them, which would be realistic and valuable.

The findings related to gender differences have support in the literature (Kimura, 2000), although it would be useful to see if they apply to recognition performance as well as recall. It has been noted that when geographically distributed groups work together via video links, those groups often nominate a 'bridge' member, who acts as a liaison between the local group and remote groups and most communication is channelled through them to reduce confusion (Mark et al., 2003). These results indicate that more research into gender differences is required, and it is possible that women may be better suited to this particular role than men on average, and decision-makers may choose to consider this while allocating staff to roles.

It is important to note that this study differs from other research in this area through its examination of IM as a mechanism for delivery of broadcast messages, such as news messages, as well as chat. All other research reviewed has used IM only for one-to-one chat.

Additionally, it is interesting to consider the effects of side-bar conversations in general, regardless of the mechanism of delivery, during group collaboration. They may not all be detrimental. For example, not all individuals will be fully engaged during every collaboration (colocated or remote), and when not fully engaged, or even interested in the discussion, side-

bar chats may, in fact, serve to keep people thinking about the topic at hand, instead of letting their minds wander (or fall asleep entirely, depending on a variety of factors).

Of course, there also remains the question of how representative of all group collaboration (co-located or remote) the experiment's video-mediated briefing concept is. What our evidence does indicate, however, pending further investigation, is that engaging in IM during a meeting may cause a person to miss what is being said, regardless of its importance to them. This is an extension of what Cherry found in his *dichotic listening* experiments (1953).

4.2.1 Implications in various contexts

The results indicate that using IM while attending to verbal communication has different implications in different contexts. Here we consider a few examples, representative of simple, intermediate, and complex collaboration contexts, and how the findings relate to them.

A one-on-one video or in-person communication with a colleague is an example of a simple collaboration and has obvious similarities to the treatment in our experiment. Attempting to respond to IM or some other written communication (e.g. email) while maintaining the conversation will have a clear impact on the conversation.

An example of intermediate collaboration is a collocated or distributed meeting of more than two individuals, where the only task is to communicate directly. These results suggest that even while simultaneously taking notes, significant details of the conversation may be missed. If it is important only to maintain a gist of the conversation the effect may not be so strong, but if the participant in question is central to the discussion and involved in the conversation, then distractions caused by note taking and IM may still have a significant impact. This might apply not only to their memory of the conversation afterwards but also to the flow of the conversation while it is in progress and their and others' associated trains of thought.

Finally, in the context of intense collaboration, where collaboration is time-critical and of great consequence (e.g. in emergency response), and it is essential to maintain the flow of new information into the collaboration environment, how distractions are dealt with is of vital importance. This study's results indicate that it is necessary for participants to choose which communication medium to use at a particular time and to refrain from multi-tasking until breaks in that communication are possible. For example, if emergency workers are being briefed by a commander, then they will attend only to that brief. Once it is over, then they may attend to other forms of communication. This highlights the need for some information to be buffered. This is obvious with media such as IM, but not so clear with forms such as television (e.g. live news coverage of a disaster area), especially when they are shared.

4.2.2 Implications for Defence

The results of this study indicate that recall and confidence in recall degrade with an increase in distraction from IM, especially when interaction (instead of mere monitoring) is required. These two issues of attention and confidence are of great importance to the military as, in times of stress, it is vitally important that personnel can act correctly without hesitation. If they are not confident, then they will hesitate, and if they are distracted, then they may

misremember and think that an incorrect action is correct or simply not know which action is correct. This suggests that further research is required to guide the development of policies regarding the use of IM during other modes of collaboration in Defence environments. The original idea of using IM as a way to 'nudge the next person' to ask them a question during a distributed meeting has been shown to disrupt what can be remembered about the discussions of the meeting, though it is quite possible that talking with an adjacent person may be just as distracting as using IM. In this way, the attention that is required to attend to what is being discussed needs to extend beyond the courtesy of not talking while others are talking to not typing or reading either. For example, it is particularly notable when a meeting's minute taker is not experienced, as they often miss details while they are taking notes. More practiced minute takers will miss less.

Furthermore, these findings suggest examining the effectiveness of using combined video/tickertape displays, such as televisions tuned to CNN, in environments where other collaboration is occurring. Movement draws attention (Bartram, 2001, Bartram et al., 2003) and tickertape displays in work environments will draw the attention of workers. Many productivity guides discourage the use of immediate alerts for non-important information (e.g. most email) (Song et al., 2007), which also applies to distracting tickertape displays. This is discouraged primarily because of the cost of context switching between tasks. Depending on the task, it may take tens of minutes or even hours to 'get back into the flow' (Fried, 2006, Medina, 2008). This kind of visual distraction (and that is to say nothing of the audio chatter of a news anchor) may reduce the efficiency of workers and should be considered when designing work environments.

4.2.3 Limitations of the study

This study could be improved in a number of ways, relating mostly to design and execution. This section briefly highlights a number of these opportunities. On the issue of analysis, more extensive analysis may have been warranted if results had been firmer thus encouraging the use of more sophisticated techniques. As the results were not particularly strong, the simple analyses performed were appropriate.

The between-subjects variation within this study's sample, compounded by its small size, strongly influenced the appearance of significant factor effects. The small sample may also not have been particularly representative of the general population. A larger sample size would help reduce that variation to produce more significant results.

Although only one scorer was used, efforts were made to randomise the scoring across IM groups[15], so that any learning effects in scoring were spread evenly among the groups. Even then, scoring will vary depending on the mood and experience of the scorer. Validity of the results would be improved by the use of more scorers, if the experiment were to be conducted again with a greater number of participants (e.g. 120 participants). Resources did not permit multiple scorers for this experiment. Another technique, to consider for future experiments involving one scorer, is for the scorer to rescore the raw data without reviewing the first

[15] Free recall scoring was done in combinations of four, with one recall sheet from each IM group, always scored in order from IM group NONE to QUESTIONS + NOISE.

scores, and then average the pairs of results; this would ameliorate learning effects in the single scorer, though probably not remove them entirely.

Another open question is how comparable the "Conversation" video is to the "Stories" video. Although the actors in the conversation were informed of the intentions of the researcher (the video was to be a fact-dense stimulus so participants would be exposed to plenty of items to recall), the conversation was free flowing and ad hoc. This alone may have made the conversation more memorable than the stories, which are obviously designed and appear slightly stilted when spoken aloud (as several participants reported). The propositional analysis (used for scoring free recall) was performed to the best of the primary researcher's abilities according to Dixon et al.'s (1989) descriptions, but not with the same rigour or testing that Dixon et al. could afford.

Close to half the participants were still writing at the end of the five minutes allocated for free recall after either or both of their videos, indicating that if the experiment were re-run more time should be allocated for the free recall.

The 7-point structure of the recognition questionnaires resulted in potentially confused recognition results. Attempting to rescore the data to remove the 'unsure' option resulted in potentially lop-sided results with twice as many false gist-equivalent statements as true gist ones and it also affected measurements of confidence removing one of four levels. The number of options on the recognition questionnaire prompted one participant to suggest it was too complex, and it meant that results of high confidence may have been conflated with a lack of interest (i.e. participants not bothering to really consider if they thought the statement was possibly, probably, or definitely true). A way around that would be to separate the questionnaire responses into a binary question of recognition and a suitable rating of confidence.

As mentioned earlier, the participants were all DSTO staff, who are atypical of the general population and military personnel in terms of education level and personality traits. This would have had an influence on the results obtained, though it is not clear if their recall would be better or worse.

Finally, as mentioned, many participants did not execute the recognition task at the required time and as such the recognition performance varied greatly. This was probably primarily because many participants were shown the videos on a Friday, requiring them to do the task on a Saturday, away from work (and thus easier to forget about). This may have been controlled for, or at least examined, if the task also included noting the time and date on which it was done.

4.3 Future Work

The findings indicate that this pilot study was valuable and that re-running it with more participants and slight modifications to its method would likely yield more significant results. If re-run, it would be sensible to lengthen the period of time for free recall, and shorten the videos to avoid any possible floor effects. Using stories from Dixon et al. (1989) for both

videos would also ensure consistency in scoring across the treatments, although this would remove the dimension of observed conversation and interaction (because all Dixon et al. stories are intended to be read or spoken by one individual). Once testing the current hypotheses is complete, variations can be made for further investigation.

4.3.1 Variations in interruption and tasking

Given this study used questions that related to the videos as IM interruptions, it would be useful to know what the effect on recall is when unrelated questions are used or when recall of news messages is also measured. Also IM is used in many other ways than for one on one chat. Often IM is used as a mechanism to communicate across groups; messages may be personal (ie intended for an individual) or related to one or many interest groups, and any individual may be a member of a large number of interest groups. In this way, questions may have different degrees of relevance, and relevance to different, potentially overlapping, interests, be they work-related, task-related, or non-work-related.

Furthermore, it would be useful to know if there is a relationship between the number of questions sent to a participant and the resulting recall of a video. Perhaps there is a rate of messages below which recall is not affected, and thus questions can be sent at or below that rate without significantly distracting an individual. However it is likely that the effect of IM interruptions is dependent on the flow of the primary verbal stimulus, which vary in propositional density over time. This is related to the issue of the effect of interruptions at different points in an arbitrary task as discussed in Section 1.2.1.

There may be merit in varying the difficulty of the questions also, as more complicated questions require more cognitive effort to answer, effort that would draw even more attention away from the primary task.

Also, as discussed by Knott et al. (2006b), using IM may have different effects with different types of primary tasks. If the primary task was a pattern matching task (e.g. watching a video of a landscape and noting when targets appear), perhaps the use of IM in combination with such a non-verbal task would not affect primary task performance significantly.

4.3.2 Variations in IM user interface

In this study, participants in IM group QUESTIONS + NOISE received questions and news messages on the same scrolling row in Sticker, forcing them to search for questions among the messages. Different configurations of Sticker may result in different recall performance, as Sticker has the ability to have multiple ticker lines and messages can be instructed to appear on different lines depending on customisable criteria. Typically news messages and personal (chat) messages are, indeed, shown on different ticker lines. This may affect people's ability to spot question messages amongst the news messages and could be tested for. Sticker has the ability to turn off the default tickertape animation and replace it with an appear-and-fade style display, as tested by McCrickard et al. (2001). Sticker is also only one implementation of an IM client, and other IM clients have very different user interfaces. Sticker was selected as it is the IM tool of choice in the BattleLabs, but other IM tools such as MSN Messenger, AIM, and IRC are more widely used, and none use a tickertape display by default.

4.3.3 Variations in involvedness of participant

Finally, as the original intention of this study was to contribute to an overall research thrust into the augmentation of VTC with a variety of collaboration tools, it would be necessary to vary the degree to which participants are involved in a collaborative effort when they are distracted. In this experiment, participants were expected to listen to someone presenting or two people coming to a decision via VTC. In typical collaboration, participants would observe local collaborators speak with remote collaborators, or would themselves interact with local and remote collaborators. Distraction under these circumstances may significantly affect their ability to collaborate successfully.

4.4 Conclusion

Overall this study's results indicate that asking people to respond to questions via IM during a VTC will significantly affect how much of the VTC they remember. It will also significantly affect their confidence in their memories of the VTC. However, the results indicate that merely having an animated tickertape display in view (which they know they can ignore) is unlikely to affect their memory of the VTC. Furthermore, the results indicate that experience with various communication technologies has no impact on recall performance, nor does age. The significance of GENDER in free recall described in the literature (Kimura, 2000) was observed with female participants consistently outperforming male participants. The lack of the expected significant effect of IM LEVEL may be due to the small sample size compounded by large between-subjects variability; however significant trends were observed in the expected directions, suggesting that further investigation with greater numbers of participants is warranted.

Of particular interest to Defence are the implications regarding paying attention during collaboration while others are speaking, not only by not speaking to others at the same time but also not using IM or e-mail. Distraction may cause incorrect memories, low confidence in those memories and subsequent hesitation and errors in decision making while relying on those memories, which could potentially affect levels of trust in the workplace. Further research in this area could guide the development of usage policies for IM and other collaboration technologies.

This work was originally inspired through the researcher's involvement with the Livespace collaboration research. However, the application domain of the study is much broader, not requiring Livespaces or even the live VTC present in Livespace environments. It applies to situations requiring individuals to divide their attention between different modes of communication, and thus remains relevant to Livespace, collaboration and human-computer interaction research.

Acknowledgement

The author would like to acknowledge the support and encouragement of Mike Broughton in preparing this report, especially with its review. It has been invaluable and the report would not have been completed without his help.

References

Adamczyk, P. D. and Bailey, B. P. (2004) If Not Now, When?: The Effects of Interruption at Different Moments Within Task Execution. In: *CHI,* Vienna, Austria: 24-29 April

Allen, D. (2002) *Getting Things Done: The Art of Stress-Free Productivity*, Penguin

Altmann, E. and Trafton, J. (2007) Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin and Review* **14** (6) p. 1079-1084

Bailey, B. P. and Konstan, J. A. (2006) On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior* **22** (4) p. 685-708

Bartram, L. (2001) *Enhancing Information Visualization with Motion*. Simon Fraser University

Bartram, L., Ware, C. and Calvert, T. (2003) Moticons: Detection, Distraction and Task. *International Journal of Human-Computer Studies* **58** (5) May 2003 p. 515-545

Bezerianos, A. and McEwan, G. (2008) Presence disparity in mixed presence collaboration. In: *CHI '08 extended abstracts on Human Factors in Computing Systems,* Florence, Italy, ACM

Cherry, E. C. (1953) Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America* **25** p. 975-979

Cole, W., Steptoe, S. and Dale, S. S. (2006) The Multitasking Generation. In: *Time* 10 April, 46-53

Cummings, M. L. and Guerlain, S. (2004) *Human Performance, Situation Awareness and Automation: Current Research and Trends HPSAA II*. Vol. I, Vincenzi, D. A., Mouloua, M.Hancock, P. A. (eds.), Routledge (Taylor & Francis Group)

Czerwinski, M., Cutrell, E. and Horvitz, E. (2000a) Instant Messaging and Interruption: Influence of Task Type on Performance. In: *OzCHI,* Sydney, Australia: 4-8 December

Czerwinski, M., Cutrell, E. and Horvitz, E. (2000b) Instant Messaging: Effects of Relevance and Timing. *People and Computers XIV: Proceedings of HCI* **2** p. 71-76

Day, R.-F., Shyi, G. C.-W. and Wang, J.-C. (2006) The Effect of Flash Banners on Multiattribute Decision Making: Distractor or Source of Arousal? *Psychology & Marketing* **23** (5) May 2006 p. 369-382

Dixon, R. A., Hultsch, D. F. and Hertzog, C. (1989) *A manual of twenty-five three-tiered structurally equivalent texts for use in aging research*. 2, Technical Report, Victoria, Canada, University of Victoria

Dobson-Keeffe, N. (2007) *Evaluation of Virtual Adviser Effectiveness in Short Story Narration*. Masters minor thesis, Adelaide, South Australia, University of Adelaide

Dunn, J. C. (2004) Remember-Know: A Matter of Confidence. *Psychological Review* **111** (2) p. 524-542

Dux, P. E., Ivanoff, J., Asplund, C. L. and Marois, R. (2006) Isolation of a Central Bottleneck of Information Processing with Time-Resolved fMRI. *Neuron* **52** (6) p. 1109-1120

Evdokiou, P. (in preparation) *TeamScope*. Technical Report, Defence Science and Technology Organisation

Evdokiou, P., Thomas, B. T. and Vernik, R. J. (2004) Augmented Synchronised Planning Spaces. In: *The Ninth International Command and Control Research and Technology Symposium,* Copenhagen, Denmark

Fitzpatrick, G., Mansfield, T., Kaplan, S., Arnold, D., Phelps, T. and Segall, B. (1999) Augmenting the workaday world with Elvin. *Proceedings of the sixth conference on European Conference on Computer Supported Cooperative Work* p. 431-450

Fried, J. (2006) Getting in too-much touch (interruption is not collaboration). *Signal vs. Noise* DOI: http://www.37signals.com/svn/archives2/getting_in_toomuch_touch_interruption_is_not_collaboration.php

Fuchs, L., Pankoke-Babatz, U. and Prinz, W. (1995) Supporting cooperative awareness with local event mechanisms: The Groupdesk system. In: *Proceedings of the 4th European Conference on Computer-Supported Cooperative Work,* Stockholm, Sweden: September 10-14

Gaver, W. (1991) Sound Support for Collaboration. In: *Second European Conference on CSCW,* Amsterdam, The Netherlands, Kluwer Academic Publishers

Gazzaniga, M. S., Ivry, R. B. and Mangun, G. R. (2002) *Cognitive Neuroscience: The biology of the mind*. 2nd ed. New York, NY, USA, W. W. Norton & Company

Greenwald, A. G. and Shulman, H. G. (1973) On doing two things at once. II. Elimination of the psychological refractory period effect. *Journal of Experimental Psychology* **101** (1) p. 70-76

Hazeltine, E., Teague, D. and Ivry, R. B. (2002) Simultaneous dual-task performance reveals parallel response selection after practice. *Experimental Psychology: Human Perception and Performance* **28** (3) p. 527-545

Hong, W., Thong, J. Y. L. and Tam, K. Y. (2004) Does Animation Attract Online Users' Attention? The Effects of Flash on Information Search Performance and Perceptions. *Information Systems Research* **15** (1) p. 60-86

Kandel, E. R. (2006) *In Search of Memory: The Emergence of a New Science of Mind*, W.W. Norton & Company, Inc.

Kang, T. J. and Muter, P. (1989) Reading Dynamically Displayed Text. *Behaviour & Information Technology* **8** (1) p. 33-42

Kimura, D. (2000) *Sex and Cognition*. Cambridge, MA, MIT Press

Knott, B. A., Bolia, R. S., Nelson, W. T. and Galster, S. M. (2006a) Effects of Collaboration Technology on the Performance of Tactical Air Battle Management Teams. In: *Symposium on Human Factors Issues in Network-Centric Warfare,* Sydney, Australia

Knott, B. A., Bolia, R. S., Nelson, W. T. and Galster, S. M. (2006b) The Impact of Instant Messaging on Team Performance, Subjective Workload, and Situation Awareness in Tactical Command and Control. In: *11th International Command and Control Research and Technology Symposium,* Cambridge: September 26-28

Lee, J. C., Forlizzi, J. and Hudson, S. E. (2002) The kinetic typography engine: An extensible system for animating expressive text. In: *15th Annual ACM Symposium on User Interface Software and Technology,* Paris, France, ACM, New York, NY, USA

Lien, M.-C., Ruthruff, E. and Johnston, J. C. (2006) Attentional Limitations in Doing Two Tasks at Once. *Current Directions in Psychological Science* **15** (2) p. 89-93

Lövstrand, L. (1991) Being selectively aware with the Khronika system. In: *Proceedings of the 2nd European Conference on Computer-Supported Cooperative Work,* Amsterdam, The Netherlands: September 25-27

Luo, S. and Wu, F.-G. (2007) Reading Dynamic Text Presented on Television. In: *13th Asia Pacific Management Conference,* Melbourne, Australia: 18-20 November

Mark, G., Abrams, S. and Nassif, N. (2003) Group-to-group distance collaboration: Examining the "space between". In: *Proceedings of the 8th European Conference on Computer Supported Cooperative Work,* Helsinki, Finland: September 14-18

McCrickard, D. S., Catrambone, R. and Stasko, J. T. (2001) Evaluating Animation in the Periphery as a Mechanism for Maintaining Awareness. In: *Human-Computer Interaction: INTERACT '01,* Tokyo, Japan: July 9-13, IOS Press

Medina, J. (2008) *Brain Rules: 12 Principles for Surviving and Thriving at Work, Home, and School*, Pear Press

Meyer, D. E. and Kieras, D. E. (1997a) A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review* **104** (1) p. 3-65

Meyer, D. E. and Kieras, D. E. (1997b) A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review* **104** (4) p. 749-791

Meyer, D. E., Kieras, D. E., Lauber, E., Schumacher, E. H., Glass, J., Zurbriggen, E., Gmeindl, L. and Apfelblat, D. (1995) Adaptive executive control: Flexible multiple-task performance without pervasive immutable response-selection bottlenecks. *Acta Psychologica* **90** (1-3) p. 163-190

Monk, C., Trafton, J. and Boehm-Davis, D. (2008) The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied* **14** (4) p. 299-313

Motluk, A. (2007) How many things can you do at once? In: *New Scientist* 7 April, 28-31

Müller-Tomfelde, C., Kellar, B. and Eades, P. (2007) HxI: An Australian Initiative in ICT-Augmented Human Interactivity. *Lecture Notes in Computer Science* **4663** p. 675

Ophir, E., Nass, C. and Wagner, A. (2009) Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences* **106** (37) 24 August 2009 p. 15583-15587

Parsowith, S., Fitzpatrick, G., Kaplan, S., Segall, B. and Boot, J. (1998) Tickertape: notification and communication in a single line. In: *3rd Asia Pacific Conference on Computer Human Interaction,* Shonan Village Center, Japan: 15-17 July

Pashler, H. E. (1992) Attentional limitations in doing two tasks at the same time. *Current Directions in Psychological Science* **1** (2) p. 44-48

Pashler, H. E. (1994) Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin* **116** (2) p. 220-244

Phillips, M. (2008) *Livespaces Technical Overview*. DSTO-TR-2188, Technical Report, Defence Science and Technology Organisation

Potter, M. C. (1984) Rapid serial visual presentation (RSVP): A method for studying language processing. In: *New Methods in Reading Comprehension Research*. Vol. 118. Mahwah, NJ, USA, Lawrence Erlbaum 91-118

Ramsey, N. F., Jansma, J. M., Jager, G., Van Raalten, T. and Kahn, R. S. (2004) Neurophysiological factors in human information processing capacity. *Brain* **127** (3) p. 517-25

Rennecker, J. and Godwin, L. (2003) Theorizing the Unintended Consequences of Instant Messaging for Worker Productivity. *Sprouts: Working Papers on Information Environments, Systems and Organizations* **3** (3) p. 137–168

Rittenbruch, M. and McEwan, G. (2007) *Awareness Survey: An Historical Reflection of Awareness in Collaboration*. HXI-TR-1, Technical Report, HxI Initiative

Rugg, M. D. and Yonelinas, A. P. (2003) Human recognition memory: a cognitive neuroscience perspective. *Trends in Cognitive Sciences* **7** (7) p. 313-319

Schremmer, C., Krumm-Heller, A., Vernik, R. J. and Epps, J. R. (2007) Design Discussion of the [braccetto] Research Platform: Supporting Distributed Intensely Collaborating Creative Teams of Teams. *Proceedings of HCI International (HCII'07)* p. 22-27

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E. and Meyer, D. E. (2001) Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science* **12** (2) March 2001 p. 101-108

Smale, S. and Greenberg, S. (2005) Broadcasting information via display names in instant messaging. In: *International ACM SIGGROUP Conference on Supporting Group Work,* Sanibel Island, Florida, USA, ACM, New York, NY, USA

Song, M., Halsey, V. and Burress, T. (2007) *The Hamster Revolution: How to Manage Your Email Before It Manages You*, Berrett-Koehler

Strauss, E., Sherman, E. and Spreen, O. (2006) *A compendium of neuropsychological tests: Administration, norms, and commentary*, Oxford University Press, USA

Strayer, D. L., Drews, F. A. and Crouch, D. J. (2006) A comparison of the cell phone driver and the drunk driver. *Human Factors* **48** (2) p. 381-391

Strayer, D. L. and Johnston, W. A. (2001) Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science* **12** (6) p. 462-466

Stroop, J. R. (1935) Studies of interferences in serial verbal reactions. *Experimental Psychology* **18** p. 643-662

Taplin, P., Fox, G., Coleman, M., Wark, S. and Lambert, D. (2001) Situation Awareness Using a Virtual Adviser. In: *Talking Head Workshop, OzCHI 2001,* Fremantle, Australia

Telford, C. W. (1931) The refractory phase of voluntary and associative responses. *Experimental Psychology* **14** (1) February p. 1-36

Vernik, R. J., Blackburn, T. and Bright, D. (2003) Extending Interactive Intelligent Workspace Architectures with Enterprise Services. In: *Evolve2003, Enterprise Information Integration,* Sydney, Australia

Weber, D. C. (2008) *Effects of instant messaging on recall during video-mediated briefings*. Masters minor thesis, Adelaide, South Australia, University of Adelaide

Welford, A. T. (1952) The "psychological refractory period" and the timing of high-speed performance—A review and a theory. *British Journal of Psychology* **43** (2) p. 2-19

Yonelinas, A. P. (2002) The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language* **46** (3) p. 441-517

Zimbardo, P. G. (1992) *Psychology and life*. 13th ed, HarperCollins

Appendix A: Story 16: First Flight

Margaret is flying in an airplane for the first time in her life. She is on her way to visit her son, Ted, in Seattle, Washington. He is an accountant there with a major ship-building firm. She left her home in Hartford, Connecticut, at 8:15 in the morning. She flew first in a small plane to Kennedy International Airport in New York City. She waited there for almost two hours until it was time to board the non-stop flight to Seattle. She was surprised at how big the Boeing 747 was. There are more than ten flight attendants to serve the passengers. Just after the plane took off the flight attendants began serving drinks. Then, about an hour later, they began serving a warm lunch. Now that lunch is over they are going to show a movie. Margaret was nervous at first but now flying seems so easy and safe. She will watch the movie and then rest until they land in Seattle. The supervisor announced that all passengers should close their window shades during the movie. The young man sitting next to her grumbled when he heard this request. He is a New York City attorney on a business trip. He said that he had planned to do some paperwork during the flight. Margaret closed her shade, even though it would be nice to occasionally look out the window. She thought it was better to be cooperative. Margaret donned her earphones and began watching the commercials. The movie in her cabin was a recent comedy hit. Her earphones were uncomfortable, occasionally falling from her ears, and she didn't think the movie was very funny. The time passed quickly and soon the lights went on and the shades went up. Her neigbor, the attorney, had fallen asleep and she too felt tired.

Appendix B: Story 17: Playing Cards

Tom and Barb's Canasta Club is meeting at their home tonight. There are three other couples, so there are two tables for playing doubles. Sometimes they draw cards for partners and sometimes they play couples. Tonight, at both tables the men are playing against the women. Barb and her partner, Pat, won the first game against Tom and his partner, Joe. Tom is a good sport but Joes is a sore loser. He is still teasing Tom about a mistake he made early in the game. Joe seems happier now, probably because he and Tom are already one thousand points ahead. Barb occasionally checks the other table to be sure that they have plenty of refreshments. Tonight they have peanuts, popcorn, pretzels, and assorted candies. Barb is also serving plenty of coffee, soft drinks, cocktails, and white wine. Usually, these Canasta parties last from eight in the evening until around midnight. Because everyone enjoys them so much they sometimes last much longer. Two weeks ago they played Canasta until 2:00 a.m. and then Crazy Eights and Gin Rummy until almost dawn. Once they played only one game of Canasta and then switched to Trivial Pursuit. Everyone comes faithfully to the meetings. Obviously, they need exactly eight people to have two full tables. Every spring there is a Canasta tournament for all clubs in their community. The categories of competition are couples, men's and women's doubles, and men's and women's singles. Last spring Tom came in third in men's singles. Barb and Pat came in fifth in women's doubles. Next year Tom and Barb will enter the couple's competition and try to improve on their seventh place finish three years ago. Tom and Barb began playing Canasta together even before they were married. That means they have been playing for over 45 years.

Appendix C: Couple in Conversation: Planning an Overseas Trip

The following is a transcript of a man (M) and a woman (F) in conversation.

M: What did you find out from the travel agent about our upcoming trip?
F: Well, it, er, was very interesting. We first started off talking about what airlines we would take. We looked at the options. We started with QANTAS, because that's...
M: ...the safest airline...
F: ...so they say, I wonder these days, but they say it still is. But unfortunately the price of the QANTAS flights from here via Japan to the USA were about $500 each more than the next flights...
M: Mm, hmm.
F: ... than the next sort of airline, so I thought that's a disadvantage. Uh, we looked at United Airlines, and that seemed a reasonable price. The times it left Sydney were awful – it means we'd need to leave Adelaide, you know, at the first crack of dawn...
M: ...Mmm, the 6 o'clock flight...
F: ...and that's not ever as nice as it might be. And then there was a bit of a wait in Sydney. That just didn't work all that well. The good part about that option was that we probably wouldn't have any trouble with security in Los Angeles with the United Airlines flight from Japan. So that was a situation we had to think about.
M: Hm, mmm.
F: The other one was Japan Airlines. They were a terrific price. Um, if we go via them we would have to stay 24 hours in Kyoto – no not Kyoto, Osaka. Osaka. It goes in to Kanzai Airport.
M: Hm, mmm.
F: Uh, it gets in about 9 o'clock in the evening and we would have to leave about 6 the next day.
M: 6 pm?
F: 6 pm the next day.
M: Mm, mmm.
F: Now, there's a possibility, since we've never been to Japan before, that we could, next morning, have a very brief overnight in the hotel in the airport, next morning catch a train into Osaka, switch to a bus, and take a half-day tour of Kyoto, the Old City in Kyoto, which would be very interesting.
M: Mmm, well that sounds good.
F: It does.
M: Why, why is the price so much cheaper than the American airline?
F: I can't imagine why.
M: Is...
F: It just is.
M: Is that because, you know, perhaps it's, uh, the typhoon season...
F: That's...
M: ... and we might get stranded in Japan.

F: Wouldn't that be awful? Um, it is the typhoon season – we're talking about going at the end of August. Erm, and, it definitely is the typhoon season in that part of the world. They, the Japanese seem very sanguine about flying in and out in spite of typhoons so, I suppose (laughter), we can assume that it's safe. Um, but it would be good to have that experience of catching a peak hour train in the morning into a Japanese city, and then travelling around an old city of Kyoto. So, that's a possibility, and then the connecting flight is direct across from, uh, to Los Angeles from Kanzai Airport in Osaka. So that works quite well, and then it fits in with the connection from Los Angeles to Denver to Jackson Hole, which is what we want to do, because we're planning to go to Yellowstone...

M: Yes.

F: National Park, and we have to go to Jackson Hole. So those connections worked well with when the Japanese Airline flight arrived. So, that seems to be a possibility...

M: Mmm. I'd, I'd select the Japan Airlines on that basis.

F: I think so, erm.

M: Where are we, we are, we are meeting up with Bernard and Joy? Where are we meeting them?

F: We're meeting them in Jackson Hole. They're going to meet our flight because they come in an hour before us, if we can this, um, connection.

M: Mmm.

F: And so that works well. They'll have a van they've hired. It seats five people, because Chris is coming too.

M: Mm mmm, Chris, Chris is their son.

F: Chris is their youngest son. So he'll do the driving probably, and that will work well. And they've booked all the accomodation for the one week in Yellowstone National Park. So, um, that will be good. Uh, we still don't know what will happen with security at the American airports, because I understand flying between ports can be, ah, problematic in that they assume that you're not there on normal business or normal tourism if you fly between American ports. We'll have to take that...

M: Mm mmm.

F: ...chance.

M: Yes, but that's, ah, they assume that all is well if you fly in on an American airline...

F: Correct

M: Yes.

F: But we can't do anything about that.

M: No.

F: Erm, The other thing is, what other national parks do you want to see?

M: Well, I certainly want to go to, um, Grand Canyon...

F: Yeah.

M: ...and Yosemite.

F: That's right, in California

M: ...California. But I, I think if we go out to the Grand Canyon we should also go to see Bryce Canyon and Zion National Park.

Appendix D: Conversation Converted to Statements

Comments made by the male are preceded by (M) and comments by the female by (F). Statements struck through (e.g. ~~the cat sat on the mat~~) were considered devoid of information and were not analysed. Statements with only one line through them were considered repeated information and were not analysed. Numbers in square brackets refer to how much time has elapsed in the video at that point in the transcript in seconds. This information was used to develop the interruption questions and recognition questionnaire.

1. (M) What did you find out from the travel agent about our upcoming trip?
2. (F) It was very interesting.
3. We first talked about airlines.
4. ~~We considered options~~.
5. We considered QANTAS first because (M) it's the safest airline.
6. I'm not sure these days, but they say it is.
7. Unfortunately the price of the QANTAS flights from here to the USA via Japan are $500 [30] more than other flights and airlines.
8. I thought that was a disadvantage.
9. We considered United Airlines, which had a reasonable price.
10. The times it left Sydney were awful.
11. We would have to leave Adelaide at the crack of dawn, catch the 6 o'clock flight.
12. ~~That's never as nice as it could be~~.
13. There's a wait in Sydney.
14. ~~It just didn't work out~~. [60]
15. The good part about that option, the United Airline flight from Japan, was that security would be no trouble in Los Angeles.
16. ~~That was the situation we had to think about~~.
17. The other option was Japan Airlines.
18. They had a terrific price.
19. If we go with Japan Airlines we would have to spend 24 hours in Kyoto.
20. No, not Kyoto, Osaka.
21. The flight lands in Kanzai Airport. [90]
22. It gets in about 9pm, and we would have to leave at (M) 6pm the next day.
23. We've never been to Japan before.
24. After a very brief overnight stay at the airport hotel, the next morning we could catch a train into Osaka.
25. We could then switch to a bus.
26. We could take a half-day tour of the Old City in Kyoto, which would be very interesting. [120]
27. (M) ~~That sounds good~~.
28. (F) ~~It does sound good~~.
29. (M) Why is the price so much cheaper than the American airline?
30. (F) ~~I can't imagine. It just is~~.
31. (M) Is it because it's the typhoon season?
32. We might get stranded in Japan.
33. (F) ~~Wouldn't that be awful?~~

34. ~~It is the typhoon season~~.
35. We're considering going at the end of August.
36. That's definitely the typhoon season in that part of the world. [150]
37. The Japanese seem very sanguine about flying in and out in spite of typhoons.
38. So I suppose we can assume it's safe.
39. It would be good to have the experience of catching a peak-hour train in the morning to a Japanese city, and then traveling around the Old City in Kyoto.
40. The connecting flight is direct to Los Angeles from Kanzai Airport in Osaka. [180]
41. ~~That works quite well~~.
42. It fits in with the connection from Los Angeles to Denver to Jackson Hole.
43. We want that connection because we want to go to Yellowstone National Park, and we have to go to Jackson Hole.
44. ~~The connections work well~~ with the Japanese flight arrival times.
45. (M) I'd select the Japan Airlines on that basis. [210]
46. (F) ~~I agree~~.
47. (M) Where are we meeting Bernard and Joy? (F) We're meeting them in Jackson Hole.
48. Their flight arrives an hour earlier, and they will meet our flight if we catch this connection.
49. ~~That works well~~.
50. They will hire a van.
51. It has five seats, because Chris is coming too.
52. Chris is their youngest son.
53. He'll probably do the driving, which is good. [240]
54. They've booked all the accommodation for the week in Yellowstone National Park.
55. We don't know what will happen with security at the American airports.
56. I understand flying between ports can be problematic.
57. They assume you're not on normal business/tourism if you fly ~~between American ports~~.
58. We'll have to take that chance.
59. (M) They assume all is well if you fly in on an American airline. [270]
60. (F) Correct, but we can't do anything about that.
61. What other national parks do you want to see?
62. (M) I certainly want to go to the Grand Canyon and Yosemite.
63. (F) That's in California.
64. (M) But if we go to the Grand Canyon we should also go to Bryce Canyon and Zion National Park. [300]

Appendix E: Scoring Sheet based on a Propositional Analysis of the Conversation Transcript

Below is the scoring sheet based on a propositional analysis, which was prepared according to descriptions and examples in Dixon et al. (1989).

Level	Recall	Proposition
1	___	1. (DISCOVER, WHAT, WOMAN)
2	___	2. (FROM, p1, TRAVEL AGENT)
2	___	3. (ABOUT, p1, TRIP)
3	___	4. (QUALIFIER, p3, UPCOMING)
4	___	5. (VISIT, WOMAN, TRAVEL AGENT)
2	___	6. (ADJECTIVE, INTERESTING, IT)
3	___	7. (QUALIFIER, p6, VERY)
1	___	8. (TALK, AIRLINES)
2	___	9. (p8, FIRST)
1	___	10. (CONSIDERED, QANTAS)
2	___	11. (p10, FIRST)
3	___	12. (BECAUSE, p10, SAFEST)
1	___	13. (SURE, WOMAN, NOT)
2	___	14. (TIME, p12, THESE DAYS)
3	___	15. (BUT, THEY_SAY, p12)
1	___	16. (PRICE, FLIGHTS, EXPENSIVE)
2	___	17. (QUALIFIER, FLIGHTS, QANTAS)
3	___	18. (QUALIFIER, p17, TO USA)
3	___	19. (QUALIFIER, p17, VIA JAPAN)
4	___	20. (QUALIFIER, p17, QANTAS TO USA)
4	___	21. (QUALIFIER, p17, TO USA VIA JAPAN)
4	___	22. (QUALIFIER, p17, QANTAS VIA JAPAN)
5	___	23. (QUALIFIER, p17, QANTAS TO USA VIA JAPAN)
2	___	24. (PRICE, FLIGHTS, $500)
3	___	25. (QUALIFIER, p24, MORE)
4	___	26. (COMPARISON, p25, OTHER FLIGHTS)
4	___	27. (COMPARISON, p25, AIRLINES)
5	___	28. (CONJUNCTION, AND, p26, p27)
1	___	29. (STATEMENT, DISADVANTAGE)
2	___	30. (THOUGHT, SHE, p29)
1	___	31. (CONSIDERED, UNITED AIRLINES)
2	___	32. (QUALIFY, p31, PRICE)
3	___	33. (QUALIFY, p32, REASONABLE)

```
1   ___  34. (LOCATION, SYDNEY)
2   ___  35. (MOTION, p31, p34, LEAVING)
3   ___  36. (TIME, p35)
4   ___  37. (QUALIFIER, p36, AWFUL)
    -------------------------------------
1   ___  38. (LOCATION, ADELAIDE)
2   ___  39. (MOTION, p31, p38, LEAVING)
2   ___  40. (TIME, EARLY)
3   ___  41. (QUALIFIER, p40, CRACK OF DAWN)
3   ___  42. (QUALIFIER, p40, 6 O'CLOCK)
4   ___  43. (CONJUNCTION, AND, p41, p42)
3   ___  44. (OBLIGATION, WE, p39) // have to
4   ___  45. (QUALIFIER, p44, WOULD)
    -------------------------------------
2   ___  46. (IN, WAIT, SYDNEY)
    -------------------------------------
1   ___  48. (LOCATION, JAPAN)
1   ___  49. (LOCATION, LOS ANGELES)
2   ___  50. (MOTION, FLIGHT, p48, FROM)
3   ___  51. (QUALIFIER, FLIGHT, UNITED AIRLINES)
2   ___  52. (GOOD PART)
3   ___  53. (QUALIFIER, p52, OF THAT OPTION)
4   ___  54. (CONCEPT, SECURITY)
5   ___  55. (CONJUNCTION, IN, p54, p49)
6   ___  56. (DESCRIPTION, p54, NO TROUBLE)
    -------------------------------------
1   ___  57. (JAPAN AIRLINES)
1   ___  58. (OTHER OPTION)
    -------------------------------------
1   ___  59. (PRICE, p57)
2   ___  60. (DESCRIPTION, TERRIFIC)
    -------------------------------------
1   ___  61. (LOCATION, KYOTO)
1   ___  62. (TIME, 24 HOURS)
2   ___  63. (IF, JAPAN AIRLINES, STAY IN p61)
3   ___  64. (OBLIGATION, p63) // would have to
    -------------------------------------
1   ___  65. (SELF-CORRECTION, "NO")
2   ___  66. (LOCATION, OSAKA)
2   ___  67. (NOT KYOTO)
    -------------------------------------
1   ___  68. (LOCATION, KANZAI AIRPORT)
    -------------------------------------
1   ___  69. (EVENT, FLIGHT ARRIVES)
1   ___  70. (TIME, 9PM)
2   ___  71. (QUALIFIER, p70, ABOUT)
2   ___  72. (CONJUNCTION, AT, p69, p70)
1   ___  73. (TIME, 6PM)
2   ___  74. (QUALIFIER, p73, NEXT DAY)
3   ___  75. (MOTION, LEAVES)
4   ___  76. (OBLIGATION, p75) // would have to
```

```
5   ___  77. (CONJUNCTION, p69, p75)
    -------------------------------------
1   ___  78. (JAPAN, VISITED)
2   ___  79. (QUALIFIER, p78, NEVER)
    -------------------------------------
1   ___  80. (VEHICLE, TRAIN)
2   ___  81. (MOTION, INTO, OSAKA)
3   ___  82. (BY, p80, p81)
1   ___  83. (TIME, DAY, MORNING)
2   ___  84. (QUALIFIER, p90, NEXT)
3   ___  85. (CONJUNCTION, CATCH, p80, p81)
1   ___  86. (AIRPORT HOTEL)
2   ___  87. (EVENT, STAY)
3   ___  88. (QUALIFIER, p94, BRIEF)
4   ___  89. (QUALIFIER, p95, VERY)
3   ___  90. (QUALIFIER, p94, OVERNIGHT)
4   ___  91. (QUALIFIER, p94, BRIEF AND OVERNIGHT)
5   ___  92. (QUALIFIER, p94, VERY AND BRIEF AND OVERNIGHT)
3   ___  93. (CONJUNCTION, AFTER, p87, p85)
    -------------------------------------
1   ___  94. (VEHICLE, BUS)
2   ___  95. (ACTION, SWITCH TO, p94)
    -------------------------------------
1   ___  96. (LOCATION, OLD CITY)
2   ___  97. (IN, p96, KYOTO)
2   ___  98. (ACTIVITY, TOUR)
3   ___  99. (QUALIFIER, p98, HALF-DAY)
4   ___ 100. (CONJUNCTION, WHICH, p98, INTERESTING)
5   ___ 101. (QUALIFIER, p100, VERY)
    -------------------------------------
2   ___ 102. (QUESTION, WHY, CHEAPER)
3   ___ 103. (COMPARISON, p102, AMERICAN AIRLINE)
    -------------------------------------
1   ___ 104. (QUESTION, IS IT BECAUSE)
2   ___ 105. (EVENT, TYPHOON)
3   ___ 106. (EVENT SERIES, p105, SEASON)
    -------------------------------------
2   ___ 107. (ACTIVITY, STRANDED IN JAPAN)
3   ___ 108. (MIGHT, p107)
    -------------------------------------
1   ___ 109. (TIME, MONTH, AUGUST)
2   ___ 110. (QUALIFIER, p109, END OF)
1   ___ 111. (ACTIVITY, GOING)
2   ___ 112. (CONJUCTION, AT, p111, p109)
3   ___ 113. (CONSIDERING, p112)
    -------------------------------------
1   ___ 114. (LOCATION, THAT PART OF THE WORLD)
2   ___ 115. (CONJUNCTION, IN, p106, p114)
    -------------------------------------
1   ___ 116. (RACE, JAPANESE)
2   ___ 117. (SEEM, p116, SANGUINE)
```

```
3   ___ 118. (QUALIFIER, p117, VERY)
1   ___ 119. (ACTIVITY, FLYING)
2   ___ 120. (QUALIFIER, p119, IN)
2   ___ 121. (QUALIFIER, p119, OUT)
3   ___ 122. (CONJUNCTION, AND, p120, p121)
3   ___ 123. (CONJUNCTION, IN SPITE OF, p116, p105)
    -------------------------------------
1   ___ 124. (STATE, SAFE, IT'S)
2   ___ 125. (ACTIVITY, ASSUME, p124)
3   ___ 126. (QUALIFIER, CAN, p125)
3   ___ 127. (QUALIFIER, SUPPOSE, p126)
4   ___ 128. (CONJUNCTION, THAT, p127, p125)
    -------------------------------------
2   ___ 129. (TIME, PEAK HOUR)
2   ___ 130. (ACTIVITY, CATCH, p80)
3   ___ 131. (CONJUNCTION, AT, p130, p129)
2   ___ 132. (TIME, DAY, MORNING)
3   ___ 133. (QUALIFIER, p129, p132)
2   ___ 134. (LOCATION, JAPANESE CITY)
3   ___ 135. (MOTION, TO, p134)
4   ___ 136. (CONJUNCTION, OF, EXPERIENCE, p130)
5   ___ 137. (QUALIFIER, GOOD, p136)
1   ___ 138. (LOCATION, OLD CITY)
2   ___ 139. (MOTION, TRAVELLING, p138)
2   ___ 140. (LOCATION, KYOTO)
3   ___ 141. (QUALIFIER, IN, p138, p141)
3   ___ 142. (QUALIFIER, THEN, p139)
    -------------------------------------
1   ___ 143. (LOCATION, LOS ANGELES)
2   ___ 144. (MOTION, TO, p143)
3   ___ 145. (QUALIFER, DIRECT, p144)
1   ___ 146. (LOCATION, AIRPORT)
2   ___ 147. (QUALIFER, KANZAI, p146)
3   ___ 148. (MOTION, FROM, p147)
2   ___ 149. (LOCATION, OSAKA)
3   ___ 150. (QUALIFER, IN, p149)
4   ___ 151. (CONJUNCTION, p148, p150)
1   ___ 152. (ACTIVITY, FLIGHT)
2   ___ 153. (QUALIFIER, CONNECTING, p152)
    -------------------------------------
1   ___ 154. (LOCATION, LOS ANGELES)
1   ___ 155. (NOUN, CONNECTION)
2   ___ 156. (MOTION, FROM, p155, p154)
1   ___ 157. (LOCATION, DENVER)
2   ___ 158. (MOTION, TO, p157)
1   ___ 159. (LOCATION, JACKSON HOLE)
2   ___ 160. (MOTION, TO, p159)
3   ___ 161. (ORDER, p157, p159)
3   ___ 162. (ADJECTIVE, FITS IN, p155)
    -------------------------------------
1   ___ 163. (NOUN, CONNECTION)
```

2 ___ 164. (QUALIFER, WE WANT, p163)
1 ___ 165. (LOCATION, YELLOWSTONE NATIONAL PARK)
2 ___ 166. (VERB, GO TO/VISIT, p165)
3 ___ 167. (CONJUNCTION, BECAUSE, p165, p163)
1 ___ 168. (LOCATION, JACKSON HOLE)
2 ___ 169. (VERB, GO TO/VISIT, p168)
3 ___ 170. (ADJECTIVE, HAVE TO/COMPULSION)
4 ___ 171. (CONJUCTION, BECAUSE, p165, p169)

---

1 ___ 172. (NOUN, AIRLINES)
2 ___ 173. (QUALIFIER, JAPANESE, p172)
2 ___ 174. (REFERENCE, SELECT, p172)
3 ___ 175. (CONJUNCTION, BECAUSE, p172, BASIS)

---

1 ___ 176. (VERB, MEETING)
2 ___ 177. (PERSON, BERNARD)
2 ___ 178. (PERSON, JOY)
3 ___ 179. (CONJUNCTION, AND, p177, p178)
1 ___ 180. (QUESTION, WHERE, p176, p179)
1 ___ 181. (LOCATION, JACKSON HOLE)
2 ___ 182. (MOTION, IN, p181)

---

1 ___ 183. (ARRIVAL, BERNARD & JOY)
2 ___ 184. (TIME, HOUR)
3 ___ 185. (QUANTITY, 1, p184)
4 ___ 186. (QUALIFIER, EARLIER, p184)
1 ___ 187. (PERSON, US)
1 ___ 188. (PERSON, THEY) // BERNARD & JOY
2 ___ 189. (MEET, p188, p187)
3 ___ 190. (VERB, CATCH CONNECTION/FLIGHT, p187)
4 ___ 191. (IF, p190, p189)

---

1 ___ 192. (VEHICLE, VAN)
1 ___ 193. (PERSON, THEY) // BERNARD & JOY
2 ___ 194. (VERB, HIRE, p193, p192)
3 ___ 195. (QUALIFIER, WILL/FUTURE, p194)
1 ___ 196. (HAS, p192, SEATS)
2 ___ 197. (QUANTITY, 5, p196)
2 ___ 198. (PERSON, CHRIS)
3 ___ 199. (VERB, COMING)
4 ___ 200. (QUALIFIER, TOO, p199)
3 ___ 201. (CONJUNCTION, BECAUSE, p198, p197)

---

1 ___ 202. (PERSON, CHRIS)
2 ___ 203. (RELATIONSHIP, THEY, p202, SON)
3 ___ 204. (QUALIFIER, YOUNGEST, p203)
2 ___ 205. (VERB, DRIVING, p202)
3 ___ 206. (QUALIFIER, PROBABLY, p205)

---

1 ___ 207. (LOCATION, YELLOWSTONE NATIONAL PARK)
2 ___ 208. (QUALIFIER, IN, p207)

```
3   ___ 209. (NOUN, ACCOMODATION, p207)
2   ___ 210. (PERSON, THEY) // BERNARD & JOY
3   ___ 211. (VERB, BOOKED, p209)
4   ___ 212. (QUALIFIER, ALL, p211)
3   ___ 213. (TIME, WEEK)
4   ___ 214. (QUANTITY, 1, p213)
    -------------------------------------
1   ___ 215. (LOCATION, AIRPORTS)
2   ___ 216. (QUALIFIER, AMERICAN, p215)
1   ___ 217. (PERSON, SECURITY)
2   ___ 218. (CONJUNCTION, AT, p217, p215)
3   ___ 219. (PERSON, WE)
4   ___ 220. (ADJECTIVE, UNSURE, p219)
5   ___ 221. (CONJUCTION, ABOUT, p220, p217)
    -------------------------------------
1   ___ 222. (LOCATION, AIRPORTS)
2   ___ 223. (MOTION, BETWEEN, p222)
2   ___ 224. (ACTIVITY, FLYING, p223)
3   ___ 225. (ADJECTIVE, PROBLEMATIC, p224)
4   ___ 226. (QUALIFIER, CAN BE, p225)
5   ___ 227. (QUALIFIER, I UNDERSTAND THAT, p226)
    -------------------------------------
2   ___ 228. (PERSON, SECURITY)
2   ___ 229. (ACTIVITY, BUSINESS)
2   ___ 230. (ACTIVITY, TOURISM)
3   ___ 231. (QUALIFIER, NORMAL, p229/p230)
4   ___ 232. (QUALIFIER, NOT, p231)
5   ___ 233. (VERB, ASSUME, p228, p232)
1   ___ 234. (LOCATION, AIRPORTS)
2   ___ 235. (MOTION, BETWEEN, p234)
2   ___ 236. (ACTIVITY, FLYING, p235)
6   ___ 234. (CONDITION, IF, p236, p233)
    -------------------------------------
1   ___ 235. (NOUN, CHANCE/RISK)
1   ___ 236. (PERSON, WE)
2   ___ 237. (VERB, TAKE, p236, p235)
3   ___ 238. (QUALIFIER, MUST/HAVE TO, p237)
    -------------------------------------
2   ___ 239. (PERSON, SECURITY)
2   ___ 240. (ADJECTIVE, ACCEPTING/PERMISSIVE, p239)
1   ___ 241. (ACTIVITY, FLY)
2   ___ 242. (QUALIFIER, IN, p241)
1   ___ 243. (NOUN, AIRLINE)
2   ___ 244. (QUALIFIER, AMERICAN, p243)
3   ___ 245. (CONJUCTIVE, ON, p241, p243)
4   ___ 246. (CONDITION, IF, p240, p245)
    -------------------------------------
1   ___ 247. (PERSON, WE)
2   ___ 248. (ADJECTIVE, POWERLESS)
3   ___ 249. (PERSON/ACTIVITY, SECURITY)
4   ___ 250. (CONJUNCTIVE, ABOUT, p247, p249)
```

```
    ------------------------------------
1   ___ 251. (LOCATION, NATIONAL PARKS)
2   ___ 252. (QUALIFIER, OTHER, p251)
2   ___ 253. (QUESTION, WHAT, p252)
2   ___ 254. (VERB, VISIT, p251)
    ------------------------------------
1   ___ 255. (LOCATION, GRAND CANYON)
1   ___ 256. (LOCATION, YOSEMITE)
2   ___ 257. (CONJUNCTIVE, AND, p255, p256)
2   ___ 258. (VERB, VISIT, p257)
1   ___ 259. (PERSON, I) // he
2   ___ 260. (VERB, DESIRE, p259, p258)
3   ___ 261. (QUALIFIER, CERTAINLY, p260)
    ------------------------------------
1   ___ 262. (LOCATION, YOSEMITE)
2   ___ 263. (LOCATION, CALIFORNIA)
3   ___ 264. (CONJUNCTIVE, IN, p262, p263)
    ------------------------------------
1   ___ 265. (LOCATION, BRYCE CANYON)
1   ___ 266. (LOCATION, ZION NATIONAL PARK)
2   ___ 267. (CONJUNCTIVE, p265, p266)
2   ___ 268. (VERB, VISIT, p267)
3   ___ 269. (QUALIFIER, ALSO, p268)
3   ___ 270. (QUALIFIER, SHOULD, p269)
1   ___ 271. (LOCATION, GRAND CANYON)
2   ___ 272. (VERB, VISIT, WE, p271)
3   ___ 273. (CONDITION, IF, p272, p268)
```

Appendix F: Experience Questionnaire

"Effect of instant messaging on recall during video-mediated briefings"

Questionnaire 1

ID: _____ Age in years: _____ Gender: M / F
(Circle the appropriate option)

These questions concern experience with technology.	Never	Very Rarely	Rarely	Occasion-ally	Frequently	Very Frequently
1. How often do you use video teleconferencing for group-to-group collaboration (as opposed to one-on-one collaboration)?	0	1	2	3	4	5
2. How often do you use video teleconferencing with military colleagues?	0	1	2	3	4	5
3. How often do you use computers during meetings (e.g. for minute or note taking)?	0	1	2	3	4	5
4. How often do you use paper for note-taking during meetings?	0	1	2	3	4	5
5. How often do you use instant messaging software (e.g. Microsoft Live Communicator, IRC, ICQ, Aim)?	0	1	2	3	4	5
6. How often do you use Sticker as your instant messaging application?	0	1	2	3	4	5
7. How often do you use the Short Messaging Service (SMS) on mobile phones?	0	1	2	3	4	5
8. How often do you watch television programs or view websites that make use of scrolling headline displays (e.g. CNN, Sunrise on Channel 7, CNNNN on ABC, Today on Channel 9, news.bbc.co.uk)?	0	1	2	3	4	5

The final question concerns likes and dislikes.

	Not at all ⟷ Very much/A lot					
9. How much do you like gadgets (e.g. new mobile phones, PDAs, cameras, GPS devices, and other electronic equipment)?	0	1	2	3	4	5

Appendix G: Story 16 Recognition Questionnaire

Questionnaire regarding flight story

Do the following statements give an idea from the story?	**Definitely Yes**	**Probably Yes**	**Possibly Yes**	**Uncertain**	**Possibly No**	**Probably No**	**Definitely No**
1. Margaret did not think the movie was very funny.	1	2	3	4	5	6	7
2. This is Margaret's first aeroplane flight.	1	2	3	4	5	6	7
3. The man sitting next to Margaret wanted to watch the movie during the flight.	1	2	3	4	5	6	7
4. More than ten flight attendants served the passengers.	1	2	3	4	5	6	7
5. The passengers were asked to close their window shades.	1	2	3	4	5	6	7
6. Margaret's son is a fisherman in Seattle.	1	2	3	4	5	6	7
7. The movie in her area was a classic Hollywood drama.	1	2	3	4	5	6	7
8. Margaret had never seen a Boeing 747 before.	1	2	3	4	5	6	7
9. Margaret had thought it would be nice to look out the window.	1	2	3	4	5	6	7
10. Margaret's son will meet her at the airport.	1	2	3	4	5	6	7
11. The man sitting next to Margaret is a lawyer.	1	2	3	4	5	6	7
12. She and her neighbour talked after the movie.	1	2	3	4	5	6	7

Do the following statements give an idea from the story?	**Definitely Yes**	**Probably Yes**	**Possibly Yes**	**Uncertain**	**Possibly No**	**Probably No**	**Definitely No**
13. Not all passengers chose to pay extra to watch the movie.	1	2	3	4	5	6	7
14. Margaret has ridden on trains before.	1	2	3	4	5	6	7
15. She was pleased that her earphones were perfectly comfortable.	1	2	3	4	5	6	7
16. She felt tired after the movie.	1	2	3	4	5	6	7
17. Margaret usually is not very talkative.	1	2	3	4	5	6	7
18. Margaret was nervous at first.	1	2	3	4	5	6	7
19. If this trip is successful she may consider another aeroplane trip.	1	2	3	4	5	6	7
20. Margaret has known many people her age who have flown.	1	2	3	4	5	6	7
21. Margaret had a non-stop flight from Hartford.	1	2	3	4	5	6	7
22. The coach passengers were served cold cuts and fruit for lunch.	1	2	3	4	5	6	7
23. Margaret is returning home in two weeks.	1	2	3	4	5	6	7
24. The man sitting next to Margaret is a New York accountant.	1	2	3	4	5	6	7

Appendix H: Story 17 Recognition Questionnaire

Questionnaire regarding card story.

Do the following statements give an idea from the story?	**Definitely Yes**	**Probably Yes**	**Possibly Yes**	**Uncertain**	**Possibly No**	**Probably No**	**Definitely No**
1. Two weeks ago they also played Crazy Eights.	1	2	3	4	5	6	7
2. There is no women's singles competition at the Canasta tournament.	1	2	3	4	5	6	7
3. There are many couples playing Canasta in their town.	1	2	3	4	5	6	7
4. The couples know how to play numerous board games.	1	2	3	4	5	6	7
5. Tom almost quit after losing the first game.	1	2	3	4	5	6	7
6. Tom and Barb are hosting the Canasta Club meeting.	1	2	3	4	5	6	7
7. Next year Tom and Barb will enter the couples' competition of the tournament.	1	2	3	4	5	6	7
8. Barb checks to be sure everyone has refreshments.	1	2	3	4	5	6	7
9. They have been playing Canasta for over 45 years.	1	2	3	4	5	6	7
10. Barb and Pat beat Tom and Joe in the first game.	1	2	3	4	5	6	7
11. The couples are all good friends.	1	2	3	4	5	6	7
12. Everyone in the club is at a somewhat similar level of skill.	1	2	3	4	5	6	7

Do the following statements give an idea from the story?	**Definitely Yes**	**Probably Yes**	**Possibly Yes**	**Uncertain**	**Possibly No**	**Probably No**	**Definitely No**
13. Barb and Pat are way ahead in the second game.	1	2	3	4	5	6	7
14. For tonight Barb is serving fresh cakes, pies and homemade icecream.	1	2	3	4	5	6	7
15. Barb and Pat won the women's doubles championship last year.	1	2	3	4	5	6	7
16. Sometimes the guests bring snacks to contribute to the refreshments.	1	2	3	4	5	6	7
17. Usually the Canasta meetings last until well after midnight.	1	2	3	4	5	6	7
18. Everyone is unhappy when someone misses a meeting.	1	2	3	4	5	6	7
19. The three couples are all middle aged or older.	1	2	3	4	5	6	7
20. Tom and Barb learned to play Canasta after they were married.	1	2	3	4	5	6	7
21. Tom makes a mistake early in the first game.	1	2	3	4	5	6	7
22. Their club rarely cancels meetings.	1	2	3	4	5	6	7
23. To drink, Barb and Tom are offering coffee, soft drinks, and wine.	1	2	3	4	5	6	7
24. They usually play the men against the women.	1	2	3	4	5	6	7

Appendix I: Conversation Recognition Questionnaire

Questionnaire regarding conversation.

Do the following statements give an idea from the story?	**Definitely Yes**	**Probably Yes**	**Possibly Yes**	**Uncertain**	**Possibly No**	**Probably No**	**Definitely No**
1. They are meeting friends called George and Mildred.	1	2	3	4	5	6	7
2. He suggests phoning the travel agent to book the Japan Airlines flights immediately.	1	2	3	4	5	6	7
3. He wants to see Yosemite in California.	1	2	3	4	5	6	7
4. They will have to stay in Los Angeles for a day due to poor connecting flights.	1	2	3	4	5	6	7
5. He is curious why the Japanese flights are so much cheaper than the others.	1	2	3	4	5	6	7
6. The good part about United Airlines was that security in Los Angeles would be less of an issue.	1	2	3	4	5	6	7
7. The United Airlines flights were $2500.	1	2	3	4	5	6	7
8. They will see Mt Fuji when they take a tour of the Old City in Kyoto.	1	2	3	4	5	6	7
9. Their friends have booked the accommodation in Yellowstone National Park.	1	2	3	4	5	6	7
10. They're considering flying through Japan in May.	1	2	3	4	5	6	7
11. She is not sure QANTAS is the still the safest airline.	1	2	3	4	5	6	7
12. She wants to fly a helicopter over the Grand Canyon.	1	2	3	4	5	6	7

Do the following statements give an idea from the story?	**Definitely Yes**	**Probably Yes**	**Possibly Yes**	**Uncertain**	**Possibly No**	**Probably No**	**Definitely No**
13. The QANTAS flights from Japan to the USA go via Hawaii.	1	2	3	4	5	6	7
14. The Japan Airlines flights required a stay in Kyoto of 72 hours.	1	2	3	4	5	6	7
15. Japan experiences at least 6 typhoons each year.	1	2	3	4	5	6	7
16. The United Airlines flights leave from Sydney.	1	2	3	4	5	6	7
17. The conversation with the travel agent started off with potential destinations.	1	2	3	4	5	6	7
18. The van hired by their friends will be a Winnebago.	1	2	3	4	5	6	7
19. They must get to Jackson Hole to get to Yellowstone National Park.	1	2	3	4	5	6	7
20. They will fly into Narita airport in Tokyo.	1	2	3	4	5	6	7
21. They are meeting Bernard and Joy in Jackson Hole.	1	2	3	4	5	6	7
22. Security will not be a problem anywhere.	1	2	3	4	5	6	7
23. American airport security will search their luggage if they travel on domestic American flights.	1	2	3	4	5	6	7
24. Chris is their friends' oldest daughter.	1	2	3	4	5	6	7

Appendix J: Interruption Questions

These are the questions sent to participants in IM groups QUESTIONS and QUESTIONS + NOISE during their videos. Each question was sent out at a pre-calculated time and relates to the previous 30 seconds of footage, and is intended to simulate a question from a co-viewer regarding what has just been seen.

J.1. Questions for "Stories" video

(S) What card game are they playing again?
(S) Who is the sore loser at the meeting tonight?
(S) What time do the parties normally start?
(S) How often is there a Canasta tournament in their community?
(S) How well did Pat and Barb do in the women's doubles?
(S) What time did Margaret leave in the morning?
(S) What is the name of the airport in New York City?
(S) What are they doing now lunch is over?
(S) What did the supervisor request people do during the movie?
(S) Where was the man sitting next to Margaret from?

J.2. Questions for "Conversation" video

(C) Who did they speak to about the upcoming trip?
(C) What time would they need to leave Adelaide if they flew American Airlines?
(C) What was the good thing about flying United Airlines from Japan?
(C) If they stayed in Japan, which hotel would they use?
(C) If they switch to a bus from the train, where will they go?
(C) What time will they catch the train to Osaka?
(C) Which national park do they want to visit?
(C) Who will meet their plane at Jackson Hole?
(C) How long will they stay in Yellowstone National Park?
(C) Where does he have family they will visit?

Appendix K: Project Information Sheet

EFFECT OF INSTANT MESSAGING ON RECALL DURING VIDEO-MEDIATED BRIEFINGS

In recent years C3ID has developed the LiveSpaces smart meeting room conceptual environment and has created several LiveSpace instances, called C2 Developmental BattleLabs, in DSTO-managed environments at operational military sites. These smart meeting rooms provide support for collaboration but are also environments for military users to experiment and evaluate tools and technologies not available in their standard operating environments. Software-based video teleconferencing and instant messaging applications are common features of the rooms.

Although much better than using only telephones, video teleconferencing in these environments does not provide the same experience for collaboration as being collocated with colleagues. My CEI Masters project is aimed at examining what software tools could be devised to make up for some of the shortfall. In particular, instant messaging is an obvious and topical choice for supporting side conversations during collaboration, and is already available in the BattleLabs, but we suspect it may inadvertently attract excessive attention while conversations or presentations are being conducted via the video links.

The purpose of this experiment is to determine how instant messaging affects the attention of people being presented to via a video link.

Firstly participants will be asked to fill out a questionnaire regarding experience with communication technologies prior to the day of the experiment. On that day they will perform a short typing speed test and watch two short videos on a computer workstation. After viewing each video, participants will perform simple arithmetic for two minutes. They will then be asked to recall as much of the videos as they can in five minutes. On the following day they will be asked to fill out a second questionnaire, which tests the recall of the videos again.

Some participants will be using an instant messaging application while watching the videos.

Other than minding the ramp up into the laboratory environment (the Intense Collaboration Space, 2.G.59, 205L), there are no significant risks for participants greater than what they experience at their desks each day.

Participants will be given ID numbers to ensure the anonymity of their data and results will only be viewed in their raw form by the principal researchers. The anonymised raw data will be retained for at least seven years, but will only be used for the purposes of this study and related publications and nothing more without explicit permission.

Participation is entirely voluntary and consent to participate can be withdrawn at any time with no detriment to career. There is no DSTO task to which to charge this activity but it is unlikely that more than a total 60 minutes of any individual's time will be required.

If you have any questions about the experiment and/or would like to participate, please contact Derek Weber (08 8259 7699) or Michael Broughton (08 8259 7125).

Should you have any complaints or concerns about the manner in which this project is conducted, please do not hesitate to contact the researchers in person, or you may prefer to contact the Australian Defence Human Research Ethics Committee at the following address:

Executive Secretary
Australian Defence Human Research Ethics Committee
CP2-7-66
Department of Defence
CANBERRA ACT 2600
Telephone: (02) 6266 3837 Fax: (02) 6266 4982
E-mail: ADHREC@defence.gov.au

Appendix L: Questionnaire Consent Form

CONSENT for questionnaire for
"Effect of instant messaging on recall during video-mediated briefings"

I,.....……….., give my consent to participate in the project mentioned above on the following basis:

I have had explained to me the aims of this research project, how it will be conducted and my role in it.

I understand the risks involved.

I understand that:

- Participation in the study is entirely voluntary and there is no obligation to take part;
- Had I not volunteered to participate there would be no detriment to my career; and
- I may withdraw at any time with no detriment to my career.

I am consenting to take part in this experiment on the understanding that my identity will not be provided to anyone other than the principal researcher and his supervisors, and that reports of this study will maintain my anonymity.

I have been given a copy of the project information sheet to keep, and a copy of DSTO's *Guidelines for Volunteers*.

Should I have any complaints or concerns about the manner in which this project is conducted I will contact the researchers in person, or contact the Australian Defence Human Research Ethics Committee at the following address:

Executive Secretary
Australian Defence Human Research Ethics Committee
CP2-7-66
Department of Defence
CANBERRA ACT 2600
Telephone: (02) 6266 3837 Fax: (02) 6266 4982
E-mail: ADHREC@defence.gov.au

…… ……
[Participant] [date] [Principal Researcher] [date]

Appendix M: Experiment Consent Form

CONSENT for experiment participation in
"Effect of instant messaging on recall during video-mediated briefings"

I,.....……..., give my consent to participate in the project mentioned above on the following basis:

I have had explained to me the aims of this research project, how it will be conducted and my role in it.

I understand the risks involved.

I understand that:

- Participation in the study is entirely voluntary and there is no obligation to take part;
- Had I not volunteered to participate there would be no detriment to my career; and
- I may withdraw at any time with no detriment to my career.

I am consenting to take part in this experiment on the understanding that my identity will not be provided to anyone other than the principal researcher and his supervisors, and that reports of this study will maintain my anonymity.

I have been given a copy of the information sheet to keep, and a copy of DSTO's *Guidelines for Volunteers*.

Should I have any complaints or concerns about the manner in which this project is conducted I will contact the researchers in person, or contact the Australian Defence Human Research Ethics Committee at the following address:

Executive Secretary
Australian Defence Human Research Ethics Committee
CP2-7-66
Department of Defence
CANBERRA ACT 2600
Telephone: (02) 6266 3837 Fax: (02) 6266 4982
E-mail: ADHREC@defence.gov.au

…….. ……..
[Participant] [date] [Principal Researcher] [date]

Appendix N: Actor Consent Form

Consent form for story reader

CONSENT

1. I, …… *(please print name)*

   consent to take part in the research project entitled

   Effect of instant messaging on recall during video-mediated briefings

   being carried out by staff in the Intense Collaboration Space (2.G.59, 205L), DSTO Edinburgh in 2008.

2. I agree to have video footage taken of me as part of this experiment for the purposes of data collection and analysis.
3. I understand that the video footage taken of me will only be used for the purposes of this experiment.
4. I AGREE / DO NOT AGREE to video footage of me which is relevant to the study being used in any presentations arising from this study.

5. ……
   (signature) (date)

6. This reader has been informed and understands the proposed procedure and restrictions on use of photographic material.

   Researcher:……
   (signature) (date)

   Witness:……
   (signature) (date)

Appendix O: DSTO Guidelines for Volunteers

GUIDELINES FOR VOLUNTEERS
DSTO PROCEDURES FOR RESEARCH INVOLVING HUMAN PARTICIPANTS

Thank you for taking part in Defence Research. Your involvement is much appreciated. This pamphlet explains your rights as a volunteer.

ADF Pamphlet 1.2.5.3 sets out guidelines for Defence research involving human beings. The Australian Defence Human Research Ethics Committee (ADHREC) exists to review Defence proposals for scientific research involving humans. The ADFP 1.2.5.3 does not require ADHREC to review all proposals for experimentation involving human participants. DSTO follows an internal review process to ensure its research conforms to ADF 1.2.5.3 and is reviewed by ADHREC when appropriate.

What is ADHREC?

- ADHREC is the Australian Defence Human Research Ethics Committee. It was established in 1988, as the Australian Defence Medical Ethics Committee (ADMEC), to make sure that Defence complied with accepted guidelines for research involving human beings.
- After World War II, there was concern around the world about human experimentation. The Declaration of Helsinki was made in 1964, which provided the basic principles to be followed wherever humans were used in research projects.
- The National Health & Medical Research Council in Australia (NHMRC) published a set of guidelines in 1982 for how human research should be carried out.
- ADHREC follows both the Declaration of Helsinki and the NHMRC Guidelines.

DSTO process

- DSTO has developed an approval process for ensuring that research involving humans complies with the comprehensive guidelines provided in Australian Defence Force Publication (ADFP) 1.2.5.3 entitled “Health and Human Performance Research in Defence – Manual for Researchers”.
- If you are told that the project has DSTO approval, what that means is that the DSTO S&T Activity Review Team has reviewed the research proposal and has agreed that, in accordance with ADFP 1.2.5.3 paragraph 1.13, ethical clearance through ADHREC is not required. In addition, a Research Leader has reviewed the research proposal and is satisfied that safety and ethical issues relating to informed consent, confidentiality and security of data have been addressed.
- DSTO approval does not imply any obligation on commanders to order or encourage their military personnel to participate, or to release military personnel from their usual workplace to participate. Obviously, the use of any particular military personnel must have clearance from their commanders but commanders should not use DSTO approval to pressure military personnel into volunteering.

Voluntary participation

- As you are a volunteer for this research project, you are under no obligation to participate or continue to participate. You may withdraw from the project at any time without detriment to your military career or to your medical care.
- At no time must you feel pressured to participate or to continue if you do not wish to do so.
- If you do not wish to continue, it would be useful to the researcher to know why, but you are under no obligation to give reasons for not wanting to continue.

Informed consent

- Before commencing the project you will have been given an information sheet which explains the project, your role in it and any risks to which you may be exposed.
- You must be sure that you understand the information given to you and that you ask the researchers about anything of which you are not sure.
- If you are satisfied that you understand the information sheet and agree to participate, you should initial every page of the information sheet and keep a copy.
- Before you participate in the project you should also have been given a consent form to sign. You must be happy that the consent form is easy to understand and spells out to what you are agreeing. Again, you should keep a copy of the signed consent form.

Complaints

- If at any time during your participation in the project you are worried about how the project is being run or how you are being treated, then you should speak to the researchers.

  Contact details:
  Mr Derek Weber
  Command, Control, Communication & Intelligence Division
  DSTO Edn
  Ph 08 825 97699
  Fax number: 08 825 95619
  Email address: derek.weber@defence.gov.au

- If you don't feel comfortable doing this, you can contact the Executive Secretary of ADHREC.

  Contact details are:
  Executive Secretary
  Australian Defence Human Research Ethics Committee
  CP2-7-66
  Department of Defence
  CANBERRA ACT 2600
  Ph: 02 62663837
  Fax: 02 62664982
  E-mail: ADHREC@defence.gov.au

More information

- If you would like to read more about ADHREC, you can look up the following references on the Defence Manager's Toolbox or on DEFWEB

  – DI(G)ADMIN 24-3 *Function, Structure and Procedures for Obtaining Clearance for Research from Australian Defence Medical Ethics Committee* (or as amended)
  – HPD 205 *Australian Defence Medical Ethics Committee* (or as amended)
  – ADFP733 *Health and Human Performance Research in the Australian Defence Organisation – Manual for Researchers*

  Or, visit the ADHREC web site at http://defweb2.cbr.defence.gov.au/dpedhs/default.htm

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE

Effects of Instant Messaging on Recall During Video-Mediated Briefings

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document	(U)
Title	(U)
Abstract	(U)

4. AUTHOR(S)

Derek Weber

5. CORPORATE AUTHOR

DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-RR-0380	AR-015-275	Research Report	April 2012

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2010/1044833/1	07/313	CDS	83	68

13. DSTO Publications Repository

http://dspace.dsto.defence.gov.au/dspace/

14. RELEASE AUTHORITY

Chief, Command, Control, Communications and Intelligence Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.shtml

CSCW, Distributed collaboration environments, Experimentation, Human Factors, Ergonomics

19. ABSTRACT

This report explores the effect of interruption from instant messaging (IM) on memory of a video-teleconference brief. A 4x2 factorial pilot study was conducted with 32 participants (gender balanced), using four levels of interruption and gender as the independent variables. Two videos were presented to participants, one of a single person speaking and one of a pair in conversation. Memory was tested with five minutes free recall after each video and recognition questionnaires 24 hours later. Analysis revealed that women performed better than men in free recall, that requiring participants to respond to questions degraded their free recall, and that interruption caused a significant downward trend in free recall performance and confidence. Recognition performance results were inconclusive, however. We recommend repeating this study with more participants. Results obtained could help advise Defence on the development of usage policies for collaboration technologies, specifically regarding limiting the types and number of sources of interaction and sensory input where possible.