Integration of Spatial and Linguistic Information in Virtual Search and Rescue Missions

Yan Shan Tai

A thesis submitted for the degree of

Doctor of Philosophy

Cardiff University

August 2025

Summary

Positive outcomes in safety- and time-critical situations, such as during fire emergencies, hinge on coordinated responses, which in turn depend on the ability of response teams to interpret new information and adapt their behaviour accordingly. However, communication habits that sacrifice directness, such as conversational implicatures, can emerge and be misinterpreted under pressure, which might in turn impair behavioural adaptation. This thesis investigated whether these effects extend to naturalistic, high-stakes, and high-stress scenarios using undergraduate participants who acted as firefighters in desktop-based and semi-immersive virtual reality simulations of search and rescue (SAR) scenarios. Chapter 2 reports three experiments that investigated the conditions under which behaviours adapt in response to goal-relevant information that contradicted prior knowledge or experience. The findings showed that in these scenarios, critical information conveyed through conversational implicatures was often misinterpreted unless supported by explicit hints, leading to ineffective behavioural adaptation during SAR missions even when they had time to process the messages and reformulate strategies. Chapter 4 reports an experiment investigating how self-generated expectations and explicitly instructed expectations about the prevalence of explosive hazards interact during visual search under high and low stress conditions. Despite direct instructions that target prevalence would be lower, search performance (i.e., false positives) was influenced by prior experience in high prevalence, which had formed self-generated expectations of high target prevalence. Together, these findings suggest that under pressure, the provision of new and critical information does not guarantee its integration into goal-directed behaviour. Instead, its uptake depends on an interplay of factors such as the directness of communication, the engagement of pragmatic reasoning, the strength of prior knowledge and experience, opportunities for deliberate decision-making, and the limits imposed by stress on executive functioning supporting adaptive behaviour. While further

research with career firefighters is needed, this work highlights the potential value of training protocols that emphasise explicit communication and awareness of the biasing effects of prior experience or expectations.

Acknowledgements

This thesis would not have been possible without the support that I received from several individuals who, in one way or another, have contributed to its completion. I am first and foremost indebted to my wonderful supervisors, Professor Rob Honey and Dr Jacques Grange, for their continuous guidance, patience, and faith in me throughout this work. Their mentorship paved the way toward completing this thesis and helped me grow into a better researcher. I am also grateful to my co-supervisor, Dr Sabrina Cohen-Hatton, for being an integral part of my supervisory team. I would like to thank Dr Phil Butler for his invaluable insights as a veteran firefighter and for his assistance in developing my simulated search-and-rescue scenarios.

To the members of the IROHMS lab past and present, thank you for the collegial atmosphere that made work incredibly enjoyable. I would also like to thank my lab mate, Victoria, for her friendship, moral support, and shared laughter, which were a constant source of positivity throughout my PhD, especially during the later stages. I am also grateful to Carly for her helpful presence in the lab, whose assistance with various tasks contributed to the smooth running of my experiments.

For the care and provision they have given me throughout my life, I would like to thank my parents and extended family. Last but not least, to my dear friends back home in Malaysia: Audrey, Sabrena, Wen Jia, Crystal, Shyuan Mey, Nigel, Hooi Fang, and more… Thank you for always being there, even when I am difficult. Without your unwavering friendship, I would not have been able to persevere through challenges over the years.

Funding: This thesis was supported by an EPSRC DTP studentship (EP/T517951/1-2598935).

Manuscripts Under Review for Publication

- Chapter 2 is based on an unpublished article: Tai, Y. S., & Honey, R. C. (2025). The interface between pragmatics and wayfinding in simulated search and rescue environments.
- Chapter 4 is based on an unpublished article: Tai, Y. S., Grange, J. A., & Honey, R. C. (2025). Self-generated expectations of target prevalence influence visual search during virtual hazard search.

Conference Presentation

- Tai, Y. S., Grange, J. A., Honey, R. C., & Cohen-Hatton, S. (2023, July 20-24). *How expectations affect performance in a simulated fire and rescue environment* [Conference presentation]. 14th International Conference on Applied Human Factors and Ergonomics (AHFE 2023), San Francisco, CA, United States.

List of Figures

Chapter 2

one of the managers was not in the building, and a third control group received redundant explicit or implicit information.

Chapter 3

List of Tables

Chapter 2

Chapter 4

Table of Contents

Chapter 1:

General Introduction

1.1 Context

On the evening of 6 April 2010, fire and smoke engulfed Flat 72 of Shirley Towers, a high-rise residential building in Southampton. Fire crews struggled to navigate the complex "scissor" layout of the building where each flat spanned three floors: To access the main living areas of a flat, one had to either ascend or descend the stairs upon entry. Signage labelling the entry system of the flat (i.e., either up or down) was obscured by thick smoke, creating disorientation and confusion for the fire crews. Amid the chaos, two firefighters, Alan Bannon and James Shears, died in the line of duty due to excessive heat exposure after becoming entangled in fallen cables encased in plastic.

While the immediate circumstances of their demise led to changes to the British Standard regulations, specifically requiring electrical cables to be supported in metal trunking, investigations also identified additional factors that compounded the firefighting response, none of which could be easily remedied by a single regulatory measure. These included challenges in navigating the atypical layout of the building, reliance on prior experience, failure to verify information, and inadequate briefing (Hampshire Fire and Rescue Service, 2013). For instance, assumptions based on prior experience and communication failures appeared to have led the incident commander to believe that Flat 72 was on the seventh floor when instead it was on the ninth. Although he requested the bridgehead to be set up two floors below on the fifth floor, it was mistakenly established on the seventh floor (i.e., two floors below the actual location, which happened to be appropriate by coincidence). Most crews remained unclear about which floor they were on throughout the incident (Hampshire Fire and Rescue Service, 2013), likely contributing to communication difficulties and confusion. This overreliance on prior experience was further exemplified by a firefighter who, relying on his knowledge of Shirley Towers, chose a right-hand search strategy that bypassed the lounge and kitchen in an attempt to reach the bathroom and bedrooms upstairs (Hampshire Fire and Rescue Service, 2013); a right-hand search strategy involves systematically searching a building with the right hand maintaining contact with the wall and "sweeping" the immediate area with the left hand for victims (Head of Operational Procedures, 2012). Despite having access to a thermal imaging camera, he did not use it to confirm whether his strategy was appropriate and, consequently, missed the fire in the lounge. The fire grew larger and engulfed the lounge as the crew moved deeper into the flat, melting the plastic trunking and causing the cables to fall, ultimately trapping Bannon and Shears and claiming their lives (Hampshire Fire and Rescue Service, 2013).

There were also two separate instances of failure to verify information during the early stages of the incident. In the first instance, the Control Operator assumed that the flat was empty simply because the caller was not the occupant (Hampshire Fire and Rescue Service, 2013). Critically, key details, such as the position of the flat within the building and the location of the fire, were not communicated to the attending incident commander, resulting in decisions being made based on incomplete or assumed information. Firefighters later met the occupant of the flat on the incident ground but did not confirm whether anyone else was in the building or where the fire was located (Hampshire Fire and Rescue Service, 2013). Had they done that, they would have known to search the lounge for the fire instead of bypassing that area and might have prevented the escalation of the incident.

These events reveal more than just operational failures. Across multiple junctures in the incident, information was not completely absent but was present and overlooked, or could have been obtained but was not. These patterns suggest that in high-stakes, high stress scenarios, new information, especially when in conflict with prior expectations or when opportunities to obtain or verify it are not utilised, might not always translate into behavioural adaptation. The death of two firefighters in the Shirley Tower fires underscores the severe consequences of such failures, but the cognitive tendencies that contributed to the tragedy are unlikely to be unique to this incident alone.

High-stakes operational settings are often defined by high levels of uncertainty and dynamic conditions, with information that is often incomplete, ambiguous, contradictory, or entirely absent. These situations are typically time-critical, forcing first responders or operators to make rapid decisions under the stress of the unfolding incidents and within a narrow margin of error. In such contexts, prior knowledge and assumptions derived from past experience form expectations that might influence whether and how new information is interpreted and utilised. As exemplified in the Shirley Towers fire, an incorrect decision could delay critical incident responses, exacerbate the situation, and lead to adverse outcomes including death. Thus, the likelihood of attaining favourable incident outcomes hinges on one's capacity to correctly interpret, evaluate, and act upon new information that might contradict their existing expectations. This process, however, is reliant on the availability of cognitive resources, which are likely constrained under physiological and psychological stress during such incidents (see Section 1.4). Given these constraints, understanding these failures – where relevant information is overlooked or unexamined – requires an examination of the conditions that support the effective use of new information in such contexts, and how behaviour can be influenced by the way in which new information is presented or framed. This question forms the basis of the research reported in this thesis.

This chapter examines a range of cognitive processes that are relevant to how information is assimilated in high stakes conditions, which is at the core of the empirical work presented in this thesis. It begins by exploring how individuals adapt their behaviour according to changing task demands, emphasising the role of executive control. It then explores the use of pragmatic reasoning in everyday communication, considering how different ways of presenting information affect their interpretation, and how stress and other cognitive constraints might disrupt this process. Next, the chapter shifts focus to visual search as a cognitive process distinct from strategic decision-making in goal-directed behavioural adaptations, emphasising the influence of expectations on search performance. Finally, the chapter identifies key gaps in the literature that this thesis seeks to fill by investigating the joint roles of cognition and communication in adapting behaviour in high-stakes, stressful situations.

1.2 Executive Control in Goal-Directed Behaviour

Goal-directed behaviours, in contrast to habitual or reflexive behaviours, are characterised by actions that are initiated and adjusted in pursuit of desired outcomes or in avoidance of undesirable ones (Balleine & Dickinson, 1998). Search and rescue (SAR) missions undertaken by firefighters appear to be a good example of a goal-directed behaviour. For such behaviours to be effective, individuals must adjust their actions flexibly in response to changing task demands which are often communicated through new information or conflicting cues. In contrast, well-practiced behaviours are triggered by stimulus-response associations stored in long-term memory, making their execution largely automatic and independent of such resources, potentially interfering with controlled processes (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

This capacity for goal-directed behaviours is underpinned by executive functions, which are generally understood to encompass a set of core cognitive abilities which are interrelated but dissociable, namely inhibitory control, working memory, and cognitive flexibility (Diamond, 2013; Miyake et al., 2000). Inhibitory control refers to the ability to suppress attentional capture, mental representations, and dominant responses when they are incongruent with task demands. For example, inhibitory control is required when participants name the colour of the ink a colour word is printed in while ignoring the word itself (Stroop task, MacLeod, 1991; Stroop, 1935), respond to a stimulus feature when its location conflicts with the response side (Simon task, Hommel, 2011; Simon & Rudell, 1967), or identify a central stimulus while suppressing interference from adjacent stimuli (Flanker task, Eriksen & Eriksen, 1974). Working memory refers to the capacity to maintain and manipulate task-relevant information (Baddeley & Hitch, 1974; Engle, 2002), which has been implicated in tasks such as complex span tasks that alternates between processing information (e.g., judging the correctness of arithmetic operations) and storing information (e.g., memorising a letter after each operation to be recalled later) (Miyake et al., 2000; Oswald et al., 2015; Turner & Engle, 1989). Cognitive flexibility refers to the capacity to switch between mental representations or response strategies in response to changing task demands. It is commonly assessed with set-shifting paradigms such as the Wisconsin Card Sorting Task, where individuals infer changes in sorting criterion based on feedback and adjust their strategy accordingly (Grant & Berg, 1948; Miyake et al., 2000).

Together, these components of executive functions, often also referred to as executive control or cognitive control, coordinate to support effective behavioural adaptation (Diamond, 2013): Working memory maintains goal representations and manipulates task-relevant information, inhibitory control suppresses distractions and task-irrelevant responses to maintain focus on the goal, and cognitive flexibility enables strategic adjustments when demands change to ensure that actions align with current goals. The more dynamic or novel a task is, the more these processes are recruited as routine responses are insufficient to meet new demands (Diamond, 2013). However, being a finite resource, substantial evidence suggests that cognitive load constraints the capacity of cognitive control, thus increasing the vulnerability to goal-incongruent behaviours, such as those that are well-rehearsed (Bissett et al., 2023; Engström et al., 2017).

1.2.1 Dual-Process Theory of Cognition

The distinction between automatic and controlled processes is also reflected in the processing of task-relevant information, which could in turn impact behavioural adaptation. The dual-process theory of cognition popularised by Kahneman (2011) describes two qualitatively distinct modes of information processing: System 1 thinking, which is fast, intuitive, heuristic-based, and operates with minimal cognitive resources, whereas System 2 thinking, which is slow, deliberative, rule-based, and resource-intensive[1]. Although both systems are active during processing, System 2 is minimally engaged most of the time, with System 1 outputs informing beliefs or actions. As such, individuals often rely on System 1 due to its efficiency and general accuracy, especially when mental models are well-calibrated (Kahneman, 2011) or judgements can be made based on incomplete but highly informative cues (e.g., take-the-best heuristic, Gigerenzer & Goldstein, 1996). If a conflict is detected between incoming information and internal models or if System 1 fails to provide a response (e.g., during complex or unfamiliar tasks), System 2 is held to intervene for more elaborated processing, provided that the motivation to mobilise the necessary resources is present (J. S. B. T. Evans & Stanovich, 2013; Kahneman, 2011). However, when cognitive resources are limited, the capacity for deliberative processing is compromised, increasing reliance on System 1 responses (De Neys, 2006; Greene et al., 2008).

[1] Recent evidence suggests that logical reasoning is not exclusively the domain of System 2, as System 1 can intuit responses based on basic logical principles. This view suggests that System 1 generates both a traditional "heuristic" and a "logical" intuitive responses, with conflict detection and further deliberative processes engaged only when both types of intuitive responses are activated to similar strength levels (De Neys & Pennycook, 2019).

Thus, in situations that constrains executive resources (such as during an emergency incident), controlled processes are hampered, potentially impairing behavioural adaptation in more ways than one. Where task-related information requires additional interpretation or inference, reduced cognitive resources can limit System 2 processing, increasing the likelihood of misunderstanding critical information. Such misunderstandings have the potential to misguide behavioural adaptation and compromise safety-critical decision. Where task-related information is explicit, deficits in cognitive control might still complicate behavioural adaptation by making it more challenging to override well-trained responses.

1.2.2 Self-Generated Expectations versus Instructed Expectations

So far, the barriers to behavioural adaptation have been discussed in terms of well-rehearsed behaviours or well-learned heuristics (both necessarily formed through extensive experience). However, it seems reasonable to suggest that individuals may also possess prior experience sufficient to form expectations about how an event will unfold, yet insufficient for (near-)automatic behaviours to develop. In such cases, these self-generated expectations, albeit formed in the absence of well-rehearsed behaviours, could still impair adaptive control of goal-directed behaviours.

Evidence that self-generated expectations may introduce barriers to behavioural adaptation comes from studies examining the consequences of expectation violations (Gaschler et al., 2014; Jiang et al., 2018; Kemper et al., 2012, 2017; Oberauer et al., 2013; Schwager et al., 2017; Umbach et al., 2012). To investigate whether the source of expectations affect goal-directed actions, researchers used expectation mismatch effects to compare the impact of violations of self-generated expectations versus violations of instructed expectations (i.e., externally informed expectations). Mismatch effects are typically observed when a stimulus does not match an individual's predictions, leading to slower reaction times (RTs) compared to when the stimulus matches their predictions. For example, in Kemper et al. (2012), participants either predicted the colour of the upcoming stimulus when prompted (i.e., self-generated expectation) or read aloud a colour word presented on screen (i.e., instructed expectation). They then indicated via keypresses whether the subsequently presented stimulus was red or yellow. When stimuli matched expectations, RTs were faster, and this facilitation was greater for self-generated expectations than for instructed expectations. Conversely, when stimuli violated expectations, RTs were slower, and this mismatch effect was larger for self-generated expectations (Kemper et al., 2012).

Subsequent research showed that the size of the mismatch effect is dependent on the validity of self-generated expectations. For instance, when self-generated expectations were met with 80% validity, greater mismatch effects were observed than when they were met at chance level (Schwager et al., 2017). Interestingly, however, mismatch effects were observed even when self-generated expectations were no better than guesses and when instructed expectations had to be processed (Kemper et al., 2017; Umbach et al., 2012). In Kemper et al. (2017), the two types of stimuli (i.e., scatterplots showing either an ascending or descending slope) appeared with equal frequency throughout the experiment, making prior experience an unreliable basis from which to generate expectations about whether the upcoming scatterplots had an ascending or descending slope. Additionally, instead of merely indicating slope directions, participants were asked to indicate via key presses whether the direction matched their expectations, ensuring that instructed expectations were fully processed (Kemper et al., 2017).

Together, self-generated expectations appear more resistant to disconfirmation than instructed expectations, suggesting they may override the latter even when it is critical for guiding goal-directed behaviour (e.g., a SAR mission). This interpretation received empirical support from Jiang et al. (2018), who showed with a task-switching paradigm that switch costs were more strongly modulated by recent trial history (up to the most recent three trials, which informed self-generated expectations) than external cue (i.e., instructed expectation) that always accurately indicated the probability of the upcoming task type. Participants were shown a cluster of 60 coloured dots in motion, and were either asked to categorise the dominant colour (purple or green) or the dominant motion direction (left or right) with keypresses. The findings replicated classic task switch cost, where longer RTs were observed for task-switches (e.g., colour categorisation followed by motion categorisation) than for task-repeats. Crucially, trial history biased the participants to expect task-repeats, leading to shorter RTs when self-generated expectations of a task-repeat was met, and longer RTs when they were violated by a task-switch. Although these self-generated expectations were not predictive of the upcoming task, their influence on categorisation RTs were three times stronger than that of instructed expectations, with five out of 22 participants showing very little or no reliance on the latter. These findings suggested that preparatory processes were guided by self-generated instead of instructed expectations when both conflicted (Jiang et al., 2018).

The stronger influence of self-generated expectations (over instructed ones) is proposed to stem from their stronger representation in working memory, enabling them greater access to attentional focus and enhancing their influence on behaviour (Gaschler et al., 2014). The evidence for this comes from event-related potential components observed during mismatch trials (Kemper et al., 2012): early detection of expectation mismatches and conflict monitoring processes indexed by the N2, and later evaluative processes of whether expectations were confirmed or violated indexed by P3. Both components showed larger amplitudes for violated self-generated expectations than violated instructed expectations. This increased magnitude likely reflected greater attentional engagement and preparatory processing for self-generated expectations, which, when violated, required additional cognitive resources to process the unexpected stimuli (Kemper et al., 2012). Nevertheless, it remains unclear how constraints on cognitive resources impact the maintenance of self-generated expectations in working memory, and how these interact with instructed expectations, especially when the latter conveys critical information in high-stakes contexts.

To sum, in high-stakes contexts (e.g., SAR missions) where high cognitive load limits cognitive control, behavioural adaptation might be vulnerable to automatic processes where controlled processes are required. This susceptibility is not limited to well-rehearsed behaviours and dominant responses but includes robust self-generated expectations that do not require extensive experience to develop. The competing influence between self-generated and instructed expectations is a key theme in this thesis and is addressed in the experiments reported in Chapters 2 and 4, primarily in Experiment 3 (Chapter 2) and Experiment 4 (Chapter 4) where self-generated expectations are informed by more extensive "practice trials" relative to Experiments 1 and 2 (Chapter 2).

The next section describes barriers to behavioural adaptation from the perspective of language use, focusing on the effectiveness of everyday pragmatics in communicating new, critical information (i.e., instructed expectations) that supports adaptive behaviour.

1.3 The Use (and Misuse) of Implicit Communication

Daily communication involves a mixture of explicit and implicit utterances, as well as various forms of ambiguity and error, all of which requires the addressee to interpret the message correctly. Consider the following dialogues (Jang et al., 2013):

(1) A: "Is Dr. Smith in his office now?"

B: "Dr. Smith is in his office now."

(2) A: “Is Dr. Smith in his office now?”

B: “Dr. Smith’s car is parked outside the building.”

Explicit communication is illustrated in (1), where the communicator’s message is directly conveyed through the literal meaning of the words. In contrast, implicit communication is illustrated in (2), where the communicator’s message must be inferred from a response that is literally irrelevant to the question (Grice, 1989; Jang et al., 2013). In pragmatic terms, this indirect reply exemplifies a conversational implicature, which is a form of implicit communication where the addressee infers the intended meaning by assuming cooperation (Grice, 1989) and relevance (Sperber & Wilson, 1986, 1987). Despite taking 300ms longer to interpret, most participants understood the implicit responses correctly when asked to assess whether the response affirmed or denied the question posed: The mean accuracy for interpretations of implicit responses was 95.19%, only slightly (but significantly) lower than the 98.63% mean accuracy observed for interpretations of explicit responses (Jang et al., 2013). This high level of accuracy demonstrates our capacity to infer intended meanings from implicit communication, specifically conversational implicatures, with remarkable ease, a feat enabled by everyday pragmatic reasoning.

1.3.1 Grice’s Cooperative Principle and Conversational Maxims

A comprehensive overview pragmatic theories is beyond the scope of this thesis. Instead, the focus here will be primarily on Gricean principles (Grice, 1989) and Relevance Theory (Sperber & Wilson, 1986, 1987) due to their direct relevance to pragmatic implicatures in simulated SAR contexts.

Grice (1989) originated the foundational framework that explains how underspecified communication is interpreted through pragmatic reasoning, based on the expectation of cooperative discourse. The Cooperative Principle summarises Grice's position that successful communication is a collaborative effort that relies on the communicator adhering to the four conversational maxims: quantity (provide the right amount of information), quality (contribute only accurate information), relation (contribute only relevant information), and manner (be clear and concise). Conversational implicature, as illustrated in (2), arises because the addressee expects that the communicator obeys the Cooperative Principle: They intend to be understood and are thus cooperative and truthful in their contribution. Therefore, when the communicator appears to have made an irrelevant remark (i.e., when they flout one or more of the maxims), the addressee exploits shared context to make an alternative interpretation of the remark that goes beyond its literal meaning to preserve the collaborative nature of the exchange. In formal terms, this inferencing process can be characterised, from the perspective of the addressee, in the following steps (Grice, 1989):

(i) The communicator has said that *p*.

(ii) There is no reason to think that the communicator is not being cooperative during the exchange.

(iii) The communicator would not have said that *p* unless they thought that *q*, and they know that I know that *q* is required to understand that *p*.

(iv) The communicator has not stopped me from thinking that *q*.

(v) The communicator has implicated that *q*.

1.3.2 Relevance Theory

Sperber and Wilson's (1986, 1987) Relevance Theory offers a cognitive framework that extends Grice's socio-cooperative understanding of pragmatic implicatures, focusing on the cognitive processes that are biased toward maximising relevance through the efficient use of processing resources. In this theory, relevance is defined as the size of positive cognitive effects (i.e., adjustments to understanding through confirmation, revision, or rejection of existing assumptions) relative to the processing effort (i.e., cognitive resources) required to interpret the information. Relevance Theory operates on the assumption that our cognitive system has evolved to automatically recognise stimuli that are likely to yield the greatest relevance for the least processing effort. Consequently, to ensure that the most relevant interpretation is achieved with the least effort, implicatures are computed only when background assumptions that support their interpretation are readily accessible in the communicative context. Relatedly, the theory also posits that the addressee expects any ostensive behaviours (i.e., behaviours that indicate the communicator's intent) to be optimally relevant, prompting them to seek the most cognitively efficient interpretation by integrating the most salient background assumptions and environmental cues (Sperber & Wilson, 1986, 1987).

1.3.3 The Effortful Processing of Conversational Implicatures

The fact that the derivation of implicatures includes an (often unconscious) consideration of processing effort trade-offs suggests that effortful cognitive processes are involved (Sperber & Wilson, 1986, 1987). Nevertheless, given that (conversational) implicatures must be inferred, the resulting inference can sometimes be incorrect. That such inferences are usually accurate or at least sufficiently reliable for effective communication suggests that they are generated with relative ease in daily interactions. Moeschler (2023) addresses this apparent paradox between the effortful nature of pragmatic processing and its seamless integration into everyday communication by borrowing from dual-process theories (Kahneman, 2011). He proposed that System 1 processes syntax, semantics, and pragmatics which govern initial comprehension and inference. These initial interpretations are biased by existing beliefs, norms, and heuristics, and are typically accepted by System 2. However, when uncertainty or conflict arises, such as when the initial inference conflicts with the assumed common ground, System 2 intervenes to evaluate the truth value of the initial inference and revise it if necessary.

As the reliability of intuitive inferences depends critically on the engagement of System 2, any limitation in the availability of cognitive resources might risk the uncritical acceptance of such initial inferences (De Neys, 2006; Greene et al., 2008). These System 2 processes are supported by the different components of executive functions. Indeed, these components have been found to directly support pragmatic reasoning, specifically the understanding of implicatures in a sample of healthy older adults (Bambini et al., 2021). Cognitive flexibility enables one to negotiate between the literal meaning of the utterance and alternate interpretations generated from perspective-taking. These elements, along with other sources of information (e.g., prior knowledge, contextual information), are maintained and integrated in working memory to construct an interpretation that is maximally relevant to the present context (Bambini et al., 2021). While Bambini et al.'s findings do not speak directly to the correction of misleading inferences intuited by System 1, they nevertheless suggest that reduced executive functioning might limit the ability to integrate goal-directed, contextual cues while interpreting communitive intent.

1.3.4 Conversational Implicatures Under Cognitive Constraints

Evidence from theoretical and experimental work has demonstrated that cognitive constraints pose significant challenges to interpreting implicature (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Grice, 1989; Nys et al., 2024). A well-studied example of this is scalar implicature, which occurs when communicators soften assertions by using weaker terms on a scale to imply that the stronger statement does not hold. Examples include saying "some" to imply "some but not all" (where "all" is the stronger statement), instead of its strict logical meaning of "at least one". From a Gricean perspective, scalar implicatures arise from the assumption of cooperative communication, which is to say that the addressee presumes adherence to conversational maxims. For instance, the maxim of quantity requires the communicator to provide the right amount of information, but no more and no less. When a communicator says, “some of my dogs are overweight”, the logical interpretation of this statement is consistent with “all of my dogs are overweight”. A cooperative communicator would have said the latter if that were true, but by using the weaker term “some” instead of the stronger term “all”, they provide less information than they could have. Unless there is a reason to believe that the communicator is being uncooperative by flouting the maxim of quality, the addressee concludes that the communicator did not say “all of my dogs are overweight” because the communicator did not believe it to be true. Consequently, this leads the addressee to infer that “some” means “some but not all” in this context. In a similar vein, Relevance Theory offers a complementary account of scalar implicatures. The addressee presumes the communicative act to be optimally relevant and engages processing effort in proportion to the positive cognitive effects expected. In this case, the meaning of “some” is pragmatically enriched beyond its logical interpretation to “some but not all” if the latter interpretation is more contextually relevant than the logical one.

There is empirical evidence for this inferential process being cognitively costly. An uninformative sentence like “some tuna are fish” is normally understood to mean “some but not all tuna are fish” (i.e., there are other types of tuna that are not fish), based on our expectations of language use in daily conversation and the addressee’s assumption about communicative intent (Bott & Noveck, 2004; De Neys & Schaeken, 2007). In other words, the sentence was interpreted pragmatically, rather than logically, as the logical interpretation does not carry any additional meaning beyond its literal sense (De Neys & Schaeken, 2007). Interpreting uninformative sentences pragmatically (e.g., judging “some tuna are fish” as false) takes significantly longer than interpreting them logically (e.g., judging the same sentence as true), suggesting that pragmatic interpretations require greater cognitive effort (Bott & Noveck, 2004). Under cognitive constraints such as visuospatial memory load (De Neys & Schaeken, 2007; van Tiel, Pankratz, et al., 2019), participants made significantly fewer pragmatic than logical interpretations of such sentences containing the words “some”, “or”, “might”, and “most”, often judging them as false statements. They also took longer to interpret these sentences pragmatically than logically under high versus low cognitive load (De Neys & Schaeken, 2007; van Tiel, Pankratz, et al., 2019).

While there seems to be broad agreement that pragmatic processing is cognitively costly, some findings challenge that assumption. For instance, reducing processing time by means of increasing time pressure was found to reduce pragmatic interpretations of scalar implicatures (Bott & Noveck, 2004), or not at all (van Tiel, Marty, et al., 2019). This discrepancy could potentially be attributed to how processing time was operationalised in both studies. In Bott and Noveck’s Experiment 4, participants responded to sentences alone, with strict response deadlines of either 900ms or 3000ms imposed, which directly limited the time available for processing and decision-making. In contrast, van Tiel, Marty, et al.’s (2019) Experiment 2 used a sentence-picture verification task where participants judged whether a sentence (e.g., “most of the apples are green”) correctly described an image (e.g., a collection of only green apples). Processing time was mainly manipulated by limiting exposure to the image. The image always appeared first, followed by the sentence, and in the “fast” condition the image was only visible for a second, but participants had as much time as they needed to process the sentence and respond. In the “normal” and “slow conditions the image was visible until participants responded, and in the “slow” condition, an additional 3-second delay was enforced before they could respond. Thus, although the brief display window in the “fast” condition might have increased working memory demands (as participants had to retain the image while processing the sentence), it did not impose the same kind of pressure to respond.

This might explain why pragmatic reasoning was more affected in Bott and Noveck's study, where time pressure directly constrained both inferencing and decision-making processes.

1.3.5 Pragmatic Cues Encourage Pragmatic Processing

Evidence still suggests that cues that encourage pragmatic processing, such as strategically positioned emphasis, might counteract the deleterious effects of cognitive resource impairment. Chevallier et al. (2008) presented participants with single words (e.g., "TABLE") and instructed them to judge whether corresponding or-statements (e.g., "There is an A or a B") were true based on the letters in the word. When the scalar word "or" was stressed, either graphically ("<u>OR</u>") or prosodically, participants made more pragmatic interpretations (i.e., one but not both) than logical interpretations (i.e., at least one) of the or-statements, compared to when it was unstressed (Chevallier et al., 2008). Although neither exposure nor response windows were restricted, these findings suggest that certain cues, such as the non-lexical cues described here, could potentially help maintain pragmatic reasoning when cognitive resources are otherwise challenged.

In contrast with previous work (Bott & Noveck, 2004; De Neys & Schaeken, 2007; van Tiel, Marty, et al., 2019), a recent study found that neither a concurrent visuomotor tracking task nor a reading span task prevented participants from interpreting atypical implicatures, which arise in informationally-redundant utterances concerning common knowledge about typical event sequences[2] (Ryzhova & Demberg, 2023) . For example, in response to the vignette "Today, Lisa went to the swimming pool […] She brought her swimming suit!",

[2] The conversational implicatures in Ryzhova and Demberg (2023) can be considered Particularised Conversational Implicatures (PCI), which are context-dependent implicatures some consider more effortful to compute than Generalised Conversational Implicatures (GCI). GCIs arise automatically (i.e., by default) and are context-independent, such as scalar implicatures (Default Inference Account; Levinson, 2000). However, evidence suggesting that GCIs are neither effortlessly nor automatically generated (e.g., Bott & Noveck, 2004; Breheny et al., 2006) challenges the validity of the distinction between PCIs and GCIs.

participants under high and low load were equally able to infer that Lisa typically did not bring her swimming suit to the swimming pool (Ryzhova & Demberg, 2023, p. 48). While the authors did not explicitly state this, it is plausible that the exclamatory intonation of the redundant information (e.g., "She brought her swimming suit!") in Experiment 1, where participants listened to the vignettes, might have served as a pragmatic cue. This likely drew attention to the redundant information and encouraged pragmatic reasoning even under high load. Given the similar null findings in Experiments 2 and 3 where participants read the vignettes, it seems plausible to suggest that the exclamation mark alone might have sufficed to serve as a pragmatic cue.

Overall, the evidence suggests that cognitive constraints hinder pragmatic processing, potentially leaving the interpretation of implicatures vulnerable to bias due to insufficient System 2 intervention. This interpretation is closely linked to limitations in executive functioning that support the deliberative processes characteristic of System 2. However, certain pragmatic cues could overcome this bottleneck. This raises a critical question: which types of cues could serve this function, especially in stressful, high-stakes situations where goal-directed behaviour is required but cognitive control is taxed? The difference in pragmatic context between laboratory experiments and naturalistic environments calls into question whether the richness of the situational context alone, which is full of environmental cues and information, can sufficiently support pragmatic reasoning in conditions where cognitive resources are limited. Experiments 1-3 in Chapter 2 investigated participants' ability to process conversational implicatures in such a context, where information critical to the success of simulated SAR missions was conveyed either explicitly or as conversational implicatures.

While previous research manipulated cognitive load using concurrent tasks (Bott & Noveck, 2004; De Neys & Schaeken, 2007; Ryzhova & Demberg, 2023; van Tiel, Marty, et al., 2019), elevated stress as experienced in emergencies is also likely to contribute to these constraints in naturalistic high-stakes settings. The next section explores this possibility in the context of the effects of stress on executive functioning.

1.4 Executive Functioning Under Stress

In emergencies or other high-stakes operational contexts, stress is typically experienced as an acute response to the intense conditions that individuals are subjected to. A comprehensive review of the effects of stress on performance and overall well-being is provided by Staal et al. (2004). One widely used framework for understanding the experience of stress is the Transactional Model, which describes the source of stress as stemming not necessarily from the intensity of external conditions, but the individual's primary and secondary appraisals (Lazarus & Folkman, 1987). Primary appraisal evaluates the relevance of what is happening to our goals, with obstacles that hinder goal achievement appraised as harmful, threatening, or challenging. In other words, primary appraisal reflects the individual's judgement about the impact of what is happening in the context of goal achievement (e.g., whether low visibility makes it more challenging to rescue a trapped person in the building). Secondary appraisal evaluates the individual's ability to cope with the demands of the situation, considering the available resources that could be used to manage them. This appraisal informs the individual's perceived control over the situation and outcomes. Thus, stress arises when the individual perceives that the demands of the situation exceed their perceived ability to cope (Lazarus & Folkman, 1987).

Although the experience of stress as defined above is based on subjective appraisal, its cognitive effects on different aspects of executive functions are well documented (Shields et al., 2016). High levels of stress are associated with deficits in executive functioning, particularly in working memory and cognitive flexibility. However, its effects on inhibitory control are more nuanced: While stress can enhance behavioural performance requiring the suppression of prepotent responses (i.e., response inhibition), it usually impairs the ability to ignore task-irrelevant information (i.e., cognitive inhibition or interference control) (Shields et al., 2016). These impairments are commonly thought to arise from the reallocation of cognitive resources to process stress-related information, which reduces the availability of resources that typically support such processes (LeBlanc, 2009; Plessow et al., 2011). Thus, while stress may facilitate the inhibition of goal-irrelevant responses, it likely undermines behavioural adaptation when task-relevant information needs further interpretation or conflicts with self-generated expectations (based on prior experience), by biasing cognitive processing towards reactivity and automaticity.

Disruptions to executive functioning under stress also undermine the higher-order cognitive processes that depend on them. In the domain of decision-making, the Stress-Induced Deliberation-to-Intuition (SIDI) framework, developed by Yu (2016) from a synthesis of existing findings, complements Shields et al.'s (2016) review of the effects of stress on executive function tasks. Such tasks include: complex span task for working memory, go/no-go task for response inhibition, flanker task for cognitive inhibition, and Wisconsin card sorting test for cognitive flexibility. According to SIDI, impairments in executive functioning can lead to suboptimal decisions in tasks that require careful evaluation of options under uncertainty (e.g., Iowa gambling task, balloon analogue risk task). Under high stress, individuals default to intuitive thinking (System 1) instead and bypass the reasoning system (System 2) that would normally evaluate whether an intuitive suggestion aligns with activated goals and the immediate environment (Yu, 2016). In the context of SAR missions, there is evidence that incident commanders are more likely to rely on standard operating procedures (i.e., rules; presumably involving System 1) than to use operational discretion (i.e., reasoning; presumably involving System 2; Butler et al., 2023).

There is, however, some evidence showing that acute stress might benefit certain aspects of executive functioning under certain conditions. Goldfarb et al. (2017) found that working memory updating, a component of cognitive flexibility, was enhanced after stress from a cold pressor task. Participants completed a delayed match-to-sample (DMS) task under both stress and control conditions. Each trial involved: (1) encoding two coloured figures, (2) either retaining them (No Interference trial), updating working memory with two new coloured figures presented (Update trial), or ignoring the new figures presented (Ignore trial), and (3) identifying whether a single image matched one of the figures currently held in working memory. While stress impaired task switching (e.g., reduced accuracy when switching between Update and Ignore trials), those with a larger cortisol response performed more accurately in Update trials than in No Interference trials compared to those with a smaller cortisol response. The impairment in task switching was an expected consequence of stress-induced prefrontal cortex disruption, but Goldfarb et al. (2017) posited that the unexpected updating flexibility might be supported by enhanced striatal processes under stress.

It is also possible that the type of stress manipulation might also contribute to why there was no deterioration in cognitive flexibility in Goldfarb et al. (2017). They noted that the use of Trier Social Stress Test (TSST) in previous studies (Alexander et al., 2007; Plessow et al., 2011) encourages post-task rumination and elevated cortisol levels due to explicit social evaluative threat (Zoccola et al., 2008). This suggests that the cold pressor task, which lacks a social evaluative component, likely did not provoke post-task rumination, potentially resulting in less disruption in cognitive flexibility relative to the TSST. More broadly, this interpretation aligns with literature showing that rumination imposes cognitive load, depleting working memory resources needed to support goal-directed behaviours (Bruning et al., 2023; Curci et al., 2013).

Taken together, it appears that stress impairs executive functioning and reasoning in ways that resemble the effects of high cognitive load observed in dual-task paradigms (Bissett et al., 2023; Diamond, 2013; Engström et al., 2017). Given that both the ability to interpret conversational implicatures correctly and to adapt behaviour in pursuit of goals also relies on System 2 processes that depend on the availability of executive resources (see Sections 1.2 and 1.3.3), it seems plausible to argue that stress also constrains pragmatic reasoning and behavioural adaptation. In high-stakes contexts which are often stressful, the integrity of these processes has important implications for goal-directed behaviour: They directly influence both the interpretation of critical information and the flexible adaptation of behaviour, both of which may rely on information that might not be always be communicated explicitly (Kurinec et al., 2019).

1.4.1 Anticipatory Stress

In high-stakes professions such as firefighting, new information is ideally communicated in conditions approximating a "sterile cockpit", a term borrowed from aviation. It describes an environment where distractions and interruptions are minimised to ensure undivided attention, so that information exchange, processing, and decision-making can be optimised in preparation for demanding tasks (Sumwalt, 1993). However, such conditions are not always possible; critical information could sometimes be delivered mid-operation when external distractions compete for the resources required for careful reasoning. Crucially, even when a relative calm window is secured to facilitate this process, the anticipation of upcoming stressors might still disrupt the cognitive functions needed to interpret and act on new but task-critical information.

As discussed earlier, sources of stress need not be temporally immediate to elicit a stress response (Alexander et al., 2007; Plessow et al., 2011; Zoccola et al., 2008). In contrast to post-test rumination, the anticipation of future demands can also be a potent trigger to a stress response. In fact, anticipatory stress is often incorporated into standardised stress elicitation protocols, such as the TSST (Kirschbaum et al., 1993). In the classic TSST protocol, participants play the role of a job applicant and are given 10 minutes to prepare a 5-minute speech to convince a hiring committee of their suitability for an advertised vacancy. They are also informed that their performance will be recorded and that the committee specialises in monitoring nonverbal behaviour (Kirschbaum et al., 1993). This preparatory interval elicited anticipatory stress responses relative to baseline, showing modest elevations in physiological measures of stress such as cortisol and heart rate (Engert et al., 2013; Kelly et al., 2007; Kirschbaum et al., 1993; Starcke et al., 2008)[3], although subjective reports of stress did not increase relative to baseline during this period (Engert et al., 2013). Similarly, elevated cortisol levels were observed in novice firefighters prior to attending a fire extinguisher demonstration, receiving training on the self-contained breathing apparatus (SCBA), or completing a search and rescue exercise (Robinson et al., 2013).

Anticipatory stress has been shown to negatively impact different cognitive processes, such as working memory (Hyun et al., 2019), decision-making (Starcke et al., 2008), and attention control (Cain et al., 2011). In the study conducted by Hyun et al. (2019), each morning participants rated their anticipated stress for the day, and performed later in the day a spatial dot memory task. Poorer working memory was associated with higher levels of stress anticipation in the morning, even after controlling for stressors actually experienced (Hyun et al., 2019). Starcke et al. (2008) induced anticipatory stress by subjecting participants to only the preparatory phase of a modified TSST after completing several neuropsychological tests. Those who received the stress manipulation made poorer decision on the Game of Dice Task

[3] Interestingly, meta-analytical evidence showed that varying the length of the preparatory interval did not increase the overall stress response elicited by the TSST, suggesting that beyond a certain point, longer anticipation might not intensify stress further (Zimmer et al., 2019).

(GDT), which required choosing between betting options with explicitly stated probabilities of rewards and losses, than those who did not. Likewise, under anticipation of electrical shocks, participants in Cain et al.'s (2011) study were more likely to miss less-salient targets after salient ones in dual-target search tasks. However, their performance in single-target searches was unaffected, suggesting that anticipatory stress narrowed attention to salient information at the expense of less salient ones (Cain et al., 2011).

These findings reinforce the view that impairments in cognitive processes can arise not only from direct experience of actual stressors but from appraisals of future stressors (Hyun et al., 2019; Lazarus & Folkman, 1987). Brosschot et al. (2006) proposed a mechanism through which this occurs: perseverative cognitions, such as ruminations and worries, sustain focus on the anticipated stressors and triggers stress-related physiological reactions (e.g., elevated cortisol) even without immediate stressors. Recent evidence shows that anticipatory stress was associated with more perseverative cognitions up to three hours later (Kramer et al., 2021). In addition to increases in negative affect (Brosschot et al., 2006), perseverative cognitions are also linked to reduced working memory capacity (Hayes et al., 2008) and reduced cognitive flexibility (Ottaviani et al., 2013), potentially due in part to the neurocognitive effects of elevated cortisol levels (Shields et al., 2015, 2016).

Overall, the extant literature converges on the view that stress has predominantly negative impacts on executive function, whether triggered by appraisals of immediate or anticipated demands. As disruptions to executive functioning might arise even during low-distraction periods, reserved for information uptake and decision-making, this highlights a critical challenge in high-stakes settings where communication might sometimes be unintentionally implicit and requires careful interpretation. Nevertheless, to optimise the conditions for pragmatic reasoning and adaptive behaviour to unfold, all experiments in this thesis (Chapters 2 and 4) delivered this information during a period of relative calm preceding the virtual SAR tasks despite the potential effects of anticipatory stress, rather than during the SAR tasks, where environmental interference and operational demands escalate substantially. Experiment 4 (Chapter 4) also contained a direct manipulation of stress to investigate its impact on the uptake of new information about target prevalence in a high-stakes visual search task to support search performance.

1.5 Behavioural Adaptation in Visual Search

Beyond tasks that require a higher degree of planning and strategizing, certain activities in high-stakes operations rely more heavily on perceptual forms of decision-making. For example, deciding where to search for hazardous items (e.g., explosive gas cylinders) depends on interpreting relevant information to inform a new search strategy (e.g., knowing where to search). However, knowing where to search might not aid hazard identification in low-visibility conditions (e.g., smoke-filled room) particularly when the hazardous items are surrounded by visually similar but harmless items. In such cases, successful identification hinges on perceptual decision-making guided by expectations about target prevalence. Whether new information about target prevalence, even when delivered explicitly, can successfully override previously formed expectations of target prevalence during subsequent visual search under operational demands warrants investigation. This issue was examined in Chapter 4.

1.5.1 Feature Integration Theory and the Guided Search Model

The traditional understanding of visual search draws a broad distinction between feature search, which describes searching for a target defined by a unique feature (e.g., a blue circle among red circles), and conjunction search, which describes searching for a target defined by a combination of features (e.g., blue circle among red circles and blue squares) (Treisman &

Gelade, 1980). According to Feature Integration Theory (Treisman & Gelade, 1980), feature search is usually characterised by rapid, parallel processing of basic visual features across the visual field, allowing salient features to "pop out" automatically. In contrast, conjunction search is more effortful, as it requires focal attention to be directed serially to different spatial locations so that multiple features of a stimulus can be integrated and compared to a target representation. Later research softened this dichotomy, suggesting that search efficiency is influenced by factors such as distractor heterogeneity, target-distractor similarity, the distinctiveness with which the visual system represents a feature, and illusory conjunctions (where features from different items were incorrectly combined) (Duncan & Humphreys, 1989; Treisman, 1991). These findings laid the groundwork for the development of Wolfe's Guided Search model which conceptualised visual search as an interaction between bottom-up and top-down processes (Wolfe, 2021; Wolfe et al., 1989).

The initial Guided Search model (Wolfe et al., 1989) proposed that both feature and conjunction searches involve similar mechanisms: Pre-attentive, parallel processes created feature maps, and the level of activation within these maps guided attention to the locations of potential (conjunction) targets in the search array. This foundational idea formed the basis for subsequent revisions to the model, with Guided Search (GS) 6.0 (Wolfe, 2021) being the latest and most substantial iteration of the model. Among the changes introduced in GS 6.0, several are particularly relevant to understanding visual search in stressful and dynamic naturalistic conditions. GS 6.0 introduced a continuously evolving spatial priority map of the visual scene that directs attention during search based not just on bottom-up salience and top-down guidance, but also on prior selection history (e.g., priming), rewarded features, and scene structure and semantics. Candidate targets selected based on guiding templates (for feature comparisons) in working memory are then compared against target templates in activated long-term memory, influencing the evidence accumulation process (this process is discussed further in Section 4.2) (Wolfe, 2021).

However, in search contexts where the signal-to-noise ratio is low, such as during a fire incident where visibility is obscured by thick smoke or when hazardous targets closely resemble their surroundings, GS 6.0 (Wolfe, 2021) suggests that parallel processes are limited in their ability to guide focal attention. This limitation can potentially lead to less efficient searches that are more prone to errors and biases, including those arising from outdated expectations of target prevalence. Related literature on prevalence effects and the acquisition of prevalence expectations is discussed in more detail in Chapter 4, where the influence of self-generated and instructed expectations of target prevalence during high-stakes visual search was investigated.

1.5.2 Subsequent Search Misses

Where there are more than one target in a search, detecting the first target reduces the likelihood of detecting another one. The term "satisfaction of search" was first used to describe this phenomenon (Berbaum et al., 1991; Tuddenham, 1962), but was more recently replaced by "subsequent search misses" (SSMs) to avoid the implication that the phenomenon was driven solely by satisfaction or premature termination (Adamo et al., 2013) . In addition to the satisfaction theory (Tuddenham, 1962), current theories explaining SSM include the perceptual set theory (Berbaum et al., 1991) and the resource depletion theory (Berbaum et al., 1991; Cain & Mitroff, 2013; for a review, see also Adamo et al., 2021). Briefly, according to the perceptual set theory, attentional focus is biased towards the features of the initial target detected, leading to reduced detection of subsequent targets that are perceptually dissimilar to the initial target (Gorbunova, 2017; Mitroff et al., 2015). The resource depletion view instead suggests that detecting the first target consumes attentional and working memory resources needed to detect subsequent targets, leading to SSMs (Cain & Mitroff, 2013; Stothart et al., 2018).

However, Cheng and Rich (2018) found that SSMs were reduced when dual-target trials were more common compared to single-target trials, than when dual-target trials were rare. The three dominant theories of SSMs failed to account for these findings (see Cheng & Rich, 2018), but they align with broader literature on prevalence-driven expectations in visual search (Cox et al., 2021; Rich et al., 2008), suggesting that prevalence expectations likely played a role in directing attentional focus and target detection in both initial and subsequent search behaviour.

1.6 Summary and Present Objectives

To summarise, high levels of stress might introduce cognitive constraints that hamper most aspects of executive functioning. In high-stakes contexts such as firefighting, the ability to interpret incoming communication correctly and incorporate them into goal-directed behaviour is paramount to the successful execution of time-critical tasks, which is not only critical for personnel safety, but also for the successful rescue of victims and the minimisation of property damage or loss. A key factor in this dynamic is the influence of prior experience or knowledge, which could be a double-edged sword. While such experience is invaluable in generating expectations quickly in familiar scenarios (which enables rapid decision making), it has been shown to bias information processing and behaviour when new but conflicting information is introduced. The extent to which individuals can override such self-generated expectations and adapt to new, instructed information under stress remains unclear, despite its relevance in high-stakes environments. Moreover, this process might itself interact with the manner in which new information is communicated. Under stress, communicators might unintentionally use implicit language to convey new information (Kurinec et al., 2019), and the addressees might misinterpret these conversational implicatures if cognitive resources required for pragmatic reasoning are constrained under stress. While pragmatic cues such as prosodic emphasis have been shown to support pragmatic reasoning under load (e.g., Ryzhova & Demberg, 2023), it is unclear whether the situational context of a fire emergency alone can itself act as a pragmatic cue for this purpose. Crucially, little is known about how stress, self-generated expectations, and the linguistic framing of new information jointly influence behavioural adaptation in naturalistic high-stakes contexts.

This thesis aimed to investigate the conditions under which individuals effectively integrate new information into goal-directed behaviour in high-stakes and frequently high-stress contexts. To address this objective, I examined four research questions in four experiments, using undergraduate students acting as firefighters in simulated search and rescue (SAR) scenarios:

- Experiment 1: How do individuals respond to potentially inaccurate new information involving goal-directed behaviour in such contexts?
- Experiment 2: How does the framing of uncertain information involving goal-directed behaviour influence its uptake in such contexts?
- Experiment 3: How does the explicit versus implicit framing of new information that contradicts prior experience influence its integration into goal-directed behaviour in such contexts?
- Experiment 4: How does the provision of explicit probabilistic information that contradicts prior experience influence its integration into goal-directed visual search in such contexts?

Chapter 2 reports Experiments 1-3, where I developed and implemented SAR missions in desktop simulations. Chapter 3 describes the development and piloting of the methodology used in Experiment 4. In Experiment 4 I implemented SAR missions using a semi-immersive virtual reality platform. Lastly, Chapter 5 provides a discussion of the findings, synthesising insights from Chapters 2-4 to address the overarching research question: How does new information interact with existing knowledge to generate adaptive behaviour in high-stakes and high-stress conditions. Chapter 5 also discusses the theoretical and practical implications of the research and identifies directions for future research.

Chapter 2:

How (Not) to Frame Information to Guide Search and Rescue Missions: Insights from Computer-Based Simulated Environments

2.1 Abstract

Pragmatic implicatures are interpreted with relative ease in daily communication, but their use in emergency situations has not been investigated. Here, we report how implicatures affect search and rescue (SAR) behaviour in computer-simulated buildings, where participants assumed the role of firefighters engaged in SAR missions. Before the SAR missions in Experiments 1 and 2, they received a floorplan the accuracy of which was implied to be uncertain. Participants search behaviour reflected the information in the floorplan, unless it was explicitly presented as potentially outdated. Prior to the SAR mission in Experiment 3, they received information about the likely whereabouts of victims, which was either conveyed implicitly or explicitly. Explicit (but not implicit) information affected their search behaviour. These results suggest that pragmatic processing of implicatures was not evident in these SAR missions, which has clear implications for how information should be framed in high reliability industries.

2.2 Introduction

Understanding and acting upon information presented verbally or in written form is not only influenced by its literal content and the logical implications it entails, but also by pragmatic implicatures suggested by the context in which the information is presented. The process of understanding and acting upon implicatures seems to be automatic or reflexive, based on heuristics derived from prior experience or assumptions about the nature of communication (e.g. Grice, 1989; Sperber & Wilson, 1986, 1987). However, this is not always the case (see below), and the constraints that determine whether or not implicatures are acted upon are important: How information is delivered in high-stakes scenarios (e.g., in search and rescue missions) needs to be informed by the likely use of conversational implicatures. Whether or not these implicatures are interpreted as intended (i.e., pragmatically) under such conditions is unknown, but there is evidence that is relevant to this issue.

High-stakes scenarios are often stressful, and acute stress can impair aspects of executive function and cognitive control (e.g., cognitive flexibility and working memory) beyond the initial experience of stress (Geißler et al., 2023; Shields et al., 2016). There is evidence that these cognitive functions directly support pragmatic processing in communication (Bambini et al., 2021; Nys et al., 2024; see also, De Neys & Schaeken, 2007; Sperber & Wilson, 1986, 1987). For example, Bambini et al. (2021) found that working memory and cognitive flexibility accounted for significant percentages of the variance in understanding pragmatic implicature, as measured by the comprehension section of the Assessment of Pragmatic Abilities and Cognitive Substrates test (40%; Arcara & Bambini, 2016) and an adaptation of the Implicatures Test (49%; Janssens & Schaeken, 2013).

Experiments that taxed these functions prior to assessing pragmatic implicatures provided direct evidence of their involvement in interpreting these implicatures. van Tiel et al

(2019) first gave participants either a complex visuospatial memory task or a simpler task, and then paired sentences containing scalar words (e.g., “The arrow might land on red.”) with an image (e.g., a spinning wheel): one was unambiguously true (i.e., one segment was red), one unambiguously false (i.e., no segment was red), and a target image whose truth value depended on whether the sentence was interpreted pragmatically (i.e., all segments were red). If participants computed the scalar implicature, then they would judge the sentence as an inaccurate description of the target image, as they have moved beyond interpreting the scalar word based on its logical meaning but instead interpreted it based on expectations of everyday language use and cooperative communication (Grice, 1989). Under the assumption of communicative cooperation, the communicator’s use of a weaker term “might” when they could have used the stronger term “will” led the listener to infer that the communicator knew the stronger alternative to be false. Thus, the sentence incorrectly describes a spinning wheel with only red segments. Similarly, under Relevance Theory (Sperber & Wilson, 1986, 1987), the addressee would interpret “might” in a maximally relevant manner: The communicator chose “might” rather than “will” to signal uncertainty, so the sentence does not accurately describe a spinning wheel that is fully red. Participants who first completed a complex visuospatial memory task provided significantly fewer pragmatic responses than those who first completed a simpler task (Bott & Noveck, 2004; De Neys & Schaeken, 2007; van Tiel, Marty, et al., 2019, Experiment 1; but see, van Tiel, Marty, et al., 2019, Experiment 2; Ryzhova & Demberg, 2023).

There is then evidence to suggest that understanding implicatures might be challenging in emergency scenarios (e.g., in a search and rescue mission, SAR), which could affect the safety of firefighters and the public. For example, when provided with a floorplan described as “might be useful” for search and rescue in a burning building, the term “might” not only signals weak evidentiality of the utility of the floorplan but also triggers a scalar implicature indicating that the floorplan should not be relied on unreservedly. This inference arises from reasoning about the communicator's choice to use weaker term "might" versus stronger term "will" (Grice, 1989) to describe the floorplan in this context, while also reflecting a maximally relevant interpretation (Sperber & Wilson, 1986, 1987). The present study used computer-based simulations of SAR environments in which undergraduate students assumed the role of a firefighter. Virtual reality (VR) and serious games (i.e., games designed primarily for training and education) are widely used in firefighting research and training, offering a safe, controlled environment for systematic manipulations while enhancing immersion and presence (Doroudian et al., 2022; Shi et al., 2021; Williams-Bell et al., 2015). For instance, VR has been used to investigate wayfinding aids (e.g., maps, directions) in low visibility conditions, showing that route and survey information were more helpful than landmarks for navigation (Shi et al., 2021). Similarly, Douroudian et al. (2022) found that dynamic floorplans, displaying live updates of fire locations and trapped individuals, improved rescue performance relative to static floorplans in a VR-based SAR operation. However, to our knowledge, pragmatic reasoning has not been studied in such SAR contexts. While computer-based simulations are less immersive than VR headsets, they are less likely to induce cybersickness (Saredakis et al., 2020; Srivastava et al., 2019) due to lower sensory conflict (Palmisano et al., 2020). In our simulations, participants received implicit (or explicit) information that was potentially relevant to their SAR missions, and the levels of self-reported stress during the SAR missions were assessed.

2.3 Experiment 1

Experiment 1 aimed to assess the impact of the provision of potentially useful spatial information about the to-be-searched building on search behaviour. There were two main stages: Familiarisation with a virtual office building followed by a SAR mission in the same building. During the Familiarisation stage, participants explored the well-lit building while moving objects from one place to another and then received a test designed to assess their knowledge of the layout of the building. During the SAR mission, participants played the role of a firefighter entering a smoke-filled and burning office building with poor visibility and additional stressors including a constant alarm and a depleting air source. Their primary task was to search for and rescue a person (the Assistant Finance Manager) who was trapped in the building and their secondary task was to remove a potentially explosive item (a portable stove).

After the Familiarisation stage and before the SAR mission, participants received a floorplan that they were informed "might be useful" in completing the SAR mission. Figure 2.1 depicts floorplans of the two virtual environments that were used (A and B) and screenshots of the entrance to the building during Familiarisation (C) and the SAR mission (D). For half of the participants, this floorplan was an accurate representation of the spatial distribution of rooms within the building (e.g., Floorplan A) and for the remainder it was an inaccurate representation of these rooms (e.g., Floorplan B; with the Finance Department and the Service point transposed with the Marketing Department and Software Hub, respectively). The accurate and inaccurate floorplans were presented to the participants either upright or rotated 180° on the computer screen and with respect to the virtual building entrance. These two manipulations (accuracy and orientation) mimicked scenarios in which firefighters accessed information (e.g., floorplans or blueprints) that might reflect the current or previous interior of a building (accurate or inaccurate), which is presented to them (or discussed with colleagues) oriented with respect to the building (upright) or not (rotated).

The impact of floorplan provision in the four groups (i.e., Accurate-upright, Accurate-rotated, Inaccurate-upright and Inaccurate-rotated) was assessed using two principal measures: Whether once inside the building they turned in the direction of the Finance Department indicated by the floorplan, and the number of redundant paths taken before they exited the building (e.g., retracing their steps). For example, a greater number of redundant paths in groups given inaccurate floorplans relative to groups given accurate floorplans would suggest an ongoing interaction between the floorplan and exploration within the virtual building. However, if the provision of a floorplan had no impact and the participants retained and used information about the virtual environment experienced during Familiarisation, then they should turn in the correct direction and should have little grounds to retrace their steps.

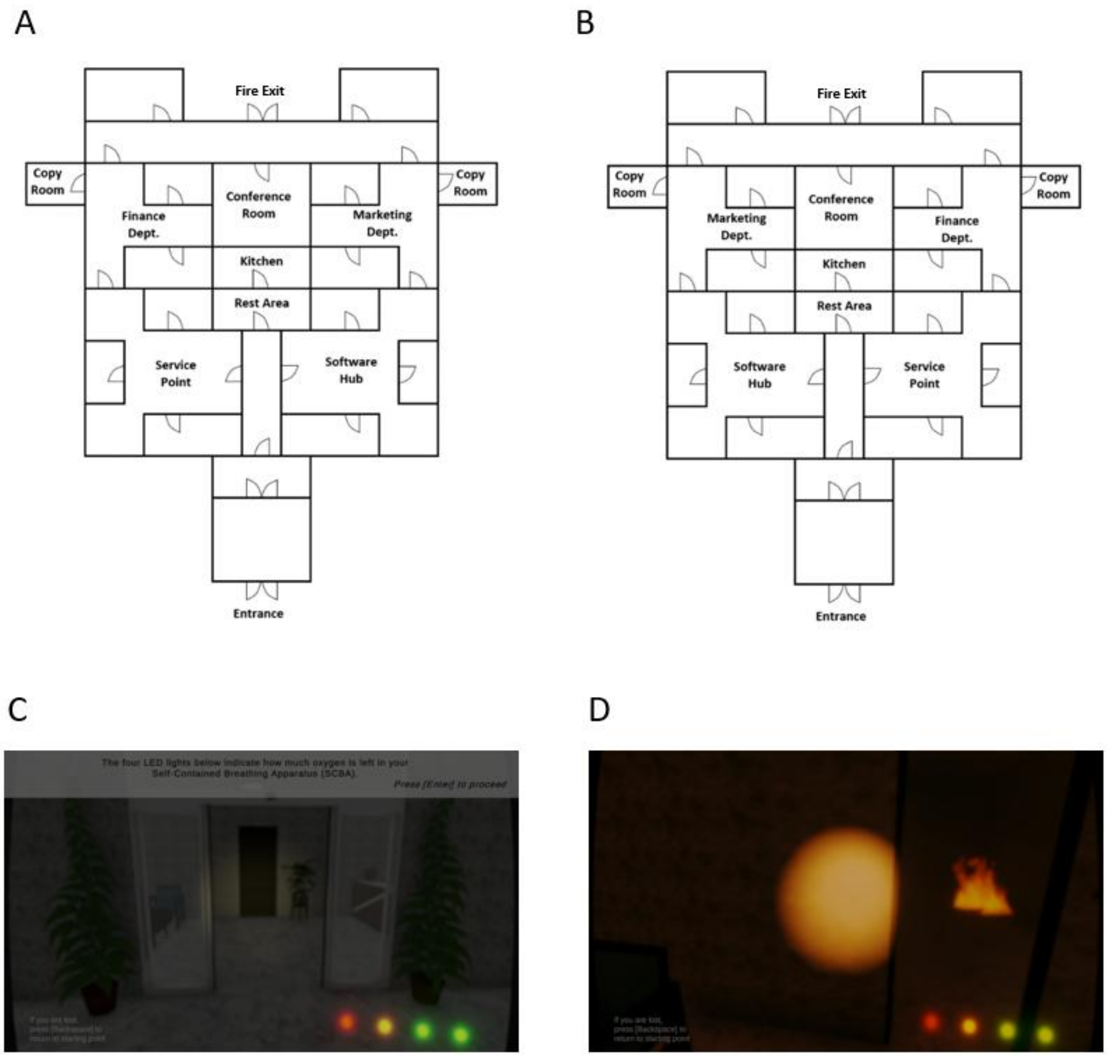


Figure 2.1. Experiments 1-3: Floorplans representing the two configurations of the virtual buildings (A and B) and a screenshot of the entrance to the buildings during the familiarization stage (C) and the search and rescue mission (D).

2.3.1 Method

2.3.1.1 Participants. Forty-eight participants (43 females, 5 males; mean age = 19.76 years, range: 18.08-27.08 years) were recruited from the student population of the School of Psychology, Cardiff University, and received course credit for their participation. All participants had normal or corrected-to-normal vision. The research reported in this paper was approved through the School of Psychology Research and Ethics Committee (EC.22.04.26.6560GRA).

2.3.1.2 Materials. Participants were instructed to carefully follow the on-screen instructions throughout the game. All ambient audio and sound effects were played through in-ear headphones at 65% of the computer's maximum volume (OEGStone BOAMOT-508). I built the Search and Rescue (SAR) computer game using Unity 2022.3.4f1 and the Unity Experiment Framework (Brookes et al., 2020), and it was run full screen on desktop computers with 23.5 x 13.2" screens. Participants started each phase of the game at their own pace by pressing the "B" key when prompted. They used the computer mouse to orient themselves with respect to the building, and pressed the W, A, S, and D keys on the computer keyboard to move forward, left, right, or backward, respectively.

In the office building, the rooms along the central axis of the building were the Common Area, Kitchen, and Conference Room (see Figure 2.1). Flanking the central axis were work areas: Finance Department and Service Point on one side, and Marketing Department and Software Hub on another. Each of these four work areas had three smaller rooms within them. Two toilets and the Fire Exit were located along the corridor at the back of the building that connected the Finance and Marketing Departments. All spaces shown in the floorplans were labelled with a door or wall plaque; and the doors to the spaces were opened by pressing the R key. Instructions were presented on-screen at different points throughout the study. The transposed floorplans of the virtual environments (Floorplans A and B) presented in Figure 2.1 showed that while the central areas of the environment remained the same, the rooms on each side were exchanged: Finance Department with Marketing Department and Service Point with Software Hub. Half of the participants in each of the four groups received one virtual environment throughout the study and the remainder received the second environment. Verbatim instructions of the game are provided in Appendix A.

2.3.1.3 Procedure.

Familiarisation. Participants were first shown how to use the computer mouse to orient themselves, and to press the W, A, S, and D keys on the computer keyboard to move forward, left, right, or backward, respectively. They then pressed "Enter" to begin training the use of these controls in an outside virtual space (unrelated to the SAR mission) to walk on translucent platforms towards six animated markers at different locations in virtual space. Reaching each marker triggered a congratulatory sound effect, whereas moving off the platforms resulted in an error sound effect and the test being restarted. This pre-training continued until they had approached each of the animated markers. The participants then explored the office building with the same layout as they would later enter as a firefighter during the SAR mission. At this point, the building was not engulfed in smoke and fire, and the participants simply explored the building with the task of retrieving 17 items (e.g., folders, pizza boxes) between different areas of the building in a fixed sequence. For example, the first two tasks were to: "*Get the broken laptop from the Quality Assurance Specialist's office in the Software Hub*." and "*Leave the broken laptop on the red tray in the IT Department in the Service Point.*". To encourage the considered opening of doors, participants were not allowed to proceed within the game for four seconds whenever they opened a door, during which they were immobilised. They could re-enter an area freely if the door had already been opened. This rule was also applied during the SAR mission. All doors were reset upon exiting the building. Appendix A contains the full set of instructions for this pre-training.

To assess whether participants had learned the building layout, they received a floorplan labelling test where they dragged and dropped nine labels from the left of the screen (e.g., Finance Department, Conference room presented in a random 3×3 array) into an unlabelled floorplan building (screen dimensions: 22cm × 26.5cm) on the right. Another 3×3 array below the labels included nine "Not sure" labels for denoting areas they were uncertain about. Five areas common to both environments (i.e., Common Area, Kitchen, Conference Room, the two Copy Rooms) will henceforth be denoted Fixed Areas, and the remaining four areas (i.e., Finance Dept, Service Point, Marketing Department, and Software Hub) will be denoted Flanking Areas.

Search and Rescue Mission Cover Story. Participants first received a cover story about playing the role of a firefighter. They were told that an electrical fault with the main electrical supply had plunged the building into darkness and their mission was to rescue a trapped person (the Assistant Finance Manager) and remove a portable stove from the Kitchen quickly. They were then introduced to features of the virtual environment: The four virtual LEDs at the bottom right corner of the screen (see Figure 2.1) that indicated the air supply level in their Self-Contained Breathing Apparatus (SCBA). Each LED would flash three times before deactivating, signalling a reduction in air supply. Participants were also informed that when the last red LED light started flashing, they had limited air left to complete the mission; and were shown to collapse and perish when the air ran out, as represented by the first-person view dropping to the floor and the surroundings fading to black. This effect (i.e., air running out), however, was not a feature of the SAR mission. Participants were also instructed to press "Spacebar" when their visor started fogging and laboured breathing sounds appeared, which temporarily removed these hazards. These effects would worsen if they were not removed, and participants tested the effect of pressing the spacebar to remove them. There were also told that if they felt lost at any point, they could press the "Backspace" key to quickly return to the starting position, with the prompt being always visible at the bottom left corner of the screen: "*If you are lost, press [Backspace] to return to starting point*". After this introduction, they could either revisit the interactive instructions or proceed to the SAR mission.

Search and Rescue Mission. Participants were first informed: "*You will be shown a floorplan that might be useful for the search and rescue mission*." For half of them, this floorplan accurately reflected the area labels inside the building; for the remainder, the positions of Flanking Areas were transposed on the floorplan (see Figure 2.1A and B). Half of those who received an accurate floorplan saw it presented upright on the screen (group Accurate-upright), and the remainder saw it rotated through 180° (group Accurate-rotated). The same applied to those who received an inaccurate floorplan (group Inaccurate-upright, group Inaccurate-rotated). The identities of the virtual buildings represented in Floorplans A and B were fully counterbalanced. Participants had a maximum of two minutes to view the floorplan and could proceed to the SAR mission at any time by pressing an on-screen button. They then entered the dark, smoke-filled burning building with a fire alarm sounding. The virtual headtorch, linked to their current orientation, provided focal illumination during navigation. Each SCBA LED deactivated when the participant entered one of the seven invisible trigger points (see Figure 2.2), until only the red LED remained flashing. Another six invisible trigger points (see Figure 2.2) controlled three instances of screen fogging and laboured breathing. When triggered, the laboured breathing audio appeared. After 15 to 27 seconds, during which the screen gradually fogged, the instruction "*Press [Spacebar] to regulate breathing and reduce fogging.*" appeared. Failure to act within three to five seconds resulted in the fogging becoming worse until the "Spacebar" key was pressed. For a random two of three instances, removal of these effects was delayed by three to five seconds. Another instance would not trigger until the previous one was resolved or after all three instances had occurred. The mission ended when participants exited via the Fire Exit regardless of tasks completion.

Participants then rated, on a 10-point scale, perceived difficulty to cope ("*How were you coping with the search and rescue tasks just now?"*), with 1 being "feeling no pressure" to 10 being "unable to cope with the pressure" (Butler et al., 2021).

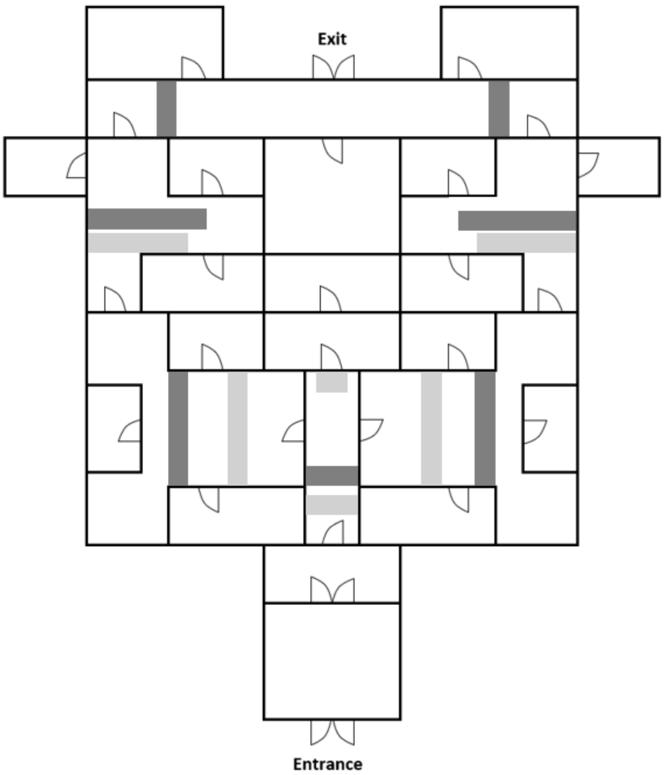


Figure 2.2. Experiments 1 and 2: Plan of the virtual building with the locations that triggered deactivation of SCBA LEDs (dark grey boxes) and laboured breathing (light grey boxes) during the SAR mission.

Response measures for the SAR mission. The mission was self-paced, and the time taken to rescue the trapped person was necessarily variable across participants and groups. Two measures assessed the impact of floorplan accuracy and orientation. The first was whether the first turn within the building (i.e., left or right) was towards the Finance Department as indicated on the floorplans, as opposed to the building layout during Familiarisation or the in situ door/wall labels. The second measure was the number of redundant paths taken before reaching the Finance Department. Entering an area through a doorway was coded as one path unit. As some participants did not search the Kitchen for the portable stove, both doors leading to it were excluded from the redundant paths analysis. Consequently, the most efficient pathway to the Finance Department was four path units, with any deviations from it considered redundant.

If the floorplan was being used, all participants should turn in the direction specified by the floorplan provided. Those given accurate floorplans should have fewer redundant paths than those given inaccurate floorplans. If participants used their experience during Familiarisation or the in situ door/wall labels, then the Accurate groups should turn in the direction consistent with the floorplan, but the Inaccurate groups should turn in the opposite direction indicated by the floorplan. There would be no reason to expect a difference in the number of redundant paths between the groups.

All statistical tests were performed using R, version 4.4.0 (R Core Team, 2024). Analyses were collapsed across the fully counterbalanced factor of virtual environment identity of the building layouts, because there were no significant differences in labelling errors between the environments (Mann-Whitney U test, $Z = 0.44$, $p = .663$) or correct identification of the general direction of the Finance Department in the building (Fisher's exact test, $p = 1.00$). Where the assumptions required for the use of parametric statistical tests were violated, non-parametric alternatives were conducted. Bayes factors (BF_{10}) using default priors were also computed for SAR mission response measures using the BayesFactors R package (Morey et al., 2024) to quantify the strength of evidence for the alternative hypothesis relative to the null based on the data. For data with non-normal distribution, log-transformed data were used if it reduced skewness; otherwise, raw data were analysed. BF_{10} values greater than one indicate stronger support for the alternative hypothesis, whereas values less than one indicate stronger support for the null hypothesis; evidence categories are based on Andraszewicz et al. (2015).

2.3.2 Results

2.3.2.1 Familiarisation. In the floorplan labelling test, which followed exploration of the virtual environment, correct answers were expressed as a percentage of the total number of blanks in the Fixed and Flanking Areas of the building (five and four, respectively). A significant proportion (n = 39) correctly identified the general direction of the Finance Department (i.e., the correct side of the building; binomial test, $p < .001$, 95% CI [67.37%, 91.05%]). One-sample Wilcoxon signed-rank tests showed that participants performed above chance levels (mu = 31.50% and 24.35% for Fixed and Flanking areas, respectively[4]) when

[4]The chance level performance (mu) for an area was computed by summing the probabilities of correctly labelling each blank in an area. Given that labelling a blank reduced the available label options for the remaining blanks, these probabilities were calculated using a ‘sampling without replacement’ approach:

$$P(\text{correct label}) = \frac{1}{\text{total options} - \text{labelled blanks}}$$

Therefore, the probability of correctly labelling the first blank is $\frac{1}{18-0}$ and the probability of correctly labelling the second blank is $\frac{1}{18-1}$, and so on. As participants could label the floorplan in any order, the chance level performance (mu) for both areas were approximated as:

$$\mathrm{Mu}_{\text{Flanking Areas}} = \frac{1}{18} + \frac{1}{17} + \frac{1}{16} + \frac{1}{15} = 24.35\%$$

$$\mathrm{Mu}_{\text{Fixed Areas}} = \frac{1}{18} + \frac{1}{17} + \frac{1}{16} + \frac{1}{15} + \frac{1}{14} = 31.50\%$$

labelling the Fixed Areas (median = 100%, IQR = 0%, $V = 1167$, $p < .001$) and Flanking Areas (median = 100%, IQR = 50%, $V = 1141$, $p < .001$).

2.3.2.2 Search and Rescue Mission: Engagement. Mean durations (in seconds, s) with access to the floorplans were also similar across the four groups: 18.56s (SEM = 2.77s) for group Accurate-upright, 28.60s (SEM = 6.13s) for group Accurate-rotated, 24.91 (SEM = 2.63s) for group Inaccurate-upright, and 27.05s (SEM=2.91s) for group Inaccurate-rotated. A factorial ANOVA showed no significant main effects of floorplan accuracy, $F(1, 44) = 0.38$, $p = .541$, $\eta^2_G = .00$, floorplan orientation, $F(1, 44) = 2.45$, $p = .125$, $\eta^2_G = .05$, and no interaction between these factors, $F(1, 44) = 1.03$, $p = .316$, $\eta_p^2 = .02$. The four groups also spent similar durations in the SAR mission: 217.44s (SEM = 34.96s) for group Accurate-upright, 282.72s (SEM = 90.00s) for group Accurate-rotated, 217.55s (SEM = 33.68s) for group Inaccurate-upright, and 243.12s (SEM = 22.71s) for group Inaccurate-rotated. A factorial ANOVA showed no significant main effects of floorplan accuracy, $F(1, 44) = 0.14$, $p = .708$, $\eta^2_G = .00$, floorplan orientation, $F(1, 44) = 0.75$, $p = .390$, $\eta^2_G = .01$, and no interaction between these factors, $F(1, 44) = 0.14$, $p = .706$, $\eta^2_G = .00$. Median perceived difficulty to cope was 5.50 (IQR = 4.00) for group Accurate-upright, 7.00 (IQR = 3.00) for group Accurate-rotated, 6.00 (IQR = 2.25) for group Inaccurate-upright, and 7.00 (IQR = 1.25) for group Inaccurate-rotated. A Kruskal-Wallis test showed no significant differences between groups, $H(3) = 5.47$, $p = .140$.

Table 2.1 shows the numbers of participants who rescued the trapped person and retrieved the portable stove in the four groups for the complete sample and the subset of 39 who correctly identified the general direction of the Finance Department on the floorplan. Binomial tests showed that a significant majority rescued the person (70.83%, $p < .01$) and retrieved the portable stove (83.33%, $p < .001$), suggesting high levels of engagement with the

SAR mission. For the subset of 39, a significant majority retrieved the portable stove (82.05%, $p < .001$) but a non-significant majority rescued the person (64.10%, $p = .108$). Fisher exact tests showed no significant group differences in these proportions in the complete sample (Person, $p = .716$; Stove, $p = .950$) and the subset of 39 (Person, $p = .775$; Stove, $p = .896$).

Table 2.1. Numbers (and Percentages) of Participants who Rescued the Person and Retrieved the Portable Stove

	Complete sample (*n*=48)		Subset (*n*=39)	
	n	Person / Stove	*n*	Person / Stove
Group				
Accurate-upright	12	10 (83%) / 9 (75%)	9	7 (78%) / 7 (78%)
Inaccurate-upright	12	7 (58%) / 10 (83%)	11	6 (54%) / 9 (82%)
Accurate-rotated	12	9 (75%) / 10 (83%)	9	6 (67%) / 7 (78%)
Inaccurate-rotated	12	8 (67%) / 11 (92%)	10	6 (60%) / 9 (90%)

2.3.2.3 Search and Rescue Mission: Navigation. The number (and corresponding percentages) of participants whose first turn was in the direction of the Finance Department indicated by the accurate and inaccurate floorplans are shown in Table 2.2 for the complete sample of 48 (left column) and the subset of 39 (right column). Across groups, a majority of participants took their first turn within the building in the direction of the Finance Department as indicated by the floorplans. Fisher's exact test showed that this proportion did not differ significantly across the four groups, $p = .431$, and the corresponding Bayes factor ($BF_{10} = 0.20$) indicated anecdotal evidence for the null hypothesis. A binomial test showed that the participants were more likely to turn in that direction than chance would predict, $p < .001$. However, it is worth noting that binomial tests conducted on the individual groups revealed that participants in groups Accurate-upright and Inaccurate-upright were more likely to turn in that direction (p = .006, BF_{10} = 12.67, and p = .039, BF_{10} = 3.83, respectively), providing strong and moderate evidence for a directional bias during navigation. In contrast, there was no significant bias in groups Accurate-rotated and Inaccurate-rotated (ps = .388), and the corresponding Bayes factors (BF_{10} = 0.90 for both groups) provide anecdotal evidence in favour of the null hypothesis. Similarly, the subset of 39 showed no significant difference in the proportion of those whose first turn was in that direction, Fisher's exact test, p = .664, BF_{10} = 0.31. Supplementary analyses revealed a significant bias with moderate and strong evidence in groups Accurate-upright and Inaccurate-upright (p = .039, BF_{10} = 3.74, and p = .012, BF_{10} = 8.25, respectively), but not in groups Accurate-rotated and Inaccurate-rotated (p = .180, BF_{10} = 1.50, and p = .344, BF_{10} = 1.02, respectively), where only anecdotal evidence for a directional bias was observed.

Table 2.2. Numbers (and Percentages) of Participants whose First Turns were Consistent with the Floorplan

	Complete sample (*n*=48)	Subset (*n*=39)
Group		
Accurate-upright	11 of 12 (91.67%)	8 of 9 (88.89%)
Inaccurate-upright	10 of 12 (83.33%)	10 of 11 (83.33%)
Accurate-rotated	8 of 12 (67.67%)	7 of 9 (77.78%)
Inaccurate-rotated	8 of 12 (67.67%)	7 of 10 (70.00%)

Table 2.3 shows the median number of redundant paths in the four groups. Participants in both Accurate groups took fewer redundant paths than those in both Inaccurate groups. A Kruskal-Wallis test confirmed a significant difference in the median number of redundant paths taken

across the four groups, $H(3) = 9.38$, $p = .025$, $BF_{10} = 1.29$. Planned comparisons using Mann-Whitney U tests showed that the Accurate-upright group took significantly fewer redundant paths than the Inaccurate-upright group, $Z = 2.67$, Bonferroni-corrected $p = .032$, BF10 = 5.47, indicating moderate evidence for the group difference. The difference between the Accurate-rotated and Inaccurate-rotated groups was not significant, $Z = 0.99$, Bonferroni-corrected $p = 1.00$, BF10 = 0.44, suggesting anecdotal evidence supporting the null. Similarly, for the subset of 39 participants, there was a significant difference in the median number of redundant paths taken across all groups, $H(3) = 9.50$, $p = .023$, $BF_{10} = 0.97$. While group Accurate-upright took significantly fewer redundant paths than group Inaccurate-upright, $Z = 2.61$, Bonferroni-corrected $p = .036$, $BF_{10} = 3.20$ (moderate evidence), there was no significant difference between groups Accurate-rotated and Inaccurate-rotated, $Z = 1.57$, Bonferroni-corrected $p = .468$, $BF_{10} = 0.64$, (anecdotal evidence).

Table 2.3. Median Numbers of Redundant Paths (IQR)

	Complete sample (*n*=48)	Subset (*n*=39)
Group		
Accurate-upright	1.00 (1.00)	1.00 (0.00)
Inaccurate-upright	2.50 (2.25)	3.00 (2.00)
Accurate-rotated	1.00 (4.00)	1.00 (3.00)
Inaccurate-rotated	3.00 (3.50)	3.00 (4.25)

2.3.3 Discussion

Experiment 1 examined whether the use of a floorplan described as "might be useful" during a virtual indoor search and rescue (SAR) mission reflected its uncertain utility. Despite this uncertainty, the search of the building reflected the floorplan that participants received: In the context of a high-stakes emergency, the implicature seemed to have failed in cautioning uncritical reliance on the floorplan. Across conditions, Bayes factors indicated moderate to strong evidence for a directional bias when the floorplan was upright, but only anecdotal evidence supporting the absence of such a bias when the floorplan was rotated. Consequently, among participants who received upright floorplans, those with the inaccurate floorplan took more redundant paths to reach the Finance Department, whereas those with accurate floorplan reached there directly. One potential explanation is that the uncertainty conveyed about the floorplan was not encoded. It is also possible that participants appeared to rely less on the rotated floorplan (regardless of its accuracy) because they had difficulty mentally rotating it correctly. Participants may have also failed to recognise the building experienced during SAR as the same as the building from the Familiarisation stage, and conflicting information on the walls and doors was not used as they only became visible at closer distances. Experiment 2 attempted to increase the salience of the uncertainty about the utility of the floorplan, while explicitly linking the building explored during Familiarisation to that during SAR.

2.4 Experiment 2

All participants received an inaccurate floorplan between Familiarisation and the SAR mission that was presented in an upright orientation. Participants in group Inaccurate-same were given similar information about the floorplan to those in the group Inaccurate-upright in Experiment 1, while participants in group Inaccurate-outdated were also informed that the floorplan “could be outdated”.

2.4.1 Method

2.4.1.1 Participants and Materials. Thirty-four participants (24 females, 10 males; mean age = 19.89 years, range: 18.54-25.25 years) with normal or corrected-to-normal vision were recruited from the student population of the School of Psychology, Cardiff University, and received course credit for their participation. The materials were identical to Experiment 1 except for a confidence rating of the answers provided in the floorplan labelling test. Also, the movements of the participants within the building were recorded automatically by a tracker, invisible to the participants, that allowed search patterns during the SAR mission to be assessed with greater granularity than was possible in Experiment 1.

2.4.1.2 Procedure. The Familiarisation stage was identical to Experiment 1 with the exception that after the floorplan labelling test, participants rated their confidence in their answers on a scale from 1 ("not confident at all") to 10 ("very confident") to the question: "*How confident are you of your final answer?*". The cover story for the SAR mission remained the same as before. Before the mission, participants in the group Inaccurate-same (n = 16) were told: "*We have found a floorplan of the same building, which might or might not be useful for the search and rescue mission.*"; while those in group Inaccurate-outdated (n = 18) were informed: "*We have found a floorplan of the same building that could be outdated, which might or might not be useful for the search and rescue mission.*". Participants then received an upright, inaccurate floorplan of the building. All remaining procedures were the same as Experiment 1.

2.4.2 Results

Figure 2.3 depicts representative examples of search patterns in groups Inaccurate-same (left panel) and Inaccurate-outdated (right panel). Appendix B contains the complete set of individual search patterns.

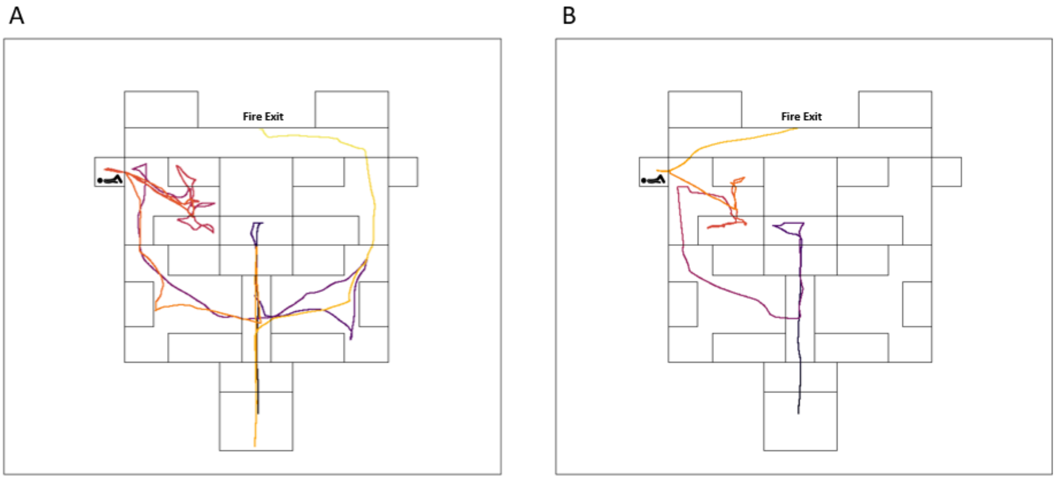


Figure 2.3. Representative patterns of search behaviour for participants in groups Inaccurate-same (A) and Inaccurate-outdated. The path showing movement through the building began in dark purple (starting point) transitioned to yellow (destination). The location of the trapped Assistant Finance Manager is indicated by the lying person icon. The exit from the building was through the Fire Exit.

2.4.2.1 Familiarisation. A binomial test showed a significant majority (n = 29) correctly identified the general direction of the Finance Department, $p < .001$, 95% CI [68.94%, 95.05%]. One-sample Wilcoxon signed-rank tests showed participants performed above chance when labelling Fixed areas (median = 100%, IQR = 0%, V = 588, $p < .001$) and Flanking areas of the building (median = 100%, IQR = 50%, V = 590, $p < .001$). There were also no significant differences in the confidence ratings provided by groups Inaccurate-same (mean = 6.75, SD = 1.39) and Inaccurate-outdated (mean = 7.44, SD = 0.92), $t(25.6) = -1.69$, $p = .102$.

2.4.2.2 Search and Rescue Mission: Engagement. A Mann-Whitney U test showed no significant differences in the durations spent (in seconds, s) in the SAR mission between the groups Inaccurate-same (median = 211.44s, IQR = 314.78s) and Inaccurate-outdated (median = 168.97s, IQR = 138.24s), $Z = 1.79$, $p = .073$. Time spent with the floorplan also did not differ significantly between the Inaccurate-same (median = 29.26s, IQR = 14.05s) and the Inaccurate-outdated groups (median = 27.49s, IQR = 15.77s), $Z = 0.79$, $p = .427$.

Table 2.4 shows the number of participants who rescued the person and retrieved the portable stove in the two groups. Binomial tests showed that approximately half of the participants (55.88%) rescued the person ($p = .608$) whereas a significant majority (88.24%) retrieved the portable stove ($p < .001$). Similar trends were observed for the subset of 29 participants, with 51.72% rescuing the person ($p = 1.00$) and 89.66% retrieving the portable stove ($p < .001$). Chi-square and Fisher's exact tests suggested no significant differences in these proportions between both groups in the complete sample (Person, $X^2(1) = 0.00$, $p = 1.00$; Stove, $p = .604$), and the subset of 29 (Person, $X^2(1) = 0.00$, $p = 1.00$; Stove, $p = .598$). Perceived difficulty to cope for groups Inaccurate-same (median = 7.00, IQR = 1.25) and Inaccurate-outdated (median = 7.00, IQR = 2.50) did not differ significantly, Mann-Whitney U test, $Z = 0.34$, $p = .736$.

Table 2.4. Number (and Percentages) of Participants who Rescued the Person and Retrieved the Portable Stove.

	Complete sample (*n* = 34)		Subset (*n* = 29)	
	n	Person / Stove	*n*	Person / Stove
Group				
Inaccurate-same	16	9 (56.25%) / 15 (93.75%)	15	8 (53.33%) / 14 (93.33%)
Inaccurate-outdated	18	10 (55.56%) / 15 (83.33%)	14	7 (50.00%) / 12 (85.71%)

2.4.2.3 Search and Rescue Mission: Navigation. Table 2.5 shows the number and percentages of participants whose first turn during the SAR mission was towards the Finance Department on the floorplan for both the complete sample of 34 participants and the subset of 29. In both samples, a greater proportion of group Inaccurate-same turned in this direction than group Inaccurate-outdated. Chi-square tests showed that the proportion of those whose first turn was in this direction did not differ in the complete sample, $X^2(1) = 3.14$, $p = .077$, Cramer's $V = .30$, but there was a significant difference in the subset of 29, $X^2(1) = 4.17$, $p = .041$, Cramer's $V = .38$. However, the corresponding Bayes factors ($BF_{10} = 3.47$ for the complete sample; $BF_{10} = 10.34$ for the subset of 29) suggested moderate and strong evidence for an association, respectively, despite the non-significant p-value in the complete sample. Participants in group Inaccurate-same were significantly more likely to turn in the direction indicated by the inaccurate floorplan ($p = .035$, $BF_{10} = 2.34$), whereas those in group Inaccurate-outdated were not ($p = .424$, $BF_{10} = 0.70$), showing anecdotal evidence for a directional bias in the first group and no bias in the second. In the complete sample, participants in group Inaccurate-same took more redundant paths (median = 2.00, IQR = 3.50) than those in group Inaccurate-outdated (median = 0, IQR = 2.00), Kruskal-Wallis, $H(1) = 202$, $p = .038$, $BF_{10} = 0.79$, suggesting anecdotal evidence for the null despite the significant p-value. This difference was also significant in the subset of 29 (Inaccurate-same, median = 2.55, IQR = 2.00; Inaccurate-outdated, median = 1.49, IQR = 0.00), $H(1) = 164$, $p = .007$, with moderate evidence for the difference ($BF_{10} = 7.56$).

Table 2.5. Numbers (and Percentages) of Participants whose First Turns were Consistent with the Floorplan

	Complete sample ($n = 34$)	Subset ($n = 29$)
Group		
Inaccurate-same	12 of 16 (75.00%)	12 of 15 (80.00%)
Inaccurate-outdated	7 of 18 (38.89%)	5 of 14 (35.71%)

The tracking data provided additional insights into the distribution of participants' search patterns, and in particular the time spent on the side of the building in which the Finance Department and Service Point (i.e., the Finance side) and the Marketing Department and Software Hub (i.e., the Marketing side) were actually located during the SAR mission; but excluding their six adjoining smaller rooms. Table 2.6 presents the time spent in these areas for both groups. Inspection of this table suggests that participants in group Inaccurate-same spent more time on the Marketing side than did group Inaccurate-outdated, and that both groups spent similar amounts of time on the Finance side. ANOVA revealed a main effect of building side, $F(1, 32) = 10.90$, $p = .002$, $\eta^2_G = 0.101$, $BF_{10} = 23.49$, no effect of message, $F(1, 32) = 2.58$, $p = .118$, $\eta^2_G = 0.05$, $BF_{10} = 0.88$, and a significant interaction effect between these factors, $F(1, 32) = 4.87$, $p = .035$, $\eta^2_G = 0.05$, $BF_{10} = 2.28$. Pairwise comparisons confirmed that group Inaccurate-same spent significantly more time on the Marketing side than did group Inaccurate-outdated ($p = .002$), with strong evidence for the group difference

(BF_{10} = 16.72). In contrast, the amount of time the groups spent on the Finance side did not differ significantly (p = .970), with anecdotal evidence for the null (BF_{10} = 0.33). Moreover, group Inaccurate-same spent a similar amount of time on both sides (p = .637), with moderate evidence for the null (BF_{10} = 0.30). However, group Inaccurate-outdated spent more time on the Finance than the Marketing side (p < .001), with extreme evidence for the difference (BF_{10} = 781.78). As the search durations were not normally distributed in two of the conditions, Mann-Whitney tests were conducted, which confirmed that there was a group difference in the amount of time spent on the Marketing side (Z = 2.88, p = .004), but not the Finance side (Z = -0.85, p = .398). Wilcoxon signed-rank tests showed no significant difference in time spent between the two sides in group Inaccurate-same (V = 82, p = .495), but a significant difference in group Inaccurate-outdated (V = 164, p < .001).

Table 2.6. Mean (SE) and Median (IQR) Time Spent in Work Areas by Building Side and Message Type

	Marketing Side		Finance Side	
	Mean (SE)	Median (IQR)	Mean (SE)	Median (IQR)
Group				
Inaccurate-same	55.93s (9.68s)	48.30s (38.44s)	62.96s (16.00s)	43.19s (61.62s)
Inaccurate-outdated	20.88s (5.86s)	8.33s (43.93s)	60.11s (6.72s)	55.00s (34.30s)

2.4.3 Discussion

Experiment 2 investigated whether highlighting the uncertainty utility of a floorplan affected its use in navigating through a virtual building during a SAR mission. Participants were informed that the floorplan they received was of the same building explored during Familiarisation, which "… might or might not be useful for the search and rescue mission" (group Inaccurate-same), or "… could be outdated, which might or might not be useful for the search and rescue mission" (group Inaccurate-outdated). Participants' first-turn direction was significantly associated with the instruction received, with moderate to strong support for the association as indicated by Bayes factors. Results from group Inaccurate-same replicated findings in Experiment 1: Participants' search patterns were influenced by the inaccurate floorplan, with moderate to anecdotal evidence for this effect. In contrast, fewer participants in group Inaccurate-outdated followed the floorplan, with anecdotal evidence suggesting no bias. Consequently, those in the Inaccurate-same group took more redundant paths to the Finance Department than those in the Inaccurate-outdated group who reached there directly. Tracking data showed that while both groups spent similar amounts of time on the SAR mission, group Inaccurate-same spent significantly more time than group Inaccurate-outdated searching on the Marketing side of the building for the Assistant Finance Manager, supported by strong to extreme evidence across pairwise comparisons. In summary, providing additional semantic cues about the (dis)utility of the (inaccurate) floorplan (i.e., that it "could be outdated") reduced the likelihood of participants using it as a basis to search the building.

2.5 Experiment 3

Experiments 1 and 2 examined the framing of floorplan information for a to-be-searched building. In both experiments, floorplans that "might (or might not) be useful" affected search patterns, irrespective of their consistency with prior knowledge of the building (gained during the Familiarisation stage) or the actual physical layout of the to-be-searched building. Experiment 3 examined whether the framing of SAR goal information itself (i.e., the presence of a person) affected search behaviour, with two stages involving SAR: training and a "real" mission. During SAR training, participants entered the smoke-filled building to rescue two people (the Finance Manager and the Marketing Manager) trapped in their respective Copy

Rooms, and retrieved a portable stove. For the SAR mission, group Explicit-message was informed that one of the managers (e.g., Finance Manager) was attending their off-site team-building event and not in the building. Group Implicit-message was informed that one of the departments (e.g., Finance Department) was holding their teambuilding event off-site, which could imply that the Finance Manager was absent. The control group received the same form of messages (Explicit or Implicit), but with reference to an IT Manager who was not part of SAR training: The participants in the Explicit-irrelevant sub-group were informed that the IT Manager was attending their teambuilding event off-site and not in the building, while those in the corresponding Implicit-irrelevant sub-group were simply informed that the IT Department was holding their teaming-building event off-site. Another point of departure from Experiments 1 and 2 was the use of indirect communication (as described in example (2), Section 1.3) rather than scalar implicatures. Nevertheless, the process of deriving an implicature was the same: The conclusion that a Manager was not in the building was reasoned through presumptions of cooperative communication and what a maximally relevant communicator would intend in a firefighting context.

The impact of message framing on the critical SAR mission was assessed with two measures. The first was whether participants visited one or both copy rooms in recovering the person during the SAR mission. The second measure was the amount of time spent in different areas of the building (Finance Department, Service Point, Marketing Department, Software Hub). If the goal of rescuing the Finance and Marketing Managers was influenced by the messaging about their likely presence, then participants should search only the relevant copy room and spend less time searching those areas where the messages had suggested that a given manager would be absent. Any such differences in exploration of the two copy rooms or search times should be absent where the messaging was about an irrelevant worker (i.e., IT Manager). As in Experiment 2, tracking data was also recorded.

2.5.1 Method

2.5.1.1 Participants and Materials. Forty-nine participants (43 females, 6 males; mean age = 19.39 years, range: 18.00–24.58 years) with normal or corrected-to-normal vision from the student population of the School of Psychology at a university received course credit for their participation. The materials were adapted from Experiment 2 with any changes described below.

2.5.1.2 Procedure.

Search and Rescue Cover Story. Participants were first informed that a company called Lemonade Tech had volunteered their office space and their Finance and Marketing Managers to facilitate a training exercise in which they would play the role of a firefighter. The first stage of Familiarisation and SAR training (described below) were portrayed as the training exercise, with a later SAR mission portrayed as a real incident that occurred in the same building two days later.

Familiarisation. The Familiarisation procedure was similar to Experiment 2 and took place in a well-lit building. The instructions were adjusted to encourage efficient wayfinding, and four more item retrieval tasks were added to the initial exploration to introduce the Finance Manager and the Marketing Manager in their respective Copy Rooms. When participants tried to exit the building through the Fire Exit, a notice appeared stating that it was jammed and awaiting repairs for several weeks, advising them to use the main entrance instead. They could disable the notice by pressing the “R” key and proceed with their task after a six-second wait. Any attempts to exit via the Fire Exit for the remainder of the experiment triggered the same procedure. After Familiarisation, participants received the same floorplan labelling test. To highlight the shortcut that linked the Marketing and Finance Departments, another space for the Back Corridor was added to the blank. Thus, new “Back Corridor” and “Not Sure” labels were added to the respective answer options. Appendix C contains the full set of instructions for Familiarisation.

Search and Rescue Training. After Familiarisation, participants were informed that virtual smoke would engulf the building to simulate a fire and that they would receive six SAR trials as part of their firefighting training. They were instructed to rescue the two volunteers (Finance and Marketing Managers) and retrieve the portable stove using the shortest route (see Figure 2.4) and in the shortest time. This stage was identical to Experiments 1 and 2, but without the option to press the “Backspace” key (to return to the starting position) to enhance realism and simplify the interpretation of tracking data. Participants were also informed that after each trial, they would receive feedback about whether they rescued the Finance and Marketing Managers, retrieved the portable stove, took the shortest route (see Figure 2.4), and were sufficiently quick in completing the trials.

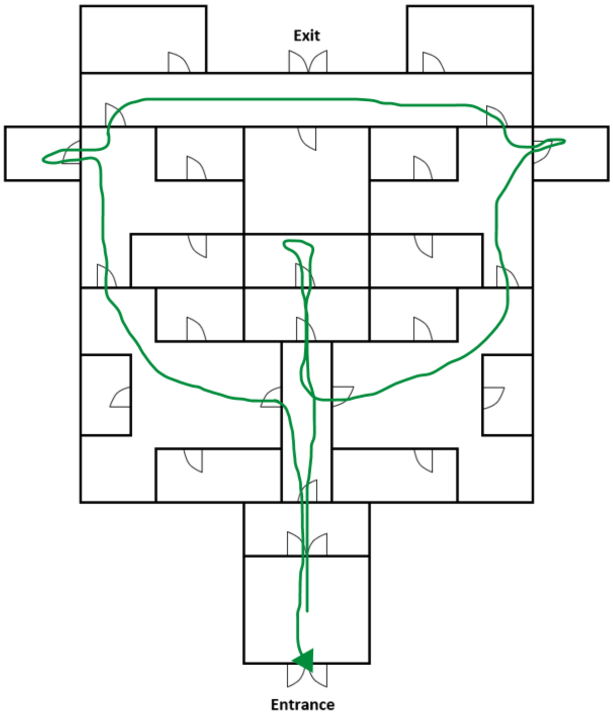


Figure 2.4. Experiment 3: Plan of the building with an example of an efficient route to rescue the Finance and Marketing Managers, retrieve the portable stove, and exit the building via the entrance.

Participants were equipped with a virtual headtorch that illuminated their surroundings in the quiet, smoke-filled building. Each of the six trials ended when they exited the building or after 4.5 min had elapsed. To encourage engagement with their training, participants were restricted from exiting the building until they had at least entered the Finance or Marketing Departments. This feature was also included in the "real" SAR mission. Forty to 80s after each trial began, one milder instance of screen fogging and laboured breathing occurred, which could be disabled by pressing the "Spacebar" key when prompted. If this key was not pressed, these effects would continue at the same intensity until the trial ended. Each of the four LED lights were also scheduled to turn off after every 55s to 70s, except for the red LED light which would continue to flash until the trial ended. These audio-visual effects were reset at the beginning of the next trial.

Feedback. At the end of each trial, participants rated their perceived difficulty to cope and were given feedback about their performance, for example:

- If participants rescued both the Finance and Marketing managers, and retrieved the portable stove using the shortest route, and were sufficiently quick, a congratulatory sound effect accompanied the feedback.
- If participants failed to exit the building in 4.5 minutes, a clock-ticking sound effect and a burning skull icon would accompany the feedback if they had rescue both target managers and retrieved the stove. Otherwise, a failure sound effect and the same burning skull icon would accompany the feedback instead.
- If participants exited the building in time but failed to complete the rescue and retrieval tasks, took longer than their "best time", and/or took a longer route than was required, a failure sound effect and a stopwatch icon would accompany the feedback.

The time taken to compete each trial was recorded only if participants rescued both the Finance and Marketing Managers and retrieved the portable stove using the shortest route (see Figure 2.4). The completion time for the first successful trial was recorded as their "best time". The time for each subsequent instance of successful trial was compared against the existing "best time". If the duration was within three seconds of the existing "best time", participants would be considered to have completed the trial quickly enough, with the shorter duration of the two replacing the "best time". If the duration was more than three seconds longer than the "best time", participants would be considered to not have completed the trial quickly enough and the "best time" would remain unchanged.

After the feedback was presented, participants proceeded to the next trial by pressing the "Enter" key after a five-second delay, until all six trials were completed. All variations of the verbatim feedback are detailed in Appendix D.

Search and Rescue Mission. Participants were told that they had performed well during SAR training and then received further instructions about the upcoming SAR mission. Presentation of the instructions/messages for the "real" SAR mission was sequenced over a minimum duration of 107 seconds to prevent skimming, with participants pressing the "Enter key" when prompted to reveal them in sequence. After a brief pause, during which three dots appeared in sequence suggesting the passing of time (see Appendix C: Search and Rescue Mission), a faint siren played in the background until the SAR mission began later. Participants were told that they had been called to the office building where they had their firefighting training two days ago, due to a fire caused by an electrical fault with the main electrical supply. They were instructed to rescue the Finance and Marketing Managers and retrieve the portable stove in the shortest time possible. They were also reminded that this was a "real incident", as opposed to the training exercise they undertook earlier.

After another pause, more messages appeared: "*We've just received a report. It is from the Receptionist of Lemonade Tech who was evacuated from the building some moments ago.*" Following that, half of the participants in group Explicit-message were told: "*The <u>Finance Manager</u> is attending their teambuilding event off-site and not here.*", and the remainder were told: "*The <u>Marketing Manager</u> is attending their teambuilding event off-site and not here.*". Half of the participants in group Implicit-message were told: "*The <u>Finance Department</u> is holding their teambuilding event in a nearby hotel.*", and the remainder were told, "*The <u>Marketing Department</u> is holding their teambuilding event in a nearby hotel.*" Thus, for those in the explicit group the absence of the Manager is explicit, but their whereabouts is not, whereas for those in the implicit group the absence of the Manager is implicit, but their whereabouts in explicit. Finally, half of participants in group Control received the same Explicit message, but with reference to the IT Manager and the remainder received the same Implicit message, but with reference to the IT Department. Participants then pressed the "Enter" key to proceed to the untimed SAR mission.

The virtual environment was identical to previous experiments, except that the deactivation of the SCBA LED lights was now time-based: The two green LEDs deactivated sequentially after 30s to 55s, followed by the orange LED 55s and 77s later; after that, the red LED started flashing 30s to 55s after the orange LED and continued until participants exited the building. The three instances of screen fogging and laboured breathing returned to the same intensity as in previous experiments. Only the Manager who was not at the teambuilding event was present and needed to be rescued. After exiting the building, the participants rated their perceived difficulty to cope and were told whether they had rescued the person and retrieved the portable stove.

Response Measures for the SAR Mission. We used two measures to investigate the impact of message framing. The first measure was time spent in the work areas of the building. The second measure was whether participants searched both Copy Rooms, clearly indicating the intent to rescue a second person. Group Explicit-message should spend the least time in these areas and search one Copy Room only, as they had no reason to search both sides of the building. In contrast, group Control should spend significantly longer in these areas and search both Copy Rooms, as they were not informed about the absent Marketing (or Finance) Manager and could have relied on initial task instructions or SAR training experience. If group Implicit-message (correctly) inferred that only one person was present, search behaviour would resemble group Explicit-message; otherwise, it would resemble group Control.

2.5.2 Results

Figure 2.5 shows representative examples of the patterns of search behaviour for a participant in groups Explicit-message (left panel), Implicit-message (centre panel), and Control (right panel). The full set of individual search patterns are presented in Appendix E.

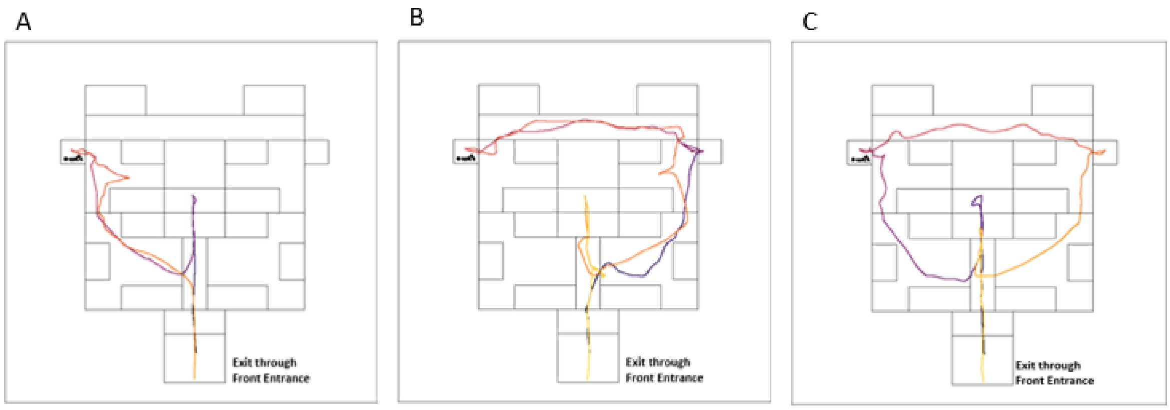


Figure 2.5. Representative patterns of search behaviour for participants in group Explicit-message (A), group Implicit-message (B), and group Control (C). The path showing movement through the building began in dark purple (starting point) transitioned to yellow (destination). The location of the trapped Manager is indicated by the lying person icon. The exit is through the Front Entrance.

2.5.2.1 Familiarisation and Search and Rescue Training. Floorplan labelling was expressed as a percentage of correct labels out of six in the Fixed Areas and four in the Flanking Areas. One-sample Wilcoxon signed-rank tests showed that participants performed above chance (mu = 34.62% and 21.70% for Fixed and Flanking areas, respectively) when labelling Fixed Areas (median = 100%, IQR = 33.33%, $V = 1187$, $p < .001$) and Flanking Areas (median = 50%, IQR = 50%, $V = 1176$, $p < .001$). Although 61.22% of the participants correctly labelled the Finance and Marketing Departments on their respective sides of the building (i.e., left or right), this was not above chance level according to a binomial test,

p = .152, 95%CI [46.24%, 74.80%]. A Kruskal-Wallis test found no significant differences in the confidence ratings between groups Explicit-message (median = 7.50, IQR = 3.25), Implicit-message (median = 6.00, IQR = 4.00), and Control (median = 7.00, IQR = 3.25), $H(2) = 1.97$, $p = .374$, $\eta^2 = 0.00$. There were also no significant differences in the median number of trials where participants rescued both managers and retrieved the portable stove during SAR training between groups Explicit-message (median = 6.00, IQR = 1.00), Implicit-message (median = 5.00, IQR = 1.00), and Control (median = 6.00, IQR = 1.00), $H(2) = 1.51$, $p = .471$, $\eta^2 = 0.01$.

Figure 2.6 shows the perceived difficulty to cope scores across the six SAR training trials and the SAR mission pooled across the three groups. The scores did not differ during the six training trials between the three groups (Explicit-message median = 2.58, IQR = 3.67; Implicit-message median = 4.00, IQR = 1.67; Control median = 3.83, IQR = 3.75; $H(2) = 1.17$, $p = .558$, $\eta^2 = 0.02$) nor during the real SAR mission (Explicit-message median = 6.00, IQR = 2.50; Implicit-message median = 8.00, IQR=2.00; Control median = 8.00, IQR = 1.25; $H(2) = 1.78$, $p = .411$, $\eta^2 = 0.00$). Inspection of the scores in Figure 6 shows that they gradually reduced from the first training trial (median = 5.00, IQR = 3.00) to the sixth (median = 2.00, IQR = 4.00) but increased again during the SAR mission (median = 7.00, IQR = 2.00). Friedman's test found a significant effect of training trial number, $X^2(6) = 137$, $p < .001$, *Kendall's W* = 0.47. While pairwise comparisons showed no significant differences between the scores for consecutive training trials, the scores for the first trial were significantly higher than those for the sixth ($p < .001$), and the scores for the sixth trial were significantly lower than those for the SAR mission ($p < .001$).

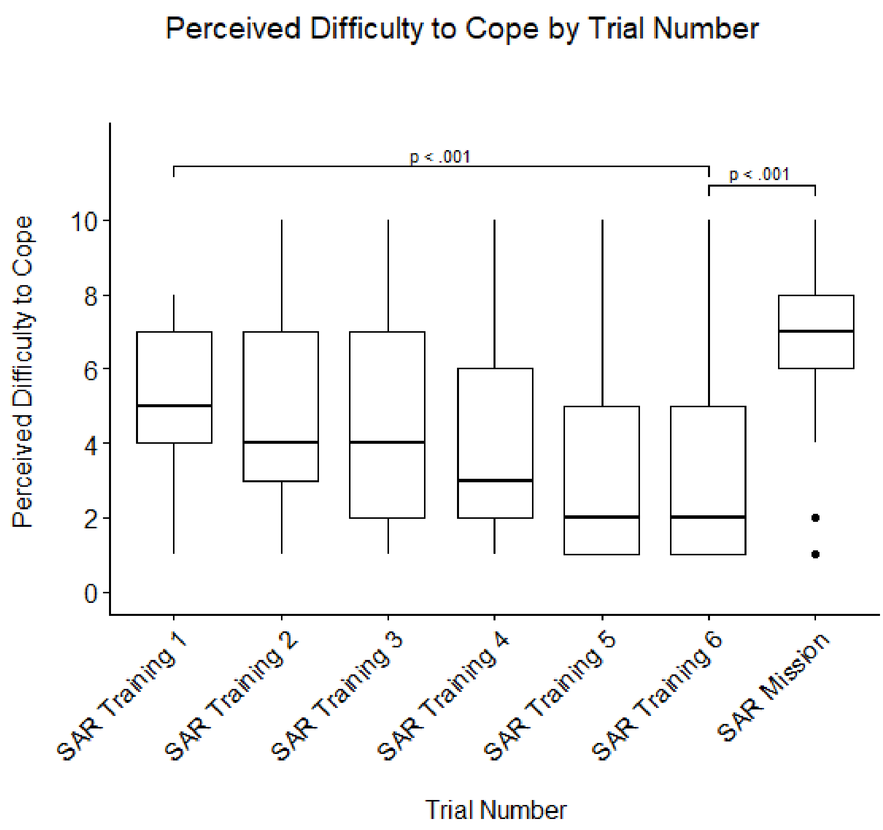


Figure 2.6. Experiment 3: Median perceived difficulty to cope with firefighter search and rescue (SAR) training trials (1-6) and the "real" SAR mission (bold horizontal lines; ±IQR). Individual participants scores are denoted by circles. During training, participants rescued the Finance and Marketing Managers in a virtual building, and retrieved a stove. The final search and rescue (SAR) mission was presented as a “real” incident in which their mission was the same, but two groups were given supplementary explicit or implicit information indicating that one of the managers was not in the building, and a third control group received redundant explicit or implicit information.

2.5.2.2 Search and Rescue Mission. Table 2.7 shows the number of participants who rescued the person and retrieved the stove in the three groups. All participants rescued the person, and the vast majority retrieved the portable stove (binomial test, $p < .001$). Fisher’s test confirmed that there were no significant group differences in stove retrieval, $p = .306$.

Table 2.7. Number (and Percentages) of Participants who Rescued the Person and Retrieved the Portable Stove

	N	Person / Stove
Group		
Explicit-message	16	16 (100%) / 15 (94%)
Implicit-message	17	17 (100%) / 17 (100%)
Control group	18	16 (100%) / 14 (87%)

Table 2.8 shows the number of participants (and corresponding percentages) in each group who visited only one Copy Room during the SAR mission in the process of successfully rescuing the person. Again, the two Control groups were combined as Fisher's exact test showed no significant differences in the proportion of those who visited one Copy Rooms instead of both, $p = 1.00$. A chi-square test revealed a significant association between message framing and Copy Rooms visited, $X^2(2) = 27.65$, $p < .001$, with extreme evidence ($BF_{10} \approx 2.08\times10^5$) for the association. Post hoc tests showed that significantly more participants in group Explicit-message visited only one Copy Room than groups Implicit-message ($p < .001$, $BF_{10} \approx 1.06\times10^4$) and Control ($p < .001$, $BF_{10} \approx 3.36\times10^4$), with extreme evidence supporting these effects. These proportions did not differ significantly between the latter two groups ($p = 1.00$, $BF_{10} = 0.32$), with moderate evidence supporting the null.

Table 2.8. Numbers (and Percentages) of Participants who Visited Only One Copy Room

Group	
Explicit-message	15 of 16 (93.75%)
Implicit-message	3 of 17 (17.65%)
Control group	2 of 16 (12.50%)

The tracking analysis revealed that the durations of the SAR missions differed between groups. For SAR mission durations (in seconds, s), groups Control-explicit (mean = 100.13s, SE = 14.58s) and Control-implicit (mean = 124.92s, SE = 22.97s) did not differ significantly, $t(11.85) = 0.91$, $p = .380$, and were therefore combined. The durations spent completing the SAR mission were shorter for participants in group Explicit-message (median = 128.15s, IQR = 77.96s) than both groups Implicit-message (median = 191.52s, IQR = 47.94s) and Control (median = 175.87s, IQR = 123.72s). A Kruskal-Wallis test confirmed a significant effect of message framing, $H(2) = 10.30$, $p = .006$, $\eta^2 = 0.18$, with strong evidence supporting this effect ($BF_{10} = 17.59$). Dunn's test confirmed that group Explicit-message spent significantly less time completing the SAR mission than groups Control ($p = .026$, $BF_{10} = 9.80$) and Implicit-message ($p = .010$, $BF_{10} = 17.99$), with moderate and strong evidence for these differences, respectively. In contrast, there was no significant difference between groups Control and Implicit-message ($p = 1.00$, $BF_{10} = 0.34$), with anecdotal evidence supporting the null. Analysis of the time spent in work areas revealed a complimentary pattern of results. The two Control groups were again combined, with an independent t-test confirming that there was no significant difference in time spent in work areas between groups Control-explicit (mean = 100.13s, SE = 14.58s) and Control-implicit (mean = 124.92s, SE = 22.97s), $t(11.9) = 0.91$, $p = .380$. The median (IQR) time spent in the work areas was 62.71s (12.88s) for group Explicit-message, 91.96s (29.99s) for group Implicit-message, and 89.53s (75.26s) for group Control. A Kruskal-Wallis test revealed a significant effect of message framing, $H(2) = 10.50$, $p = .005$, $\eta^2 = 0.19$, with moderate evidence supporting this effect ($BF_{10} = 9.43$). Dunn's test confirmed that the amount of time spent in the work areas was shorter in group Explicit-message than in groups Control ($p = .021$, $BF_{10} = 4.16$) and Implicit-message ($p = .010$, $BF_{10} = 10.25$), with moderate evidence for these differences. There was no significant difference between groups Control and Implicit-message (p = 1.00, BF_{10} = 0.33), with anecdotal evidence supporting the null.

2.5.3 Discussion

Experiment 3 examined whether implicit or explicit framing of linguistic information about the potential absence of one of two potential people who needed to be rescued influenced search patterns for people during a virtual SAR mission. Participants first received six SAR training trials where they practiced rescuing both the Finance Manager and Marketing Manager, and retrieving the portable stove, and then participated in a "real" SAR mission in a burning building. Before this mission, they received information that either explicitly or implicitly suggested that one of the previously rescued managers was absent or an irrelevant person (not present during training) was absent. Search patterns were significantly associated with message framing, with extreme evidence for this effect. Participants given an explicit message about the absence of one of the previously rescued managers searched in the location where the remaining manager was located: Their search behaviour was driven by the goal of rescuing the manager they were told was still in the building. The implicit message about the remaining manager did not show a clear change on search behaviour: Participants continued to search in the locations where the two managers had been located during training. In fact, participants given an implicit message were no less likely to visit both locations than those given an (explicit or implicit) message about a manager who had not been present during SAR training, with moderate evidence supporting this null difference.

2.6 General Discussion

Search and Rescue (SAR) emergencies involve high stakes and are consequently stressful. An effective SAR response relies on communication between members of the response team (e.g., the incident commander and members of their crew). We investigated pragmatic reasoning in simulated (computer-based) SAR emergencies where undergraduate students played the role of a firefighter. This reasoning involved the intended meaning of a communication provided before an SAR mission as opposed to its literal meaning. In Experiments 1 and 2, participants received accurate or inaccurate floorplans of the building in which their primary task was to search for and rescue a person who was trapped. They were told that these floorplans “might be useful” in Experiment 1 and “might or might not be useful” in Experiment 2. In both experiments, participants headed in the direction of the likely location of the person indicated by the floorplan irrespective of whether it reflected their prior experience with the building or was consistent with the signage in the building during the mission. That is, the scalar word “might” did not show a clear effect on the use of the floorplan by participants. It was only when it was suggested that the floorplan could be “outdated” that participants became less reliant on the floorplan. Experiment 3 assessed how the provision of explicit or implicit information about the likely presence of to-be-rescued people affected search behaviour. Explicit information had a marked impact, but implicit information did not appear to influence search behaviour relative to a group of participants who received (explicit or implicit) information about an irrelevant person. To summarise, Experiments 1-3 provided little evidence that participants interpreted conversational implicatures as intended in virtual SAR scenarios, unless a salient semantic cue was provided to guide pragmatic reasoning (e.g., “could be outdated” in Experiment 2). We shall consider in turn the theoretical frameworks that provide potential insights into these observed limitations in pragmatic reasoning.

Our virtual SAR missions induced relatively high levels of stress in participants, and stress is known to disrupt working memory and cognitive control (Bambini et al., 2021; Cain et al., 2011; Diamond, 2013; Hyun et al., 2019; Lazarus, 1993; Shields et al., 2016), cognitive processes which have been shown to support pragmatic processing in a variety of contexts

(Bott & Noveck, 2004; De Neys & Schaeken, 2007; van Tiel, Marty, et al., 2019; van Tiel, Pankratz, et al., 2019). Indeed, the Stress Induced Deliberation-to-Intuition model (Yu, 2016) posits that acute stress has a negative impact on higher-order cognitive processes, including working memory and cognitive control, and this impact results in a shift from deliberative processes (System 2) to intuitive processes (System 1) during decision-making. To the extent that implicatures require deliberative processes, the stress induced by our SAR missions would be expected to disrupt the processing of implicatures and affect a reliance on heuristics (e.g., use the floorplan in Experiments 1 and 2; see Moeschler, 2023). Consistent with this analysis, when firefighters are under increased stress (induced by different simulated emergency incidents) they are more likely to rely on standard operating procedures than to use deliberative processes to respond to an emergency incident (Butler et al., 2023). This was most evident in Experiment 3, where participants generally continued to search for both Managers, as they had been trained, despite the implicature indicating that only one needed to be rescued. The more general idea that cognitive economy is prioritised under stress (Bogdanov et al., 2021; Plessow et al., 2012; Schwabe & Wolf, 2010) is also consistent with the observation that wayfinding in Experiments 1 and 2 was based on floorplan provision, despite available information (i.e., wall plaques) that the floorplans were inaccurate (cf. Hölscher et al., 2011). Previous research has found that route maps were recoded into description-based directions rather than in terms of spatial representations (Padgitt & Hund, 2012) and it seems possible that route planning based on the floorplan in our SAR missions could have been reduced to a simple directional rule (e.g., "Keep to the right to reach the Finance Department").

Another approach to understanding the processing of implicatures is provided by Relevance Theory (Sperber & Wilson, 1986, 1987). This theory proposes that implicatures are not processed automatically but are only processed when the communicative context provides the necessary background assumptions for interpreting them. These assumptions include contextual knowledge and shared expectations between interlocutors (Breheny et al., 2006; Chevallier et al., 2008; Moeschler, 2023; Sperber & Wilson, 1986, 1987). For example, Chevallier et al. (2008) reported a study in which participants judged the truth value of "or" statements (e.g., "There is an A or a B" in the word "TABLE"). If interpreted pragmatically, then "or" would take on an exclusive interpretation to mean "one but not both", which is more typical in everyday communications and optimally relevant to the context. Conversely, if interpreted logically, then "or" would take on the inclusive interpretation of "at least one". Compared to when the scalar word "or" in the statements was unstressed, participants were significantly more likely to interpret the scalar word pragmatically when it was stressed graphically ("OR") or prosodically (Chevallier et al., 2008; see also, Jang et al., 2013; Ryzhova & Demberg, 2023).

In the case of the SAR missions, it seems possible that the computer-based environment might have been insufficient to activate expectations about the intent of the implicatures. For example, in Experiments 1 and 2, while the word "might" or the phrase "might or might not" did not generate behavioural evidence of pragmatic processing (perhaps because it was not stressed), "could be outdated" was sufficient to do so. Despite containing an implicature itself, "could be outdated" likely functioned as a semantic cue to guide the interpretation of the overall message as cautioning against complete reliance on the floorplan. Similarly, the framing of "might be useful" as part of the instructions may have encouraged its interpretation as a (polite) directive to utilise the floorplan, rather than communicating uncertainty. This interpretation aligns with the observation that context determines whether probability expressions are understood as communicating uncertainty (Bonnefon et al., 2011; Bonnefon & Villejoubert, 2006; Brown & Levinson, 1987; Holtgraves, 2014). For example, Bonnefon and Villejoubert (2006) showed that the interpretation of the word "possibly" depended on the severity of the medical condition that it was linked to. In the case of the statement "The doctor tells you, you will possibly suffer from insomnia soon”, “possibly” was interpreted as signalling the doctor’s uncertainty. However, for a more severe condition (“deafness”), participants instead believed that the doctor was more certain than uncertain about the prognosis but was being tactful when delivering the bad news. Similarly, in collaborative settings or when giving directives, probability expressions are often used to reduce forcefulness to be polite and manage social face (Bonnefon et al., 2011).

The results of the present study suggest that pragmatic reasoning in simulated search and rescue missions may be constrained, and that communication processes that are effective and commonplace in everyday life might be less robust in high-stakes situations unless additional semantic cues were provided to guide pragmatic reasoning. They suggest the need to investigate further the nature and efficacy of communications in real missions, the results of which could inform policy, guidance, and training.

Another experiment examined whether implicit or explicit messaging hinting at the potential unreliability of an eyewitness influenced participants’ reliance on their report. As there were no systematic effects of messaging on reliance during search and rescue, this study is reported in Appendix F rather than in the main chapter to maintain narrative flow.

Chapter 3:

Visual Search in Semi-Immersive Virtual Reality Environment: Methodological Development

This chapter describes the development of the methodology used in Chapter 4 to investigate the influence of self-generated expectations of target prevalence and explicit instructions about changes in target prevalence during high-stress, high-stakes visual search. This extends the theme of the experiments described in Chapter 2, which examined the use of new information to support navigation and decision-making during emergency incidents. Experiment 4 involved participants assuming the role of firefighters tasked with removing explosive hazards (i.e., medium-width cylinders) amongst visually similar stimuli (i.e., thinner or wider cylinders) in a smoke-filled building. Briefly, the experiment assessed their reliance on self-generated and instructed expectations of target prevalence under stress within this search context. A virtual-reality (VR) platform was employed to enhance immersion and presence beyond that afforded by the desktop-based simulations used in previous experiments. The following sections describe the following key considerations: The selection of a semi-immersive VR environment as the most appropriate platform for the research conducted in Chapter 4 and the rationale for the multimodal stressors used in the experiment. I then report a feasibility study that evaluated the effectiveness of the stressors within the virtual search context, and conclude with a description of the Hazard Search Task featured in Chapter 4.

3.1 Choosing Between Immersive and Semi-Immersive Virtual Reality Systems

VR provides the opportunity to study the effects of stress during emergency operations in immersive, safe, highly controlled, and standardised environments. It is most widely implemented in head-mounted devices (HMDs) given their immersive audio-visual input, commercial availability, affordability and portability. However, a common side effect of

HMDs is cybersickness, which is characterised by disorientation-related symptoms (e.g., dizziness), oculomotor-symptoms (e.g., eye-strain), and to a lesser extent, nausea-related symptoms (Saredakis et al., 2020). These symptoms result from a temporal delay between the user's actual head movement and the display output in the HMD, otherwise known as display lag (Palmisano et al., 2020, 2022). The discrepancies between the user's physical and virtual head pose (DVP) lead to conflicting visual, vestibular, and other proprioceptive signals about their head position and orientation, which promotes the onset of cybersickness (Palmisano et al., 2020, 2022). Increases in DVP have been associated with greater perceived scene instability, reduced spatial presence, and greater cybersickness severity in the virtual environment (Palmisano et al., 2022). Indeed, rapid head movements worsened cybersickness by increasing display lag, which then exacerbated the visual consequence of DVP (Feng et al., 2019; Palmisano et al., 2022). Interestingly, even when display lag was not consciously perceptible, HMD users also reported cybersickness symptoms (Stauffert et al., 2018). Therefore, prolonged use of HMDs might pose logistical challenges during data collection due to the risk of cybersickness (Martirosov et al., 2022), particularly among those with increased susceptibility due to a history of motion sickness (Laessoe et al., 2023).

Unlike HMDs, which encompass the user's field of vision, semi-immersive VR systems induce less cybersickness (Martirosov et al., 2022). One example is the Cave Automatic Virtual Environment (CAVE) (Pan & Hamilton, 2018). In a typical CAVE, stereoscopic images are projected onto (at least) three walls surrounding the user. Like HMDs, the projections are continually updated based on the user's head position. Therefore, CAVEs could still induce cybersickness, but to a much lesser degree than HMDs (Kwok et al., 2018; Martirosov et al., 2022). While healthy users strongly preferred HMDs over CAVEs for brief tasks (e.g., under three minutes, Elor et al., 2020), HMDs are less suitable for lengthier experiments, particularly those involving navigation. In an experiment involving searching for red balls in a maze, a majority of the HMD group abandoned the task before the 10-minute mark due to severe cybersickness symptoms, whereas 83% of the CAVE group completed the task (Martirosov et al., 2022). As the study reported in Chapter 4 required at least 20 minutes of VR exposure, excluding breaks and calibration processes, semi-immersive VR was deemed the better option.

The Igloo Immersive Cylinder (hereafter called the Igloo) is a 6-metre diameter semi-immersive VR system from Igloo Vision Ltd. It is similar to CAVEs with some notable exceptions. The setup in the laboratory incorporated five Epson EH-LS500B projectors mounted on its ceiling to project 2D panoramic content onto a cylindrical screen (Figure 3.1A). Critically, the virtual environment in the Igloo is updated based on locomotor input from an omnidirectional motion platform rather than the user's head position. This eliminates head-rotation induced display errors that underlie cybersickness. The Igloo is less immersive than CAVEs as the floor receives no projection and hides part of the virtual environment from the user, which likely reduces immersion relative to a CAVE (Cruz-Neira et al., 1993). Despite these limitations, its relative affordability (i.e., 90% to 95% less costly), ease of use, and straightforward setup (IglooVision.com, 2020) make the use of the Igloo an attractive alternative to CAVEs.

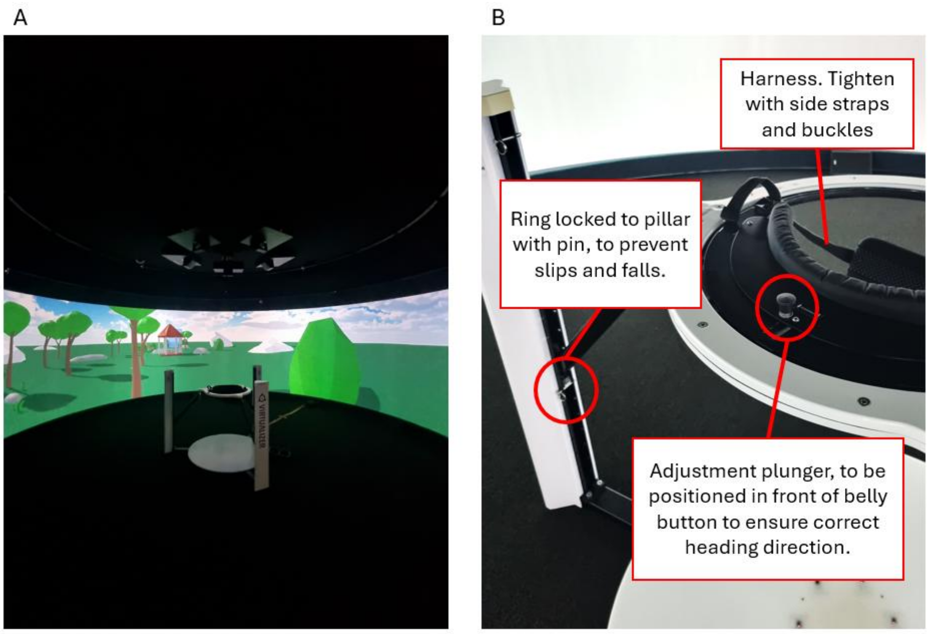


Figure 3.1. Interior of the Igloo Immersive Cylinder and the Cyberith Virtualiser R&D Kit.

3.1.1 The Igloo Environment Setup

Within the Igloo, a surround sound system supported immersive audio with five Cambridge Audio MINX MIN12 loudspeakers fixed to the circular frame inside and at the top of the cylinder, with a Cambridge Audio SX-120 subwoofer outside of the frame, and a Denon AVR-X550BT amplifier. Three SteamVR Base Station 2.0 mounted on the circular frame supported communication between the VIVE Pro HMD on the floor, the Steam VR software, a Vive Tracker 3.0, and a Steam Controller. Together with the Igloo Wrapper and Igloo Playback software, this setup allowed the projection of the virtual environment onto the panoramic screen and enabled interaction with it. Two 10-metre-long active USB extender hung down from the top of the cylinder, connecting a mechanically integrated emteqPRO mask and Pupil Core eye tracker to the main computer next to the Igloo. A portable Tripp Lite air-conditioning unit was installed for user comfort. The lightweight motion capture device, Vive Tracker 3.0, was fastened on top of the emteqPRO mask to control a virtual headtorch in the smoke-filled virtual environment. Users interacted with the virtual environment with the Steam Controller gamepad, with button mappings detailed in Figure 3.2.

Figure 3.2. Button mappings on Steam Controller gamepad.

To increase realism, locomotion in the virtual environment was controlled by an omnidirectional motion platform in the middle of the Igloo: The Cyberith Virtualiser R&D Kit. It had a low-friction, circular baseplate and a ring that secures the participant in place. Calibrating the platform involved turning the ring three times, lowering it, and raising it up to its original position. To use it, participants wore low-friction overshoes provided by Cyberith GmbH and got onto the platform when the ring was lowered from its usual position. The ring was then lifted to the participant's hips and locked with three pins on the pillars, preventing slips and falls. While the researcher tightened the harness, the user would be told to keep the adjustment plunger in front of their belly button to ensure correct heading (Figure 3.1B). Six optical motion sensors in the baseplate recorded movement speed at 1000Hz, while one in the ring tracked rotation, translating movement data into the virtual world for a near-natural walking experience. Movement speed could also be increased or decreased by adding a movement speed multiplier to the "CVirtPlayerController" script component (*Effective VR Treadmill - Cyberith Virtualizer R&D Kit*, n.d.).

User safety in the in Igloo was continually monitored with a GoPro Hero10 clamped onto the circular frame. The camera was configured to stream live video to a laptop computer using an RTMP server implemented with the NGINX Gryphon module (an open-source web server) and FFmpeg (a multimedia processing software) to receive and display the streamed video. Throughout the experiment, observable signs of distress such as freezing, trembling, or other clear physical indicators were monitored.

3.2 Stress Manipulation in VR

Standardised VR protocols have been shown to reliably induce stress in HMDs and CAVEs, namely the VR-Trier Social Stress Test (VR-TSST) (Helminen et al., 2021; Kirschbaum et al., 1993; Shiban et al., 2016) and the immersive multimodal virtual environment stress test (IMVEST) (Rodrigues et al., 2021). In the VR-TSST, participants prepared and delivered a speech to convince a virtual committee of their suitability for a job and completed an arithmetic task before the virtual committee (Helminen et al., 2021; Kirschbaum et al., 1993; Shiban et al., 2016). This procedure cannot be integrated with the experiment without being a separate component. IMVEST is a recently developed protocol using multimodal stressors to introduce socio-evaluative pressure, uncontrollability and unpredictability, audiovisual stressors, and cognitive workload (Rodrigues et al., 2021). A key stressor in IMVEST is the threat of falling, which threatened physical self-preservation. However, in the Igloo, the visual impact of exploding tiles and the resulting floor gaps could not be appreciated as the Igloo floor lacked projection. Moreover, as with the VR-TSST, adding an additional protocol would extend VR exposure by at least 10 minutes (Rodrigues et al., 2021), increasing the risk of cybersickness (Martirosov et al., 2022).

An obvious alternative to the above was to integrate contextually relevant stressors within the main experiment. In fully immersive HMDs, carefully selected stressors have proven to be effective in influencing physiological and psychological measures of stress compared to baseline. For example, firefighter trainees who assessed car crash victims and marked those needing medical attention in a virtual training scenario experienced significantly higher mental workload, time pressure, frustration, and heart rate than those who merely explored a similar virtual environment without a crash (Czarnek et al., 2020). These findings were unsurprising as the rescue task (i.e., stressor) was compared to a control scenario with no performance goal. In another study, participants searched for and extinguished a potential fire on a virtual International Space Station in low, medium, and high stress conditions before time ran out (Finseth et al., 2022). Stress was manipulated by varying the intensity of contextually relevant stressors, such as alarms, flickering lights, and smoke density. These stressors resulted in significant differences in most psychological and physiological measures of stress between low and high stress conditions (Finseth et al., 2022). Such effects were also observed in less immersive setups, such as the one used in Meng and Zhang's (2014) study, which mimicked the Igloo. Their system displayed the virtual environment on six 47-inch flat-screen monitors linked end-to-end and positioned at eye-level around the seated participant. Participants under stress located the exit in a virtual hotel while exposed to virtual fire and smoke, fire alarm, and "real smoke" from a generator, where the control group completed the task without these stressors. Compared to the control group, participants exposed to these events had significantly higher stress levels, heart rate, and skin conductance, and took longer to locate the exit (Meng & Zhang, 2014).

Overall, the appropriate combination of contextually relevant, multimodal stressors can induce stress responses in virtual emergency scenarios in different VR systems. To examine the feasibility of using such stressors in the large-scale study reported in Chapter 4, the following stressors were evaluated:

(i) The risk of an explosion during the trial that would prematurely end it;

(ii) A loud alarm that beeped in a loop throughout a trial;

(iii) A reduction of movement speed by approximately 40% relative to the low stress condition.

The explosion occurred randomly for half the High Stress trials to introduce time pressure to complete the trials quickly. Additionally, the random nature of the explosions removed the participants' ability to control or predict their occurrence, which generates stress (Mineka & Hendersen, 1985; Rodrigues et al., 2021). The loud alarm provided unpleasant auditory stimulation and served as a reminder of the potential explosion. Compared to quieter sounds, exposure to such sounds can increase perceived workload, annoyance, cortisol levels, and reduced energy. Performance on mental tasks requiring concentration also suffer as a result of such stimuli (Radun et al., 2022). To identify an alarm sound that was sufficiently unpleasant, stress-inducing, but not overwhelming, a small group of volunteers evaluated candidate sounds in the Igloo and selected one that continuously looped three loud beeps followed by a brief pause. Lastly, reducing movement speed should impede task completion and increase frustration, which approximated to disabling the inputs in the affective Pacman game (Reuderink et al., 2009). Similarly, in a 3D puzzle game designed to induce stress within the Generic Automatic Stress Induction and Control Application, intermittent movement restriction was used to induce frustration (van der Vijgh et al., 2014), and has been shown to increase stress levels (Edwards & Kelly, 2017). To maintain realism, movement speed was reduced by 40% instead to achieve an effect similar to intermittent movement restriction. Visual stressors, such as flames, were excluded to avoid interference with visual search.

3.2.1 Physiological Measures of Stress

Stress is often measured with physiological markers such as heart rate (HR), heart rate variability (HRV), and cortisol levels (Dammen et al., 2022). However, increases in these measures might not always reflect the negative affective states associated with stress responses (Lazarus, 1993). In fact, Campbell and Ehlert (2012) reported that only 25% of the studies reviewed observed an association between biological markers and subjective appraisals of stress. Facial electromyography (fEMG) could capture the affective component of stress by detecting facial muscle activation patterns associated with different facial expressions. As negative affective states emerge in response to obstructed goal attainment (Lazarus, 1993), the valence of facial expressions might indicate stress levels. The corrugator muscles which sit between the brows draw them together during frowning, while the bilateral zygomaticus muscles, which run from the cheeks to the corners of the mouth, pull them upwards when smiling. Both are associated with negative and positive affective states, respectively, during passive exposure to affective stimuli (Dimberg, 1990; Lang et al., 1993; Larsen et al., 2003). Greater corrugator activation is linked with negative valence (Lang et al., 1993), such as when experiencing a virtual fall threat scenario (Baker et al., 2020), during setbacks in racing games (Hazlett, 2006), when viewing affective photos (Mayo & Heilig, 2019), and when being evaluated by others (Kroll et al., 2021). Zygomaticus activation, however, follows a J-shaped profile, whereby extremely aversive stimuli (e.g., vomit) activated the muscles slightly more than middling stimuli, but less than positively valenced stimuli (Lang et al., 1993).

For the study reported in Chapter 4, fEMG has an advantage over other physiological measures of stress such as HR, HRV, and electrodermal activity. Participants controlled locomotion in the virtual environment by walking on the motion platform, making physical exertion inevitable. Electrodermal activity is sensitive to mild physical activity like walking on a treadmill (Posada-Quintero et al., 2018). While HR and perceived exertion increased linearly with different exercise intensities, corrugator muscle activity increased significantly only during severe-intensity exercises (de Morree & Marcora, 2012; Huang et al., 2014), making it less susceptible to the influence of low- to moderate-intensity physical exertion.

3.2.1.1 EmteqPRO Open Face Mask. The emteqPRO Open Face mask collects fEMG data at a sampling rate of 2000Hz (Gnacek et al., 2022). It has seven pairs of dry EMG electrodes and a photophlethysmogram (PPG) sensor embedded into a foam layer which records facial muscle activity from the corrugator, frontalis, orbicularis (henceforth referred to "orbicularis"), and zygomaticus muscles. The side inserts with the orbicularis and zygomaticus electrodes are adjustable to improve mask fit and sensor signal quality for different face widths (Gnacek et al., 2022). This can be monitored live in the Data Viewer component of the EmteqVRManager within the EmteqVR SDK (emteq labs, 2022), visible under the "Game View" of the Hazard Search Task in the Unity software. For each electrode, good sensor signal quality depends on good contact with the skin, indicated by a green sensor icon on the Data Viewer (see Figure 3.3); if contact is poor, the sensor icon appears grey (emteq labs, 2022).

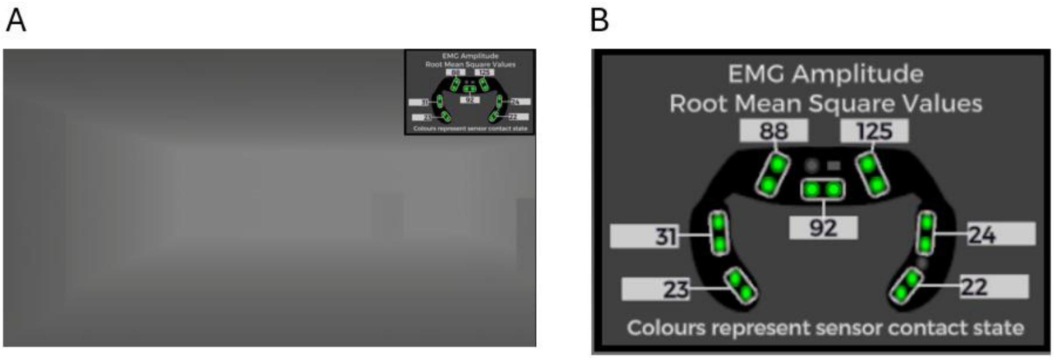


Figure 3.3. (A) "Game View" of the Hazard Search Task in Unity displayed on the main computer outside the Igloo, visible only to the researcher during the experiment. (B) Close-up of the Data Viewer, showing good sensor contact state for all electrodes (green icons) across all muscle groups. Numbers indicate the root mean square EMG amplitudes.

Each recording produces a JSON file containing timestamps and event labels and a DAB file containing sensor data, which were uploaded to the SuperVision App for further processing to generate a CSV file with readable data. Other data measures can also be computed by the app, but only "facial valence scores" were of interest here. These scores can be computed based on the activity ratio of the corrugator and zygomaticus muscles, and they were updated every 500ms and ranged from -1 (negative valence) to 1 (positive valence).

Calibration. For each participant, the mask has to be calibrated, with the first phase involving deep breathing for two minutes to obtain baseline HR and HRV. After that, at the researcher's instruction, participants alternated between a neutral expression and three maximum expressions (smile, frown, raised eyebrows), returning to neutral between each expression. Pilot testing of the mask on the researcher in the Igloo revealed that PPG signal quality fluctuated wildly, prompting a more systematic investigation into its reliability. Three sets of recordings were collected in the Igloo, each including different stages of movement: staying idle (between movements), walking only, walking while looking around, walking in circles without looking around, and walking in circles while looking around. Consistently, between 84.34% and 100% of the signal quality was above 80% when staying idle or when walking only, whereas only between 49.30% and 64.18% was above the criterion (i.e., signal quality above 80%) when walking in circles while looking around. This indicated that a separate HR monitor was required, because participants would be moving frequently during the main study, unreliable PPG signal quality would yield invalid HR or HRV data.

3.2.1.2 Polar OH1+. The PolarOH1+ is a 6-LED, arm-worn PPG HR monitor. There is a high level of agreement between its HR measure and those produced by electrocardiogram-derived HR, with intraclass correlation between 0.95-0.99 when worn on the upper arm, lower arm, or temple (Hettiarachchi et al., 2019). HR data updates every second and can be monitored live in the Polar Flow app on an Android tablet. Three volunteers, including the researcher, assessed the responsiveness of the armband to ensure that the data synchronised accurately with the user's physiological state. Each volunteer performed a simple test three times: they would jump on the spot for approximately 30 seconds while wearing the armband to increase their heart rate and stood still afterward. Immediately after they stopped jumping, an observer recorded the time taken for the HR reading to reach its peak and begin decreasing. A 17-second lag was consistently observed across the nine sets of data, which was accounted for during data analysis.

3.2.1.3 Pupil Core. While not a measure of stress, eye-tracking data could reveal changes in visual search strategy under stress. The Pupil Core headset (Kassner et al., 2014) includes a front-facing world camera and two 200Hz infrared eye cameras oriented towards the eyes which support monocular or binocular eye-tracking. Its lightweight, modular hardware design enabled mechanical integration with the emteqPRO mask (carried out by Emteq Labs), allowing simultaneous collection of eye-tracking and fEMG data.

Prior to calibration, the eye cameras were adjusted to capture clear images of the participants' eyes. To ensure robust pupil detection (and thus, useable data), participants were asked to move their eyes around until the blue circle in each eye window (3D model of the eye) fit the eyeball and the red circle with the red dot aligned with the pupil (see Figure 3.4) in the Pupil Capture software (Pupil Labs, n.d.). Further adjustments to the parameters in the Pupil Capture software were made when the pupil detection confidence was below 0.60. After that, calibration of the eye-tracker was conducted in the Igloo to ensure that viewing distance and conditions were consistent during calibration and data collection (Pupil Labs, n.d.).

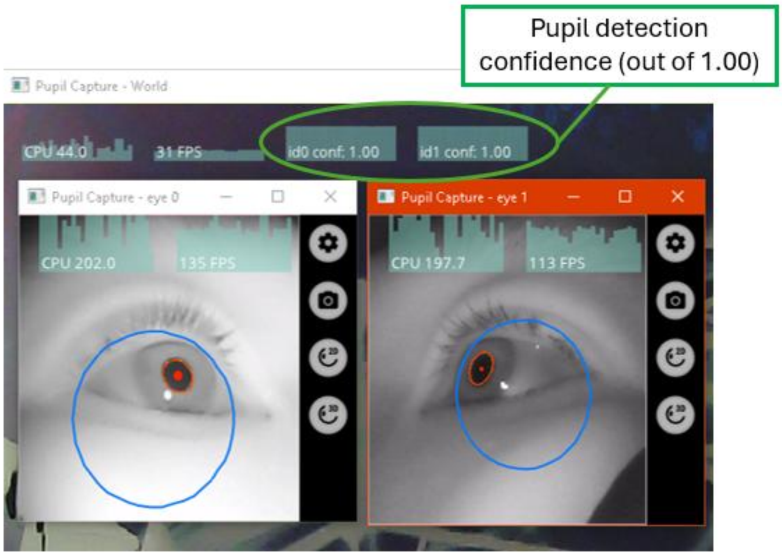


Figure 3.4. Screenshot from the Pupil Capture software showing robust pupil detection (red circle with red dot on the pupils), accurate 3D eye model mapping (blue circle), and high pupil detection confidence for both eyes (green outline).

Calibration. In the Igloo, the researcher adjusted the world camera to capture the participants' field of vision. They were then asked to orient their body towards an orange rectangle on the screen and reminded not to move their head or body during the calibration process. As the researcher explained the process, the instructions were also displayed on the screen (see Appendix G for verbatim instructions). The participant was asked to roll their heads a few times and assume a natural head position, after which the researcher adjusted the focus and position of the world camera to align with the participant's line of sight. This prevented participants from moving their body during calibration due to unnatural posture. When ready, they pressed the "Big Left Button" on the Steam Controller to start recording. To stabilise the eye model in the Pupil Capture software, the participant rolled their eyes for four seconds (Pupil Labs, n.d.), stopping when they heard "stop", and prepared for calibration. Six bullseyes-like markers appeared in sequence – centre, top right, top left, bottom left, bottom right, and back to the centre of the screen – each visible for two seconds before disappearing. The participant looked at each marker until it disappeared. At the end of the calibration, they could press the left trigger on the Steam Controller to end the process if they judged that they made no mistakes during calibration, or the right trigger to repeat the calibration if they had not followed the calibrations instructions correctly.

3.3 Feasibility Study: Evaluating the Stressors

To ensure that the combination of stressors described earlier could elicit an appreciable difference in stress responses within the Igloo environment, a feasibility study was conducted before Experiment 4. A secondary aim was to assess whether participants could comfortably complete the experiment, considering its lengthy duration and the requirement to wear the emteqPRO mask throughout. Due to the effortful nature of visual search tasks (Anderson, 2024; Anderson & Lee, 2023; Attar et al., 2016) and its likely interaction with physiological measures of stress, the hazard search task was replaced with a simple search and rescue (SAR) task, where participants searched for a trapped person in a smaller building with the same hexagonal layout.

Compared to control conditions, greater corrugator activity (Baker et al., 2020; Hazlett, 2006) and increased heart rate (HR) (Finseth et al., 2018; Jönsson et al., 2010; Rodrigues et al., 2021) have been observed during stressful tasks. Moreover, given that movement speed was artificially reduced under High Stress, participants would have to move more effortfully than under Low Stress, potentially contributing to negative affective states associated with perceived inability to cope with task demands (Lazarus, 1993). Therefore, under High Stress (and compared to Low Stress conditions), the following outcomes were hypothesised:

(i) Significantly higher perceived difficulty to cope ratings

(ii) Significantly greater negative affect, indicated by:

   a. Increased corrugator muscle activity

   b. Decreased zygomaticus muscle activity

   c. Lower facial valence scores

(iii) Significantly higher HR data

3.3.1 Participants

Sixteen participants (13 females, 3 males) with a mean age of 20.02 years (range = 18.67-22.08 years) were recruited from the student population of Cardiff University. Participants had normal or corrected-to-normal vision, did not wear prescription glasses, had no adverse experience with fire personally or through loved ones, had not experienced motion sickness in the past three months, and were not currently ill, taking psychotropic medication, suffering from hypertension, or diagnosed with mood/anxiety disorders. Approval for this research was granted by the School of Psychology Research and Ethics Committee (EC.22.04.26.6560GRA).

3.3.2 Materials

The experiment took place in the Igloo setup described earlier, complete with the surround sound system, live monitoring of the participant via the GoPro Hero10 camera, and naturalistic locomotion on the Cyberith R&D Kit omnidirectional platform. All audio was played at 70% of the maximum amplifier volume. The emteqPRO mask and PolarOH1+ armband was used to collect fEMG and HR data, respectively. The Vive Tracker 3.0 was attached to the top of the mask to control a virtual headtorch during the experiment, and the Steam Controller gamepad was used to interact with the virtual environment.

3.3.3 The Search and Rescue Task

I built the feasibility study with Unity 2020.3.6f1 and the support of the Unity Experiment Framework (UXF; Brookes et al., 2020). Other Software Development Kits (SDKs) and plugins such as the EmteqVR SDK (emteq labs, 2022), Igloo Toolkit (Selly, 2022), OpenVR XR Plugin (*SteamVR Unity Plugin*, 2017/2023), and Cyberith Virtualiser SDK (Cyberith, 2020) were integrated to support the functionality of their associated equipment. The smoke intensity in the virtual building was decided in consultation with a retired firefighter, with over 30 years of experience, to simulate low-visibility conditions realistically during fire emergencies in a way that was still conducive for visual search. The building layout for this study was based on the initial design for the Hazard Search Task, the rationale behind which is presented in Section 3.4.1. From the participants' perspective, the smoke-filled building had six rooms in a circular arrangement (see Figure 3.5), with their inner walls forming a central hexagonal space where each trial began. Each room contained an explosive hazard, and the trapped person was in one of these rooms.

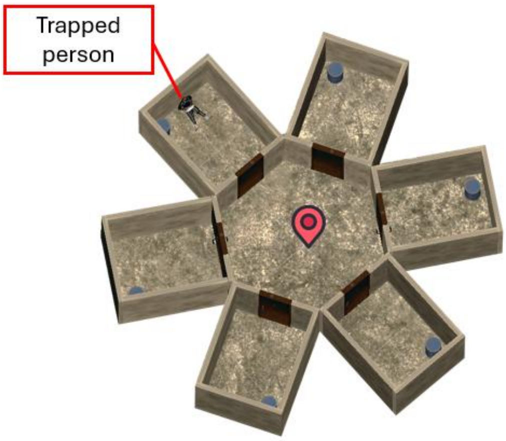


Figure 3.5. Building layout for the feasibility study, showing six rooms arranged in a circle, each containing an explosive hazard, with a trapped person located in one of the rooms. The red marked indicates the starting point in each trial.

Each trial began at the same starting point marked in Figure 3.5. The participant initiated each trial by responding to the same prompt: “Press a [small smooth button] to begin.”. To create the illusion that the trapped person was in different rooms each trial, the participant’s orientation was randomised at the beginning of every trial. To locate and rescue the trapped person, the participant inspected each room by moving close to the doors to open them automatically. While they were informed that the cylinders were explosive hazards, they were not instructed to remove them. Thus, if the trapped person was not in a room, they should quickly move on to the next room. A trial ended either when the trapped person was found and rescued, regardless of the number of rooms explored, or when the 30-second time limit was reached. If the person was rescued, a feedback message was displayed: “Congratulations! You’ve rescued the victim!”. Otherwise, no feedback was given and the prompt to begin the next trial appeared. At the end of each block, the participant supplied a rating to the question, “How were you coping with the search and rescue tasks just now?”, with 1 representing “I felt no pressure” and 10 representing “I was unable to cope with the pressure.”. After submitting their rating, they received their performance feedback: "You have rescued *x* out of 6 people who were trapped in the building.", with *x* being the number of successful rescues.

3.3.3.1 High Stress and Low Stress Conditions. In the High Stress condition, the temperature reading was red and fluctuated between 90°C – 95°C. Participants were informed that this, along with an ongoing alarm sound, signalled the risk of the cylinder in each room exploding mid-trial which would end the trial prematurely. The explosion was programmed to occur 10±2.5 seconds into a trial if it occurred. Unbeknownst to them, movement speed was approximately 40% slower than that in the Low Stress condition. This condition was quiet, and the temperature reading was green and fluctuated between 60°C – 65°C, signalling low risk of explosion.

3.3.4 Procedures

Stress level was manipulated during the SAR task within-subjects (Low Stress; High Stress). The experiment was run in two phases over two days, with a mean of three days apart (SD = 2.00). Day One was for acclimation to the VR environment and learning the game mechanics, and Day Two was for the formal experiment and data collection.

3.3.4.1 Day One: Acclimation and Training. The participant wore the PolarOH1+ on their bicep while the researcher fitted the emteqPRO mask and the Vive Tracker on them to familiarise them with wearing the equipment during the experiment. After entering the igloo, they wore the low-friction overshoes. Safety instructions were provided before they were assisted onto the motion platform, secured in place, and taught to walk using video demonstrations and verbal instructions. They were given up to 20 minutes to practice walking towards various objects in a virtual park environment until:

(i) They rated at least 7 to "On a scale of 1 to 10, with 1 being completely unnatural and 10 being almost as natural as walking in real life, how would you rate your experience now?"

(ii) They reported "yes" to "If you were given mental arithmetic tasks to do now while walking on this platform, do you think you will be able to do them without having to think about walking?"

(iii) The researcher judged objects in the virtual environment as moving smoothly when the participant was walking.

They were then trained to use the Steam Controller gamepad and learned the button names used in the experiment for interacting with in-game elements (see Figure 3.2). After that, a three-minute calibration process for the mask was carried out. However, the fEMG data for Day One was not analysed.

The training phase for the feasibility study was similar to the SAR task with some exceptions: Each block had three untimed trials instead of six timed trials, and participants did not rate their perceived difficulty to cope at the end of a block. Participants were briefed that they would be acting as a firefighter tasked with searching for and rescuing a person trapped in the virtual building. There were four blocks of trials: The first and third blocks were Low Stress blocks, whereas the second and fourth were High Stress blocks. Participants were informed that an ongoing alarm and high room temperature reading in a red font signalled that the cylinder in each room might explode and end the trial prematurely (i.e., during High Stress blocks), so they had to rescue the person quickly. Otherwise, the likelihood of explosions was very low (i.e., during Low Stress blocks). Between each block, they could take a short break if needed. For this training, explosions were not programmed to occur during High Stress trials but triggered manually in the very last trial to assess participants' comfort level with the level of stimulation. After that, participants left the laboratory for the day.

3.3.4.2 Day Two: Search and Rescue Task. At the main computer, next to the Igloo, participants were fitted with the same equipment. After ensuring good skin contact for the fEMG electrodes, the mask was disconnected from the main computer, and the participant was led into the Igloo to practice walking on the motion platform until they were comfortable with the process. The mask was reconnected via the active USB extender cables for calibration and data collection. The formal experiment had two blocks with six 30-second trials each. Half of the participants first completed the SAR task under High and then Low Stress conditions, and the remaining half first completed it under Low and then High Stress conditions. In the High Stress block, three random trials were set as explosion trials.

Participants were again briefed about the task and reminded of the features for Low and High Stress conditions. They were also informed that the researcher would only enter the Igloo during breaks to check on them and make adjustments (e.g., retighten motion platform harness and disentangle cables). They began the experiment once researcher exited the Igloo and signalled to start. After the first block of trials, participants took a mandatory three-minute break, during which a countdown circle appeared against a background of nature scenery and sounds to promote relaxation. The break could be extended upon request. When the experiment concluded, participants were debriefed and thanked.

3.3.4.2.1 Data Preparation. One participant encountered technical malfunction between the second and third trial of the High Stress block. As they stood idle when the problem was being resolved, the intervening fEMG and HR data were removed. The CSV file downloaded from the SuperVision App for each participant was processed using a Python script adapted from the "One-user Analysis Script 1" sample provided by Emteq Labs (emteq labs & Mavridou, 2025) to extract mean amplitudes (μV) for each trial block. The amplitudes were not normalised since comparisons of the same muscle groups were made within-subjects, with the same electrode placements, and on the same day (Halaki et al., 2012). A modified JSON file that excluded 5-second periods of explosions was used in the script to reduce bias from abrupt facial expressions. Values from bilateral muscle groups (frontalis, orbicularis, zygomaticus) were then averaged to produce one value per muscle group. To obtain mean HR (bpm) for each block the CSV file produced by the PolarOH1+ was subjected to another Python script. The 17-second lag observed in prior testing was corrected by shifting the trial block timestamps by 17 seconds. Likewise, the 5-second exclusion for each explosion was applied before calculating mean HR per block.

All statistical tests were performed using R, version 4.4.0 (R Core Team, 2024). Where the assumptions required for the use of parametric tests were violated, non-parametric alternatives were conducted. Bayes factors (BF_{10}) using default priors were also computed using the BayesFactors R package (Morey et al., 2024). For skewed data, log-transformed data were used if it reduced skewness; otherwise, raw data were analysed. BF_{10} values greater than one favour the alternative hypothesis, whereas values less than one favour the null hypothesis. Evidence categories follow Andraszewicz et al. (2015).

3.3.5 Results

All participants completed the feasibility study without reporting symptoms of cybersickness or general discomfort. Bonferroni correction at $\alpha = .017$ was applied to tests on corrugator and zygomaticus activity, and facial valence scores as they jointly addressed a single hypothesis (Rubin, 2021). Descriptive statistics for fEMG data were converted from volts (V) to microvolts (μV).

Perceived difficulty to cope was greater under High Stress (median = 5.50, IQR = 2.25) than Low Stress conditions (median = 3.00, IQR = 2.00). This difference was significant according to a Wilcoxon signed-rank test, $V = 136$, $p < .001$, $r = 0.89$, with Bayes factor indicating extreme evidence supporting the first hypothesis ($BF_{10} = 5.47 \times 10^4$). Median corrugator activity was higher under High Stress (median = 3.45, IQR = 3.44) than Low Stress conditions (median = 2.40, IQR = 2.54); zygomaticus activity was also higher under High Stress (median = 8.70, IQR = 9.58) than Low Stress conditions (median = 3.58, IQR = 3.67); and facial valence scores in both Stress conditions were not only positive, but also higher under High Stress (median = .01, IQ = .04) than Low Stress conditions (median = .00, IQR = .01). Wilcoxon signed-rank tests revealed that the differences were significant for corrugator muscles ($V = 116.00$, $p = .011$, $r = .62$) and zygomaticus muscles ($V = 127.00$, $p = .001$, $r = .76$), with moderate and very strong evidence in favour of the difference, respectively (corrugator: $BF_{10} = 3.15$; zygomaticus: $BF_{10} = 31.79$). However, the difference in facial valence scores was not significant at $\alpha = .017$, $V = 112.00$, $p = .021$, $r = 0.57$, with anecdotal evidence supporting the null ($BF_{10} = 0.66$). The second hypothesis was partially supported by corrugator activity data. However, the zygomaticus activity and facial valence measures were in the opposite direction to what was hypothesised, possibly due to a grimacing expression in response to negative affect. HR was higher under High Stress conditions (mean = 120.16, SE = 2.22) compared to Low Stress conditions (mean = 119.90, SE = 2.38). This difference, however, was not significant, $t(15) = 1.77$, p = .097, Cohen's D = 0.44, with anecdotal evidence in support of the null for the third hypothesis ($BF_{10} = 0.91$).

3.3.5.1 Additional Analyses: Frontalis and Orbicularis Muscles. Considering the unexpected zygomaticus activation patterns, additional analyses were conducted for the frontalis and orbicularis muscles. Greater frontalis and orbicularis activity were observed under High Stress conditions (frontalis muscles, median = 5.01, IQR = 9.55; orbicularis muscles, mean = 6.37, SE = 0.70) than Low Stress conditions (frontalis muscles, median = 4.09, IQR = 5.34; orbicularis muscles, mean = 4.06, SE = 0.47). Wilcoxon signed-rank tests confirmed that the differences were significant for frontalis activity, $V = 108.00$, $p = .039$, $r = 0.52$, and a paired t-test confirmed the same for orbicularis activity, $t(15) = 4.26$, $p < .001$, Cohen's D = 1.06. Bayes factors indicated anecdotal and very strong evidence for the difference, respectively (frontalis: $BF_{10} = 2.83$; orbicularis: $BF_{10} = 52.63$).

3.3.5.2 Order Effects. A 2(Stress) x 2(Sequence: Low-High, High-Low) mixed ANOVA was conducted to investigate potential order effects. A robust alternative (Maechler et al., 2006; Mair & Wilcox, 2020) was conducted when parametric assumptions were violated. Normalised fEMG data were used for valid between-group comparisons (Halaki et al., 2012). The normalised data were provided by the SuperVision App, with muscle activation levels expressed as a percentage of maximum activation recorded during mask calibration. The absence of significant interaction effects for self-reported stress ($F(1, 7.00) = 0.02$, $p = .894$), normalised corrugator muscle activation ($F(1, 9.82) = 0.13$, $p = .725$), normalised zygomaticus muscle activation ($F(1, 7.99) = 0.18$, $p = .685$), normalised orbicularis muscle activation ($F(1, 14) = 1.87$, $p = .193$), normalised frontalis muscle activation ($F(1, 8.55) = 0.20$, $p = .668$), facial valence scores ($F(1, 7.27) = 0.00$, $p = .996$), and heart rate ($F(1, 14) = 2.63$, $p = .127$) suggested a lack of order effects.

3.3.6 Discussion

This feasibility study investigated the joint effectiveness of a loud, beeping alarm that looped throughout the experiment, the risk of trial-ending explosions, and artificially reduced movement speed at eliciting stress responses within a firefighting context in the semi-immersive Igloo environment. Physiological and self-report data were collected as participants searched for a person trapped a building under these stressors (High Stress) or without them

(Low Stress). As hypothesised, perceived difficulty to cope was significantly higher under High Stress than Low Stress conditions, with extreme evidence supporting this effect. The second hypothesis, which predicted greater negative affect under High Stress, was partially supported. Corrugator muscles activity was significantly higher under High than Low Stress, with moderate evidence supporting this effect. However, zygomaticus activity and facial valence scores were in the opposite direction as hypothesised, with very strong evidence for the former and anecdotal evidence supporting the null for the latter. Lastly, the third hypothesis was not supported as HR data did not differ significantly between Stress conditions, with anecdotal evidence for the null.

The combination of time pressure (Rodrigues et al., 2021), unpredictability and uncontrollability (Mineka & Hendersen, 1985) induced by explosions, the loud alarm, and reduced movement speed (Reuderink et al., 2009; van der Vijgh et al., 2014) successfully elicited stress responses, possibly exacerbated by interfering with the goal of rescuing the trapped person within the 30-second time limit (Lazarus, 1993). The observed increase in corrugator activity under High Stress indicated higher levels of negative affect in response to the stressors (Lang et al., 1993; Larsen et al., 2003). Greater physical exertion to compensate for reduced movement speed in the High Stress condition was unlikely to explain this increase. The present task was unlikely to reach severe-intensity levels, considering that each block required 180 seconds of movement, yet the mean HR under High Stress conditions was only around 120bpm. This was substantially lower than the average HR of approximately 160bpm recorded after 180 seconds of heavy-intensity exercise, which was still insufficient to increase corrugator activity (de Morree & Marcora, 2012). Instead, increased mental effort (Cohen et al., 1992; Van Boxtel & Jessurun, 1993) and negative valence likely contributed to this increase. Suppression of the loud alarm required cognitive control, and the associated cognitive effort might have contributed to greater corrugator muscle activity (Berger et al., 2020).

Interestingly, zygomaticus activity levels and facial valence scores were significantly higher under High than Low Stress conditions. Contraction of the zygomaticus muscles are typically associated with positive affect (Lang et al., 1993; Larsen et al., 2003). Coupled with higher facial valence scores, this seems to suggest greater levels of positive affect under High Stress. This interpretation is incompatible with past literature and conventional assumptions about affective responses to stress, especially when considering the feedback participants provided informally post-experiment. While the participants found the overall VR experience to be novel, enjoyable, and engaging, most confirmed that the loud alarm and threat of explosions were especially frustrating and stressful. Therefore, the greater activation of zygomaticus muscles under High Stress conditions must be interpreted within this broader context.

Beyond smiling, other facial expressions that pull the lips laterally also activate the zygomaticus muscles. In fact, their activation profile follows a J-curve whereby highly unpleasant stimuli (e.g., images of mutilated bodies) could also trigger their activation, producing a grimace (Lang et al., 1993; Larsen et al., 2003). For example, when participants experienced the risk of a 200-metre fall in the virtual world, greater zygomaticus activity was observed, potentially indicating grimacing in response to the negative affect induced (Baker et al., 2020). Moreover, elevated levels of activation of the orbicularis and frontalis muscles under High Stress lent support to the negative affect interpretation. The orbicularis muscles surround the eye sockets and close the eyelids when activated. Its coactivation with the zygomaticus muscles reflects a "Duchenne smile", which is genuinely felt and expressed during enjoyment (Ekman et al., 1990; Wolf et al., 2005). Its coactivation with the corrugator muscles, however, is associated with negative affective states such as disgust (Wolf et al., 2005). The frontalis muscles on the forehead raise the eyebrows when contracted to create a "surprise" expression, which Ekman (1992) proposed helps widen the eyes to better perceive unexpected stimuli. Its coactivation with the corrugator muscles accompanied increased concentration or mental effort, such as during challenging auditory discriminations, where widening the eyes might reflect an attempt at allowing greater sensory stimulation (Cohen et al., 1992). Furthermore, increased frontalis tension could also index stress directly during stressful as opposed to relaxing mental imagery (Passchier & Helm-Hylkema, 1981).

The overall picture, therefore, is consistent with findings that faces become more expressive under stress and high workload (Dinges et al., 2005). As zygomaticus activity has lower specificity at indexing valence than corrugator activity (Lang et al., 1993; Larsen et al., 2003), further insights from orbicularis and frontalis activity provided additional support to suggest that the present stress manipulations successfully elicited stress responses, including negative effect. Likewise, the facial valence score, which was computed as a ratio of zygomaticus and corrugator activation levels, should be interpreted with caution and contextualised. Its development was based on straightforward, affective stimuli designed to evoke specific emotional responses (I. Mavridou, personal communication, February 23, 2023), and might be oversimplistic when applied to tasks that are more complex and demanding than passive exposure to valenced stimuli.

Lastly, the lack of notable differences in HR between the stress conditions might be taken to suggest that any compensatory physical effort due to reduced movement speed under High Stress did not significantly elevate HR relative to Low Stress conditions. Additionally, effects of stress (unrelated to physical exertion) on HR might be overshadowed by the effects of physical demands. While stress could increase HR, it occurred in the context of brief physical activity such as walking short distances (Finseth et al., 2018) or when remaining stationary (Jönsson et al., 2010; Rodrigues et al., 2021). Participants in the present study were subjected to continuous physical activity (i.e., walking) on the omnidirectional platform, which likely influenced HR more significantly and masked the effects of stress alone.

The present study demonstrated that contextually-relevant, multimodal stressors effectively manipulated stress levels within a virtual firefighting context. Specifically, an alarm that emitted three beeps followed by a brief pause on a continuous loop, the risk of explosions that ended trials prematurely which introduced additional time pressure, and an artificial 40% reduction in movement speed were effective at inducing stress that could be reliably detected using both self-report and fEMG data. As these stressors were embedded into the main experimental task, the duration of the experiment (and thus, exposure to VR) could be minimised. This contrasts with standardised stress induction VR protocols like the VR-TSST (Helminen et al., 2021; Kirschbaum et al., 1993; Shiban et al., 2016) or the IMVEST (Rodrigues et al., 2021), which would require additional VR exposure, potentially increasing the risk of cybersickness (Martirosov et al., 2022). The fEMG findings also contribute to a growing body of evidence showing that zygomaticus activity is not solely associated with positive valence (Lang et al., 1993; Larsen et al., 2003), particularly during complex tasks that involved more than passive exposure to stimuli (Baker et al., 2020). Thus, interpretations of zygomaticus activity should be contextualised in relation to task demands and complemented by analyses of other facial muscle groups. While corrugator activity indexes negative affect and stress, it is also sensitive to mental load unrelated to stress manipulations (Berger et al., 2020). This sensitivity might extend to other facial muscles, such as the zygomaticus, which could be activated when a person frowns and grimaces during a difficult reasoning task. In this regard, fEMG data might be confounded by cognitively demanding tasks, such as those involving conscious effort or problem-solving.

Considering this line of evidence, the aforementioned stressors were used in the large-scale study in Chapter 4 to manipulate stress levels within a virtual firefighting context. The effectiveness of stress manipulation in Chapter 4 was also assessed with self-report data, because it was not known whether their effects on fEMG (from corrugator, zygomaticus, orbicularis, and frontalis muscles) and HR data would be obscured by the inherent effortfulness of visual search (Anderson, 2024; Anderson & Lee, 2023; Attar et al., 2016) and physical activity. Facial valence scores were not analysed in Chapter 4 given their ambiguity, and their unsuitability for the present purpose. The following section outlines the design of the Hazard Search Task featured in Chapter 4, as well as the rationale behind key design decisions.

3.4 Visual Search in VR: Hazard Search Task Design and Development

The visual search task in the smoke-filled building involved participants acting as firefighters who identified and removed target cylinders, which were explosive hazards, amongst similar-looking distractor cylinders in the first five of six rooms explored. The trial concluded once the trapped person in the sixth room was rescued, who was included to raise the stakes and provide a rationale for the search task. I built the experiment with Unity 2020.3.6f1 and the support of the UXF (Brookes et al., 2020) package, alongside the integrated SDKs and plugins described in the feasibility study.

3.4.1 Building and Cylinders

The same level of smoke density was used in this task to create a low-visibility environment. As mentioned, the building in the feasibility study was based on the initial design of the building for the Hazard Search Task, where six rooms were arranged in a hexagonal formation around the participant (see Figure 3.6). This layout, whilst unconventional, offered several advantages over layouts of typical real-world buildings. Compared to the office building layout used in Experiments 1-3, the hexagonal layout ensured that the distance between each room and the starting point was equal. It also reduced the likelihood of excessive exploration by streamlining navigation and minimising variability in travel distances between rooms.

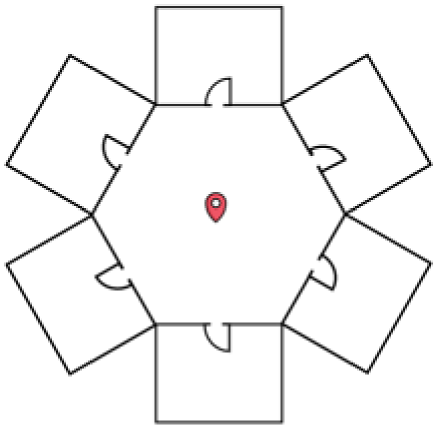


Figure 3.6. Ariel view of initial building layout. The red marker indicated the starting point in each trial.

An alternative layout where the rooms were sequentially arranged with their doors opening onto a shared corridor was considered. If the starting point was at one end of the corridor and the trapped person was always in the last room, the priority to rescue the person would conflict with the instructions: One should immediately rescue the trapped person from the last room and exit quickly, rather than first searching the five rooms along the corridor for explosive hazards as instructed. Thus, the counterintuitive instructions might reduce immersion. While the seemingly obvious solution to this problem is to vary the trapped person's location (i.e., to appear in different rooms each trial), doing so would require participants to remember where the trapped person was if they found him before they finished searching all the rooms. In addition to increasing cognitive load, travel distances between the last room and the room with the trapped person would be variable across trials, creating inconsistent time pressure to complete each trial in 2.5 minutes. For example, if the trapped person was in the second room from the starting point, the participant would continue searching for the target cylinders in rooms three to six before returning to room two again to rescue the person. In another trial, the trapped person might appear in the fifth room, so the participant only had to search the sixth room before returning to rescue the trapped person. Time pressure would differ between these trials because the distance and time needed to return to the trapped person would vary. Moreover, if the trapped person was found early in a trial, participants must remember their location while continuing the search, whereas if found later less backtracking would be needed. Such inconsistencies could inadvertently vary task demands within the same block of trials. Therefore, despite the trade-off with ecological validity, the hexagonal layout ensured minimal navigation variability, similar travel distances, and reduced non-search-related cognitive load compared to a corridor-based design.

Each of the first five rooms visited had five cylinders arranged in an arc formation (Figure 3.7C), and the trapped person always appeared in the remaining room. Compared to a haphazard arrangement, spacing the cylinders apart in a standardised manner reduced unintended variability in spatial layout that could create visual clutter and modulate visual search, such as when cylinders occluded one another (Bennett et al., 2021; Botch et al., 2023). Similarly, placing the trapped person in a separate room (i.e., without cylinders) ensured that visual search conditions would not be confounded by his visual salience or proximity to the cylinders in the room. In total, 25 cylinders of the same height (2.6 Unity-units) and three different widths were shown in each trial. The target cylinders (targets) were 1.50 Unity-units wide, the small distractors (easy distractors) were 30% less wide and the large distractors (difficult distractors) were 20% wider than the targets (Figure 3.7A-B). While the building had at least one target in any of the five rooms, not every room necessarily contained targets; each room could have up to three targets. Each room also had two types of cylinders, allowing participants to compare their sizes to identify the target(s) without resorting to picking the middle-sized cylinder, which would be possible of all three types of cylinders were present. The total number of easy distractors in the building was always 12 across trials, whereas the combined number of targets and difficult distractors was 13, with their distribution varying depending on target prevalence. When target prevalence was low, there were between one and three targets and between 10 and 12 difficult distractors (e.g., two targets and 11 difficult distractors). When target prevalence was high, there were between five and seven targets and between six and eight difficult distractors (e.g., five targets and eight difficult distractors).

Figure 3.7. (A) From left: Small/easy distractor, target (i.e., explosive hazard), large/difficult distractor. (B) The same stimuli in low-visibility. (C) Ariel view of cylinders arranged in an arc within a room.

After the researcher tested the initial building in the Igloo, the dimensions of the initial building design were adjusted. Each room was enlarged to dimensions of 13.80 (width) by 7.30 (depth) by 3.00 (height) Unity units to accommodate the more spread-out arrangement of cylinders within the space (Figure 3.7C), which resulted in a large central hexagonal space. The final redesign of the virtual building introduced a smaller hexagonal space with teleporting doors surrounded the starting point to reduce walking time and exhaustion from covering large room-to-room distances (Figure 3.8). When the participant entered a door (by standing close to it) in the smaller hexagonal space, they were teleported (seamlessly) to the door location of the corresponding room.

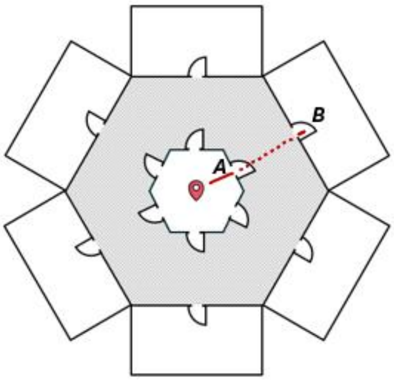


Figure 3.8. Ariel view of building layout. The red marker indicated the starting point in each trial. Doors in the smaller hexagonal area teleported participants to the corresponding rooms to save time and reduce fatigue. The greyed-out area was not visible to participants, providing a seamless walking experience. The solid and dotted red lines show a participant entering door A and getting teleported to location B.

3.4.1.1 High Stress and Low Stress Conditions. In the High Stress condition, the temperature reading was red and fluctuated between 90°C – 95°C (see Figure 3.9). Participants were informed that this, along with an ongoing alarm sound, signalled the risk of an explosion mid-trial which would end the trial prematurely, but could be prevented if all targets were removed in time. The explosion was programmed to occur 90±2.5 seconds into a trial if it occurred. Unbeknownst to them, movement speed was approximately 40% slower than that in the Low Stress condition. This condition was quiet, and the temperature reading was green and fluctuated between 60°C – 65°C, signalling low risk of explosion.

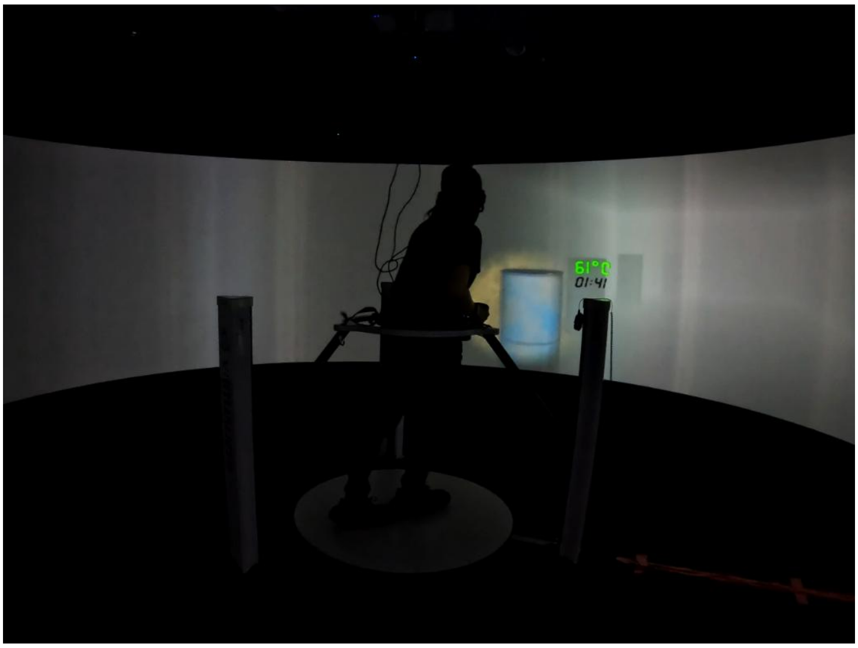


Figure 3.9. Virtual headtorch illuminating parts of a smoky room, virtual environment temperature reading, and countdown timer.

3.4.2 Learning to Identify the Hazard

The day before the Hazard Search Task, participants underwent a learning procedure in the same virtual environment to learn the identity of the target cylinders. This learning procedure, together with a familiarisation phase where participants learned to walk on the omnidirectional platform (described in the feasibility study) were scheduled the day before the main experiment to allow for learning to consolidate overnight in preparation for the experiment proper. This schedule also standardised the time elapsed between learning and testing to ensure that visual search performance was not influenced by variations in retention intervals between participants. Moreover, conducting the full experiment over two days reduced discomfort and fatigue from prolonged VR exposure, thereby minimising the risk of performance impairments during the experiment proper. A combined approach of passive and active learning was employed to speed up the learning process. Learning through observation alone, such as through observing which cylinders were pointed out to be targets, might be insufficient for retention (Shea et al., 2000). Thus, an active learning segment was included to reinforce learning through practice.

Participants completed a passive learning block and an active learning block with four trials each under Low Stress conditions to learn to identify the target cylinders. Participants began each trial at the same starting point (see Figure 3.8). A virtual headtorch followed their head movements, while a temperature reading for the virtual environment and an 85-minute countdown timer followed the headtorch along the horizontal plane (see Figure 3.10). Each learning trial was functionally untimed as it was impossible for participants to exceed the 85-minute duration (given that pilot participants never took more than half a second to make a perceptual decision for each stimulus). To start a trial, participants were prompted to "Press a [small smooth button] to begin". Walking close to a door opened it immediately.

3.4.2.1 Passive Learning Block. During passive learning trials, the participant was prevented from approaching the cylinders and instead observed them from the doorway. They were informed that after a four-second wait at the doorway, any cylinder with an orange arrow appearing above it accompanied by a "ding!" was a target. Only cylinders of that width (i.e., targets) should be removed to prevent explosions (Figure 3.10). No arrows or "ding!" after the wait signalled that there were no targets in the room. Regardless of whether targets were present, they were reminded to learn the identity of the cylinders through observation and size comparisons before leaving for another room. They were also informed that they would have to identify the cylinders on their own during the active learning trials later. These instructions were given as the researcher guided the participant through the first passive learning trial, using it as a live demonstration. After four trials and an optional break, the active learning trials commenced. Participants were informed that these trials were the same as the trials they would experience the following day, except that there would be 2.5-minute time limit imposed for each trial the following day.

Figure 3.10. Orange arrows pointing at targets during the passive learning trials in the learning phase. The participant was unable to move towards the cylinders but was instead instructed to learn the identity of the target (and distractors) cylinders through observation and size comparisons.

3.4.2.2 Active Learning Block. Upon entering the room, the participant was immobilised for two seconds to discourage them from removing the cylinders without due consideration, after which they were allowed to move again. To remove a cylinder, participants were required to move towards it until a prompt appeared ("Press the [big smooth button] to remove the cylinder.") and pressed the button specified. The prompt appeared when participants were approximately 4.5 Unity-units from the cylinder. When a cylinder was removed it would disappear, accompanied by a "ding!" sound if it was a target, or no sound if it was a distractor. Moving away without pressing the button removed the prompt, but not the cylinder. Opened doors stayed open to help participants keep track of the rooms that they had already entered. Movement restriction did not apply to the last room visited given that they only had to rescue the person at that point. Moving close to the victim prompted the message, “Press the [big smooth button] to rescue the person”. After pressing the button, the trial ended. If all targets were removed, the feedback would read, “Congratulations! You have rescued the trapped person and removed all the explosive hazards in the building!”, accompanied by a congratulatory sound. Otherwise, it would read, “You reached the trapped person but the explosive hazards you missed exploded when you were helping him out of the building”, accompanied by a failure sound.

After another optional break, two additional untimed High Stress trials followed. These trials were introduced to demonstrate the level of stimulation participants would be exposed to during the experiment proper. They were told the nature of these trials, and that although the explosion setting was removed for the first trial, it would happen randomly during the experiment proper. They completed the first trial as they did in the active learning block, and in the second trial they were warned before an explosion was manually triggered. After participants confirmed that they would be comfortable with this level of stimulation during the experiment proper, they left the laboratory for the day.

3.4.3 The Hazard Search Task

As in the learning session, participants began each trial at the same starting point (see Figure 3.8). The countdown timer below the temperature reading now indicated a 2.5-minute duration. This time limit was imposed to generate mild time pressure and was determined through trial and error by the researcher, followed by pilot testing with three participants. The remaining procedures were identical to the active learning trials the day before, except that each trial now had to be completed within the time limit. Participants received the same feedback at the end of each trial depending on whether they had removed all targets and rescued the person. Two additional feedback messages were introduced for cases in which the participants failed to rescue the trapped person in time. In such cases, the trial would also end, with a failure sound accompanying the feedback: “You removed all the explosive hazards in the building but did not reach the trapped person in time.”, or, “You did not reach the trapped person in time and the explosive hazards you missed exploded as well.”.

A full description of the experimental design is provided in Chapter 4. Briefly, to investigate how self-generated expectations and instructed expectations of target prevalence influenced visual search, participants initially completed a series of trials under High and Low Stress conditions to develop their self-generated expectations of target prevalence. After that, they were instructed that the target prevalence would either change or remain the same in the subsequent trials, which were also conducted under both stress conditions. The experiment initially included 12 trials in the pre-instruction blocks and 10 trials in the post-instruction blocks. However, the first pilot participant reported that the task became too monotonous and fatiguing at the end of the 12 initial trials. In response, the number of trials were reduced to eight in the initial blocks and six in the post-instruction blocks. This revised structure was then successfully piloted with two additional pilot participants who completed the task successfully and did not have the same complaints. Despite the reduced number of trials in the initial blocks, the pilot participants were still able to acquire expectations of target prevalence that correctly reflected the actual distribution of targets. This structure was used for the experiment proper in Chapter 4.

Chapter 4:

Self-Generated Expectations of Target Prevalence Influence Visual Search During Virtual Hazard Search

4.1 Abstract

To understand how prior expectations and instructions about hazard prevalence affect high-stakes visual search in a semi-immersive virtual environment, where participants take on the role of firefighters in search and rescue missions. Information about target prevalence influences visual search in standard laboratory studies. However, little is known about how prior expectations and new information about target prevalence interact in simulated emergency scenarios. Participants (*n* = 48) received training where the average number of hazards (explosive cylinders) amongst similar distractors was varied (two or six) before participants rescued a trapped person. They were then instructed that hazard prevalence would increase, decrease, or stay similar during test blocks. Stress was manipulated by an ongoing alarm, the threat of trial-ending explosions, and reduced movement speed. Search performance was measured by the number and type of stimuli removed and stress was assessed using self-report and physiological measures. Across high and low stress conditions, more hazards were removed and more false positives occurred (i.e., more distractors removed) when test prevalence was lower than during training, compared to when prevalence levels remained similar. False negatives were consistently low across conditions. Acquired hazard expectations can override explicit instructions, leading to persistent search errors, likely due to difficulties in adjusting decision criteria. These results suggest that training in high-stakes hazard search should incorporate the use of tools and techniques to help mitigate the persistent influence of outdated expectations on search performance.

4.2 Introduction

Visual search is an everyday perceptual task that entails searching for targets (e.g., car keys) among distractors (e.g., clutter in drawer). Common laboratory visual search paradigms might involve searching for a perfect “T” among “L”s or “near-T”s in search arrays (Duncan & Humphreys, 1989; Peltier & Becker, 2016; Rich et al., 2008). Variables known to affect the efficiency of visual search in such studies include target-distractor similarity, the heterogeneity of distractors (Duncan & Humphreys, 1989), the number of stimuli to search through (i.e., set size), and target prevalence (Peltier & Becker, 2016; Rich et al., 2008). The efficiency of visual search is clearly important in real-world settings. For example, in accuracy-critical situations like airport security, failing to detect a weapon during baggage screening could be disastrous. Studies using x-ray baggage images also showed that target salience (Biggs et al., 2014), set size, and target prevalence (Wolfe et al., 2005, 2007) influence visual search. In time-sensitive scenarios, such as during a fire, misidentifying an innocuous item as an explosive hazard diverts attention from and delays emergency response directed to the immediate threat. Aside from standard prevalence effects, search errors might arise from mismatches between expected and true target prevalence. In accuracy-critical tasks, such as baggage screening, prevalence expectations affect high-stakes visual search differently depending on whether they were based on experience (i.e., self-generated) or on instructions (Ishibashi et al., 2012; Lau & Huang, 2010). However, how these factors interact when both speed and accuracy are required (e.g., in search and rescue missions) remains underexplored.

The low prevalence effect describes the robust finding that infrequent targets are more likely to be missed compared to frequent targets (Wolfe et al., 2005, 2007). Prevailing explanations for this effect include conservative perceptual decision thresholds (Wolfe et al., 2007; Wolfe & Van Wert, 2010) and premature search termination during target absent trials

(Rich et al., 2008). Conservative decision thresholds and premature search termination are represented in the Multiple-Decision Model (MDM) for visual search as perceptual and a decisional problems, respectively (Wolfe & Van Wert, 2010). The evidence accumulation process described in the MDM offers a formalised account of target determination in the Guided Search 6.0 model (Wolfe, 2021) (see Section 1.5.1). The MDM model posits that a selected stimulus from the search array undergoes a two-alternative forced choice process, where evidence accumulates towards the “target” or “distractor” decision boundary. If classified as a target, a “target present” decision is made (cf. Schwark et al., 2013). Otherwise, the process repeats for another stimulus until the quitting threshold is reached, leading to a “target absent” decision. This quitting threshold increases under high prevalence but decreases under low prevalence, contributing to the low prevalence effect (Wolfe & Van Wert, 2010). Likewise, when evaluating a stimulus, high prevalence shifts the starting point of evidence accumulation towards the “target” decision boundary, and low prevalence shifts it toward the “distractor” boundary, thereby influencing the time taken for evidence accumulation towards a decision boundary (Peltier & Becker, 2016). Thus, the perceptual decision process is less conservative under high prevalence for “target” decisions, but more conservative under low prevalence (Peltier & Becker, 2016; Wolfe & Van Wert, 2010).

Prevalence Expectations and Search Expectations

In real life, expectations of target prevalence are often informed by experience and could be considered self-generated. For example, baggage screeners typically expect a low prevalence of weapons and medical professionals typically anticipate low prevalence of malignant lesions in diagnostic screenings due to their relative rarity. Laboratory experiments using x-ray images of baggage or mammograms show that expert participants’ expectations can be increased, and thereby attenuate the low prevalence effect (K. K. Evans et al., 2011, 2013; Nakashima et al., 2015; Wolfe et al., 2013). This manipulation highlights the role of direct experience not only in generating expectations of target prevalence, but also in modifying highly established ones among experts. Similarly, in artificial search tasks without real-world context, direct experience, such as during a practice session, has been used to generate expectations about target prevalence in visual search (Hon & Jabar, 2018; Ishibashi & Kita, 2014; Peltier & Becker, 2016).

Direct instructions can also shape prevalence expectation to produce prevalence effects (Rich et al., 2008; Taylor et al., 2022; Zhang & Houpt, 2020). When informed that targets would be rare in "T among L" searches, more targets were missed and search was terminated earlier for target-absent trials (Rich et al., 2008; Zhang & Houpt, 2020), even after participants had to view the search array for at least two seconds before responding (Rich et al., 2008). However, decision criterion and response times differed depending on whether prevalence information was learned through experience or informed directly (Zhang & Houpt, 2020). When targets were rare, participants were more likely to make a target-present response when prevalence information was learned than when instructed. When targets were common, target-present responses were more likely when prevalence information was instructed than when learned. According to the authors, this bias stemmed from how experience and instruction formed expectations of target prevalence: Through experience, participants relied on recent sampling of targets (i.e., local prevalence) to form expectations on whether subsequent trials would contain a target, whereas through instruction, global prevalence information was readily available to guide visual search directly (Zhang & Houpt, 2020).

Instructions about target prevalence also influence visual search behaviour and performance regardless of true prevalence. Subtle wording differences in search instructions produce effects similar to the low prevalence effect, where shorter searches and lower hit and false alarm rates were observed for those expecting "0, 1, or 2 targets" than those expecting "1 or 2 targets" (Cox et al., 2021). Beyond expectations of prevalence, instructions emphasising speed or accuracy also modulate speed-accuracy trade-off in search performance (Lawrence et al., 2023; McCarley, 2009). When accuracy is prioritised, saccade frequencies increased and correlated with longer RTs, which, along with enhanced target detection in visual periphery, improved hit rates (McCarley, 2009). Notably, analysis of fixation durations revealed that search strategy adjustments emerged only after initial orientation to the search scene (McCarley, 2009).

Despite the strong influence of instructions, self-generated expectations of prevalence guide visual search when the two sources of information are in conflict (Ishibashi et al., 2012; Lau & Huang, 2010). Participants who initially experienced low target prevalence took longer to search for targets in target-absent trials within blocks described as high target prevalence. However, even with lengthier search durations (i.e., search behaviour), the elevated miss error rates (i.e., search performance) reflected the self-generated expectation that target prevalence would be low (Lau & Huang, 2010). Nevertheless, there is evidence that despite experiencing the same global prevalence rate, a high prevalence cue slightly increased false positives compared to a low prevalence cue (Ishibashi et al., 2012), and an "extremely low prevalence" cue (e.g., 3%) reduced false alarm rates and increased miss rates (Ishibashi & Kita, 2014).

The broader literature on the reliance on self-generated expectations (relative to instructed expectations) in other visual tasks has been taken to suggest that its influence reflects available online cognitive capacity (Kemper et al., 2012) and involves differential attentional and preparatory processes compared to instructed expectations (Gaschler et al., 2014; Jiang et al., 2018; Kemper et al., 2012, 2017; Oberauer et al., 2013; Schwager et al., 2017; Umbach et al., 2012). Mismatch effects, where slower RT to stimuli that violate predictions than those that confirm predictions, were significantly larger when self-generated expectations were violated than when instructed expectations were violated (Kemper et al., 2012). This has been observed even when self-generated expectations were akin to guesses and instructed expectations were deeply processed (Kemper et al., 2017; Umbach et al., 2012). Neurophysiological evidence suggested that self-generated expectations involve greater top-down attentional engagement and preparatory processes, possibly making their violations more demanding to process than violations of instructed expectations (Kemper et al., 2012). These findings raise the possibility that the influence of self-generated expectations in visual tasks might be sensitive to the availability of cognitive resources, which are often depleted under stressful conditions (Hyun et al., 2019; Qin et al., 2009; Shields et al., 2016). Whether stress shifts reliance between self-generated and instructed expectations during visual search is unknown.

Stress and Visual Search

Executive function is affected by stress, with impairments in working memory, cognitive flexibility, and cognitive control (Diamond, 2013; Hyun et al., 2019; Plessow et al., 2011; Shields et al., 2016). These impairments are often attributed to the reallocation of cognitive resources that normally support executive function to managing stressors (e.g., by inhibiting stress-related interference; Hyun et al., 2019; Qin et al., 2009; Shields et al., 2016), potentially reducing reliance on anticipatory strategies in favour of less demanding reactive strategies (Steinhauser et al., 2007). There is direct evidence showing that visual search is compromised by stress. For example, in visual search tasks that model visual noise in mammograms (Rieger & Manzey, 2024) and mimic baggage screenings (Rieger et al., 2021), time pressure impairs search performance. These tasks used a procedure in which a computer mouse was used to reveal small sections of a fully hidden search image. Search was less thorough on target-absent trials under high time pressure, as indicated by the percentage of the image that was uncovered (i.e., search amount) and the rate of uncovering the image per second (i.e., search speed). The response criterion was more conservative under high time pressure, with a reduction in accuracy characterised by greater miss rate and false positives (Rieger et al., 2021; Rieger & Manzey, 2024). The observed difficulty to maintain search performance under time pressure suggests that visual search relies on the availability of cognitive resources, which are taxed under stress (see Anderson, 2024; Anderson & Lee, 2023). Under such conditions, whether previously acquired self-generated expectations of prevalence would continue to overshadow recently Instructed Prevalence (Ishibashi et al., 2012; Lau & Huang, 2010) during visual search is unknown.

To the best of our knowledge, the use of different prevalence information in speeded, high-stakes visual search contexts is unexplored. Virtual reality (VR) is well-suited to address this problem as it enables systematic manipulation of the environment to increase immersion and presence without endangering participants. Visual search in VR produced comparable search performance as their 2D counterparts on desktop computers (Beitner et al., 2024; Olk et al., 2018) and real life search (Van Den Oever et al., 2022). Naturalistic VR search tasks include searching for singleton objects in living spaces (Beitner et al., 2024; Botch et al., 2023) or tracking a moving person in a crowded corridor (Bennett et al., 2021). In VR-simulated search and rescue missions, standard findings from laboratory visual search tasks have been replicated when participants searched for the source of a fire in a smoke-filled room with virtual thermal imaging cameras or heat distribution helmet displays (Feder et al., 2024).

The present study used a semi-immersive VR system to implement a hazard removal task, where participants assumed the role of a firefighter navigating a burning building to identify and remove explosive hazards among visually similar distractors. While less immersive than VR headsets, this system limits cybersickness risk (Saredakis et al., 2020;

Srivastava et al., 2019) because there is less sensory conflict (Palmisano et al., 2020), and thereby allows extended sessions in the virtual environment. The experiment aimed to assess the effects of self-generated expectations of target prevalence, Instructed Prevalence, and stress on high-stakes visual search. In low stress conditions, timed visual search was carried out with low levels of background noise, without interruptions, and at a normal movement speed. In contrast, in high stress conditions there were randomly occurring trial-ending explosions (which added time pressure), loud alarms, and restricted movement speed.

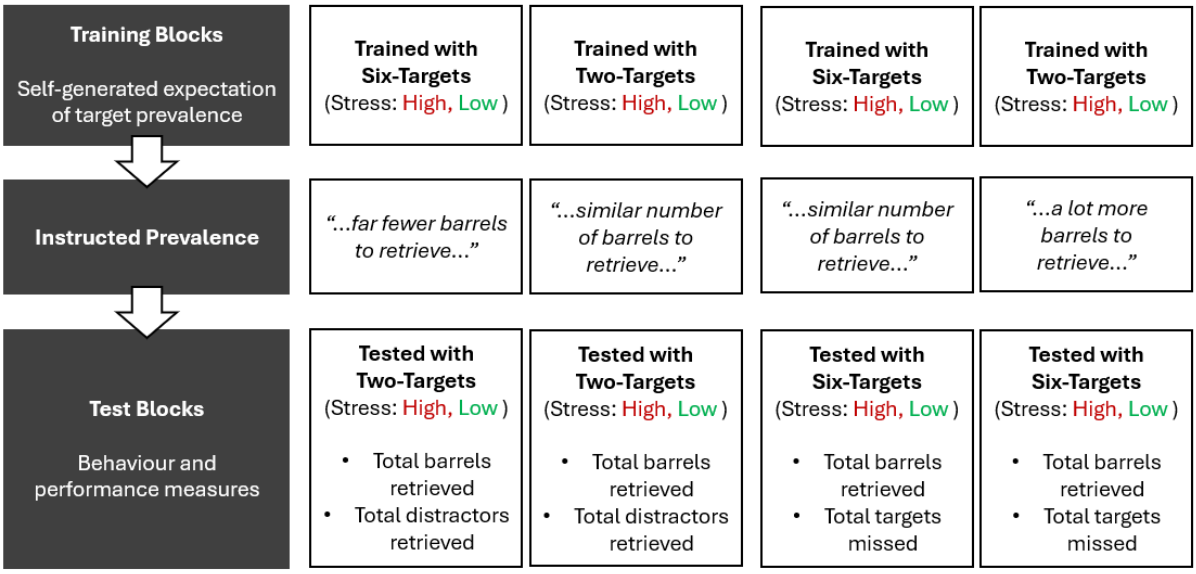


Figure 4.1. Schematic plan of the experiment design showing the target prevalence in training and test blocks, and the Instructed Prevalence for each condition. High and Low Stress blocks were counterbalanced.

On day one of the study, participants were familiarised with the VR and task setups, and day two was divided into two stages: training and test. Half of the participants received training with an average of two or six targets (explosive hazards), and then were tested with an average of six targets. Prior to the test, they were given instructions that were consistent with whether target prevalence would change or not during testing (i.e., a lot more or similar number of targets). If training was the dominant influence, then the groups, each with their own training experience that informed their self-generated expectations of target prevalence, should behave differently:

- Participants tested with two targets: Those trained with two targets are predicted to remove fewer potential hazards, make fewer false positives (i.e., removing distractors), and make more false negatives (i.e., misses) than those trained with six targets.
- Participants tested with six targets: Those trained with six hazards are predicted to remove more potential hazards, make more false positives, and have fewer false negatives than those trained with two hazards.

If instructions had a greater influence, then these measures should reflect these instructions, which correspond to the actual number of targets presented at test; with high prevalence during the test resulting in fewer misses than low prevalence (cf. Wolfe et al., 2005, 2007).

Stress was manipulated within-subjects: Both training and test stages contained one low-stress and one high-stress session in a counterbalanced order. Low stress sessions involved a quiet environment, a 2.5-minute time limit per trial, and normal movement speed. High stress sessions involved an ongoing alarm, explosion threats (which increased time pressure), and reduced movement speed. If high stress reduces the influence of self-generated expectations, then the pattern of errors in the high stress sessions might be expected to be dominated by the instructions that participants received; whereas in the low stress sessions the influence of self-generated expectations should dominate.

4.3 Methods

4.3.1 Participants

Forty-eight participants with equal sex distribution (mean age = 20.51 years, range = 18.17-21.59 years) were recruited from the student population of Cardiff University and received course credits or £25 for their participation. Participants had normal or corrected-to-normal vision, did not wear prescription glasses, had no adverse experience with fire personally or through loved ones, had not experienced motion sickness in the past three months, and were not currently ill, taking psychotropic medication, suffering from hypertension, or diagnosed with mood/anxiety disorders. Approval for this research was granted by the School of Psychology Research and Ethics Committee (EC.22.04.26.6560GRA).

4.3.2 Materials

4.3.2.1 Virtual Reality Setup and Other Equipment. The experiment took place in a six-metre diameter Igloo Immersive Cylinder (hereafter called the Igloo), complete with a surround-sound system. A Cyberith R&D Kit omnidirectional platform positioned in the middle of the Igloo supported locomotion in the virtual world. Interaction with the virtual world was facilitated by a Steam Controller gamepad. A Vive Tracker 3.0 motion tracker (attached on top of the emteqPRO mask) controlled a virtual headtorch. A mechanically integrated emteqPRO mask and Pupil Core eye-tracker recorded facial electromyography (fEMG) data and eye-tracking data, respectively, while a PolarOH1+ armband recorded heart rate (HR) data. Further details about the VR setup and equipment are presented in Sections 3.1.1 and 3.2.1.

The smoke-filled virtual building had six rooms arranged in a hexagonal formation around a mid-point where the participants began each trial (see Figure 4.2). Five of these rooms consisted of five cylinders each arranged in an arc (thus adding up to 25 cylinders in the building, per trial), and the trapped person was in the sixth room. Crucially, these cylinders were similar in height and appearance, but differed in width: The smallest and largest cylinders were (harmless) distractors, whereas the middle-width cylinders were the explosive targets. In each room, only two types of cylinders were present to prevent participants from relying on a “middle-size” selection strategy, which would be possible if all three widths were present. Further details about the building and cylinders are presented in Section 3.4.1.

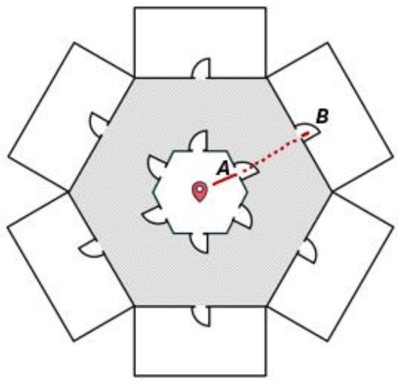


Figure 4.2. Ariel view of building layout. The red marker indicated the starting point in each trial. Doors in the smaller hexagonal area teleported participants to the corresponding rooms to save time and reduce fatigue. The greyed-out area was not visible to participants, providing a seamless walking experience. The solid and dotted red lines show a participant entering door A and getting teleported to location B.

4.3.2.2 Hazard Search Task: Learning and Test. A brief recap of the training process on Day One and the test on Day Two is provided below, with detailed descriptions provided in Sections 3.4.2 and 3.4.3.

Learning. All trials were untimed during training. Briefly, participants underwent a passive learning block and an active learning block with four trials each under Low Stress conditions (see Section 3.4.1.1). In the passive learning block, upon entering a room, they were prevented from approaching the cylinders in each room and instead learned their identity by observing from the doorway. After a four-second delay, target cylinders were marked by orange arrows above them and accompanied by a “ding!”, whereas distractors were not indicated by either cue.

In the active learning block, participants were immobilised for two seconds upon entering a room before they were allowed to approach the cylinder. Moving close to a cylinder triggered a prompt that gave them the option to press a button to remove it. If the cylinder removed was a target, it disappeared with a “ding!”; if it was a distractor, it disappeared but without the sound. When participants entered the last room, they could immediately approach the victim and rescue him with a button press when prompted. Search performance (i.e., whether all targets were removed) was provided after each rescue. Two additional trials under High Stress conditions (see Section 3.4.1.1) followed to familiarise participants with the level of stimulation they would experience during the experiment proper.

Test. The testing phase is identical to the active learning phase on Day One, but each trial had a 2.5-minute time limit. If the rescue exceeded this time limit, this was reflected in the end-of-trial feedback.

4.3.3 Procedures

Stress level was manipulated within-subjects during training and test. Target prevalence during training and test was manipulated between-subjects using a 2 x 2 factorial design (an average of two or six targets during training and an average of two or six targets at test; see Figure 4.1). The first day involved acclimation to the VR environment and learning the game mechanics, and the next day included the training and test sessions.

4.3.3.1 Day One: Acclimation and Learning. Participants were equipped with an emteqPRO mask (mechanically integrated with Pupil Core), PolarOH1+ on their bicep, a Vive Tracker, and low-friction overshoes from Cyberith GmbH. The researcher provided safety instructions, assisted participants onto the motion platform, secured them in place, and guided them in walking using video demonstrations and verbal instructions. They were given up to 20 minutes to practice walking towards various objects in a virtual park environment until:

(i) They rated at least 7 to "On a scale of 1 to 10, with 1 being completely unnatural and 10 being almost as natural as walking in real life, how would you rate your experience now?"

(ii) They reported "yes" to "If you were given mental arithmetic tasks to do now while walking on this platform, do you think you will be able to do them without having to think about walking?"

(iii) The researcher judged objects in the virtual environment as moving smoothly when the participant was walking.

The participants were then trained to use a Steam Controller gamepad to learn the button names used in the experiment for interacting with in-game elements (see Figure 3.2 in Section 3.1.1). A three-minute mask calibration (Section 3.2.1.1) was then carried out, but Day One fEMG data was not analysed.

Participants read the on-screen information while the researcher set up the hazard learning phase: "A person is trapped in a burning building with explosive hazards. You have 150 seconds to reach them before their health is compromised. To rescue them successfully, you also have to look for and remove all explosive hazards along the way. Otherwise, they will explode when you try to carry the person out of the building." They were briefed that they would be roleplaying a firefighter and learning to identify target cylinders amongst similarly sized distractors in preparation for the experiment proper the following day, where the time limit would only apply. The learning phase (Section 4.3.2.2; for a full description, see Section 3.4.2) began after that.

4.3.3.2 Day Two: Experiment Proper. At the main computer next to the Igloo, participants were fitted with the same equipment. After ensuring good skin contact for the fEMG electrodes (see Section 3.2.1.1) and robust pupil detection for the eye-tracker (see Section 3.2.1.3), the mask and eye-tracker were disconnected from the main computer. The participants were then led into the Igloo to practice walking on the motion platform to regain familiarity, after which the mask and eye-tracker were reconnected via USB extenders. The participants relaxed and took deep breaths for two minutes, during which their baseline HR was recorded. After that, they displayed maximum expressions (smile, frown, raised eyebrows), each alternating with a neutral expression for mask calibration. They were again briefed on their role as a firefighter, who must remove all targets and rescue the victim in 2.5 minutes. They were also reminded of the features in Low and High Stress conditions. They were informed that the researcher would only enter the Igloo during breaks to check on them and make adjustments (e.g., retighten motion platform harness and disentangle cables). Participants were guided through the first eye-tracking calibration session (see Section 3.2.1.3) and informed that a new calibration would be carried out at the start of each block of trials (for a total of four eye-tracking calibrations throughout the experiment). They began the experiment once the researcher exited the Igloo and signalled for them to start.

Figure 4.1 shows the design of the experiment. Participants first received two training sessions, which each included four trials. Half of them received one to three targets (i.e., Two-target training) and the remainder received five to seven targets (i.e., Six-target training). In a counterbalanced order, one session was conducted under Low Stress and the other under High

Stress, including two trials set for explosions 90±2.5 seconds into the trial. During each of the two test sessions, participants received three trials. Of the participants who received training with six targets, half received six targets during the tests and the remainder received two targets. Similarly, for those who received training with two targets, half received six targets during the tests and the remainder received two targets. Again, one of the two test sessions was conducted under High Stress, with the second trial containing an explosion, and the other test session was under Low Stress. The order in which a given participant received the two types of session was the same as during training. During training and test, a three-minute break separated each session. The break featured a countdown circle against a background of nature scenery and sounds to promote relaxation, and could be extended upon request.

At the start of the test sessions, participants received the Instructed Prevalence, "Before you start, I need you to know that there is/are ___ to remove in the building now as/than there were before." When training and test prevalence were the same, the blank was replaced with "a similar number of cylinders"; otherwise, the blank was replaced with "a lot more cylinders" or "far fewer cylinders" when the number was increased or reduced, respectively. At the end of each session, participants provided a rating to the question, "How were you coping with the search and rescue tasks just now?", with 1 representing "I felt no pressure" and 10 representing "I was unable to cope with the pressure." At the conclusion of the experiment, participants were led out of the motion platform and Igloo.

However, for a large number of participants, pupil detection was inaccurate. As shown in Figure 4.3, the border of the pupils and the iris was not clearly demarcated, which prevented accurate pupil detection during both live recording (Figure 4.3A) and post-hoc processing (Figure 4.3B). This remained the case even after implementing suggestions provided by Miguel from Pupil Labs (personal communication, November 25, 2022), such as adjusting the exposure settings on the eye cameras to enhance the pupil-iris contrast and adjusting gain, brightness and contrast during post-processing. As this issue affected a substantial proportion of the recordings, the eye-tracking data was ultimately not analysed.

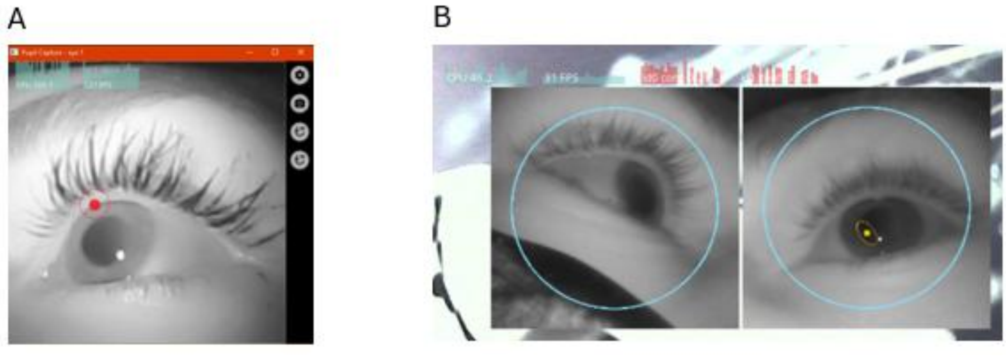


Figure 4.3. Examples of poor pupil detection: (A) Image from the right eye camera of a female participant during live recording in the Pupil Capture software. The red circle and dot, meant for pupil detection, did not align with the pupil. (B) Images from the eye cameras of a male participant during post-hoc pupil detection in the Pupil Play software. The red dot and circle were absent, and the red bars above the eye images indicated low pupil detection confidence (below the 0.60 threshold).

Response measures for the Hazard Search Task. The effects of Training Prevalence (i.e., self-generated expectations of prevalence) and Instructed Prevalence on search behaviour were assessed using the average number of cylinders removed in the Test sessions. Their effects on search performance were assessed using the number of false positives (i.e., distractors retrieved) and false negatives (i.e., targets missed). Analyses were not conducted on hits (i.e., targets correctly removed) as they are the inverse of false negatives. For a given Test Prevalence condition (e.g., six targets), if search behaviour or performance were similar between those who experienced different Training Prevalence (six or two targets), then it would suggest an influence of their instructions or actual prevalence during the test. Whereas if search behaviour or performance was affected by whether training involved six or two targets, it would suggest that the expectations generated by training were influential.

Data Preparation. The emteqPRO mask produced a JSON file containing timestamps and event labels and a DAB file containing sensor data. These were uploaded to the SuperVision App for processing to generate a CSV file, from which mean amplitudes (μV) for each session were derived with a Python script adapted from the "One-user Analysis Script 1" sample provided by Emteq Labs (emteq labs & Mavridou, 2025). A modified JSON file that excluded 5-second periods of explosions was used in the script to reduce bias from abrupt facial expressions. Values from bilateral muscle groups (frontalis, orbicularis, zygomaticus) were averaged by muscle group.

The mean HR (bpm) for each session was calculated from the CSV file produced by the PolarOH1+ using another Python script, accounting for a 17-second lag for the armband (see Section 3.2.1.2). This lag was corrected by shifting the session timestamps backwards by 17s. Likewise, the 5s-exclusion for each explosion was applied before calculating mean HR for each block. Session averages were computed for total number of cylinders removed, together with whether they were targets or distractors. Trials on which participants failed (i.e., did not rescue the victim or when an explosion occurred) were excluded from this analysis.

All statistical tests were performed using R, version 4.4.0 (R Core Team, 2024). Where assumptions for parametric tests were violated, non-parametric alternatives were conducted or robust methods were used (Maechler et al., 2006; Mair & Wilcox, 2020). For the response measures, Bayes factors (BF_{10}) using default priors were also computed using the BayesFactors R package (Morey et al., 2024). For skewed data, log-transformed data were used if it reduced skewness; otherwise, raw data were analysed. BF_{10} values greater than one favour the alternative hypothesis, whereas values less than one favour the null hypothesis. Evidence categories follow Andraszewicz et al. (2015).

4.4 Results

4.4.1 Training

Average trial durations for each block were computed for trials when participants did not exceed the allocated time limit (i.e., 2.5 minutes). As movement speed was artificially slowed down in the high stress relative to the low stress conditions, the analysis of trial duration was separated by stress group. Table 4.1 shows the median (IQR) trial durations of the training and test blocks, grouped by target prevalence and stress. In the low stress condition, trial duration was shorter in the two targets than the six targets conditions for the training and test blocks. Mann-Whitney U tests confirmed that the difference was significant on training blocks ($Z = 2.89$, $p = .004$, $r = .42$) and test blocks ($Z = 2.76$, $p = .006$, $r = .40$). The same patterns of results were evident in the high stress condition for training blocks ($Z = 2.94$, $p = .003$, $r = .43$) and test blocks ($Z = 2.12$, $p = .033$, $r = .31$).

Table 4.1. Median (IQR) Trial Duration (in Seconds) for Training and Test Blocks by Test Prevalence and Stress

	Training Blocks	Test Blocks
Target Prevalence	Two Targets /Six Targets	Two Targets /Six Targets
Stress		
Low Stress	98.38 (21.71) / 117.13 (19.40)	89.00 (27.83) / 104.17 (11.09)
High Stress	98.92 (21.00) / 112.50 (27.50)	94.00 (38.25) / 116.25 (16.50)

4.4.2 Stress measures

Paired t-test showed that mean perceived difficulty to cope rating was significantly lower in the low stress condition (mean = 2.94, SE = 0.21) than in the high stress condition (mean = 4.92, SE = 0.26) conditions, $t(47) = 9.82$, $p < .001$, Cohen's $D = 1.42$. Due to technical issues, facial EMG data was missing from two participants. Table 4.2 shows the median amplitudes (μV) for the corrugator, frontalis, orbicularis, and zygomaticus muscles in the low and high stress conditions ($n = 46$). Mann-Whitney tests confirmed what inspection of Table 4.2 suggests, that there were no significant differences in muscle activation between stress conditions for the corrugator ($Z = 0.91$, $p = .37$), frontalis ($Z = 0.42$, $p = .67$), orbicularis ($Z = 1.03$, $p = .30$), and zygomaticus, ($Z = 0.50$, $p = .62$) muscles. Average heart rate in the low stress condition (median = 109.85bpm, IQ = 14.75bpm) and high stress condition (median = 111.84bpm, IQR = 15.65bpm) did not differ significantly, $Z = 1.06$, $p = .288$.

Table 4.2. Median (IQR) Facial Muscle Amplitudes (μV) in Low and High Stress Conditions

	Low Stress	High Stress
Muscle group		
Corrugator	3.51 (1.85)	3.98 (2.17)
Frontalis	6.91 (5.09)	7.02 (6.19)
Orbicularis	4.24 (2.35)	4.95 (2.44)
Zygomaticus	6.35 (14.78)	7.08 (11.20)

4.4.3 Test Blocks: Total Cylinders Removed

Figure 4.4 shows that when tested with two targets, the groups trained with six targets removed more cylinders than those trained with two (upper panels); and there was no such difference in the groups tested with six targets (lower panels). This pattern of results was evident in high stress (left panels) or low stress (right panels) sessions. There was no effect of training when participants were tested with six targets. A 2(Training Prevalence: 2 or 6 targets) x 2 (Test Prevalence: 2 or 6 targets) x 2(Stress: high or low) mixed ANOVA found significant main effects of Training Prevalence ($F(1, 44) = 12.00$, $p = .001$, $\eta^2_G = .16$) and Test Prevalence ($F(1, 44) = 94.54$, $p < .001$, $\eta^2_G = .61$), but no effect of Stress ($F(1,44) = 0.56$, $p = .458$, $\eta^2_G = .00$). There was a significant interaction between Training and Test Prevalence, $F(1, 44) = 5.30$, $p = .026$, $\eta^2_G = .08$. Simple effects analyses confirmed a significant difference between the groups tested with two targets was significant ($p < .001$), but not between the groups tested with six targets ($p = .319$). There were no other significant interactions (ps = .308-.868). The corresponding Bayes factors showed strong evidence for Training Prevalence ($BF_{10} = 27.37$), extreme evidence for Test Prevalence ($BF_{10} \approx 4.53 \times 10^9$), moderate evidence supporting the null for Stress ($BF_{10} = 0.28$), and anecdotal evidence for the interaction between Training and Test Prevalence ($BF_{10} = 2.35$). All other interactions showed anecdotal to moderate evidence for the null ($BF_{10} = 0.29 – 0.42$).

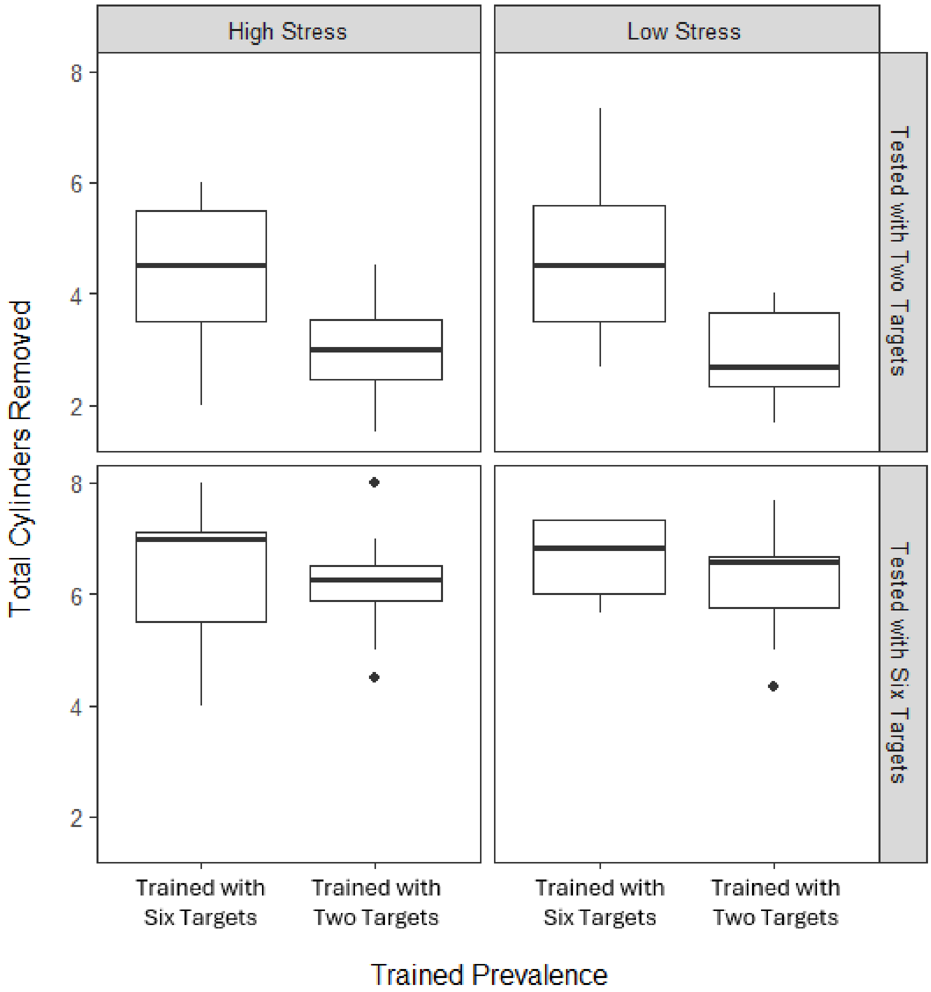


Figure 4.4. Boxplot showing the median number (bold line) and interquartile range (box) of cylinders removed in the test in both stress conditions (high and low), separated by whether training and test involved two or six targets. Individual points represent outliers.

As heterogenous variances and non-normality were observed, a robust between-subjects ANOVA was conducted to confirm the interaction between training and test prevalence, collapsing across stress conditions. Figure 4.5 shows a boxplot of the number of cylinders removed across groups. There was a significant main effect of Training Prevalence (X^2 = 8.72, p = .009), test prevalence (X^2 = 69.13, p = .001), and a significant interaction between these factors (X^2 = 5.20, p = .036). Bayes factors indicated strong evidence for Training Prevalence ($BF_{10} = 24.78$), extreme evidence for Test Prevalence ($BF_{10} \approx 3.90 \times 10^9$), and anecdotal evidence for their interaction ($BF_{10} = 2.35$). For the groups tested with six targets, robust independent t-tests found no significant effect of training ($F(1, 13.96) = 0.41$, Bonferroni-corrected $p = 1.00$, $\xi = .19$), with anecdotal evidence for the null ($BF_{10} = .054$). For those tested with two targets, those trained with six targets removed significantly more cylinders than those trained with two targets ($F(1, 8.93) = 9.47$, Bonferroni-corrected $p = .027$, $\xi = .72$), with strong evidence supporting the difference ($BF_{10} = 17.81$).

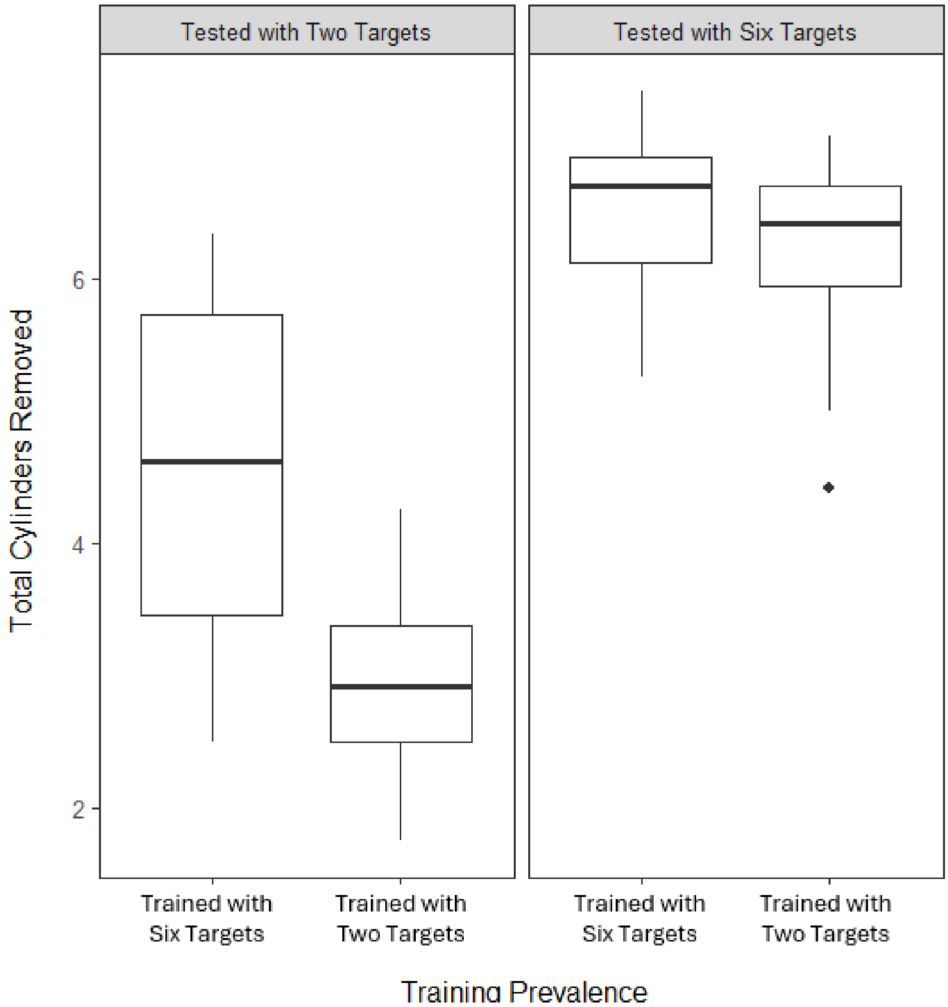


Figure 4.5. Boxplot showing the median number (bold line) and interquartile range (box) of cylinders removed in the test by training and test prevalence. Individual points indicated outliers.

4.4.4 Test Blocks: False Positives

Figure 4.6 shows that when tested with two targets, the groups trained with six targets made more false positives than those trained with two (upper panels); and there was no such difference in the groups tested with six targets (lower panels). This pattern of results was also evident in high stress (left panels) or low stress (right panels) sessions. A mixed ANOVA found that there were significant main effects of Training Prevalence ($F(1, 44) = 12.98$, $p = .001$, $\eta^2_G = .20$) and test prevalence ($F(1, 44) = 22.22$, $p < .001$, $\eta^2_G = .30$), but not stress ($F(1, 44) = 0.37$, $p = .546$, $\eta^2_G = .00$). There was also a significant interaction effect between training and test prevalence ($F(1, 44) = 12.86$, $p = .001$, $\eta^2_G = .20$). Simple effects analyses confirmed a significant difference between the groups tested with two targets ($p < .001$), but not six targets ($p = .987$). There were no other significant interactions (ps = .093-.920). Strong evidence for Training Prevalence ($BF_{10} = 29.57$), extreme evidence for Test Prevalence ($BF_{10} = 511.38$), moderate evidence supporting the null for Stress ($BF_{10} = 0.23$), and strong evidence for the interaction between Training and Test Prevalence ($BF_{10} = 24.78$) were observed. All other interactions showed anecdotal evidence for the null ($BF_{10} = 0.35 – 0.87$).

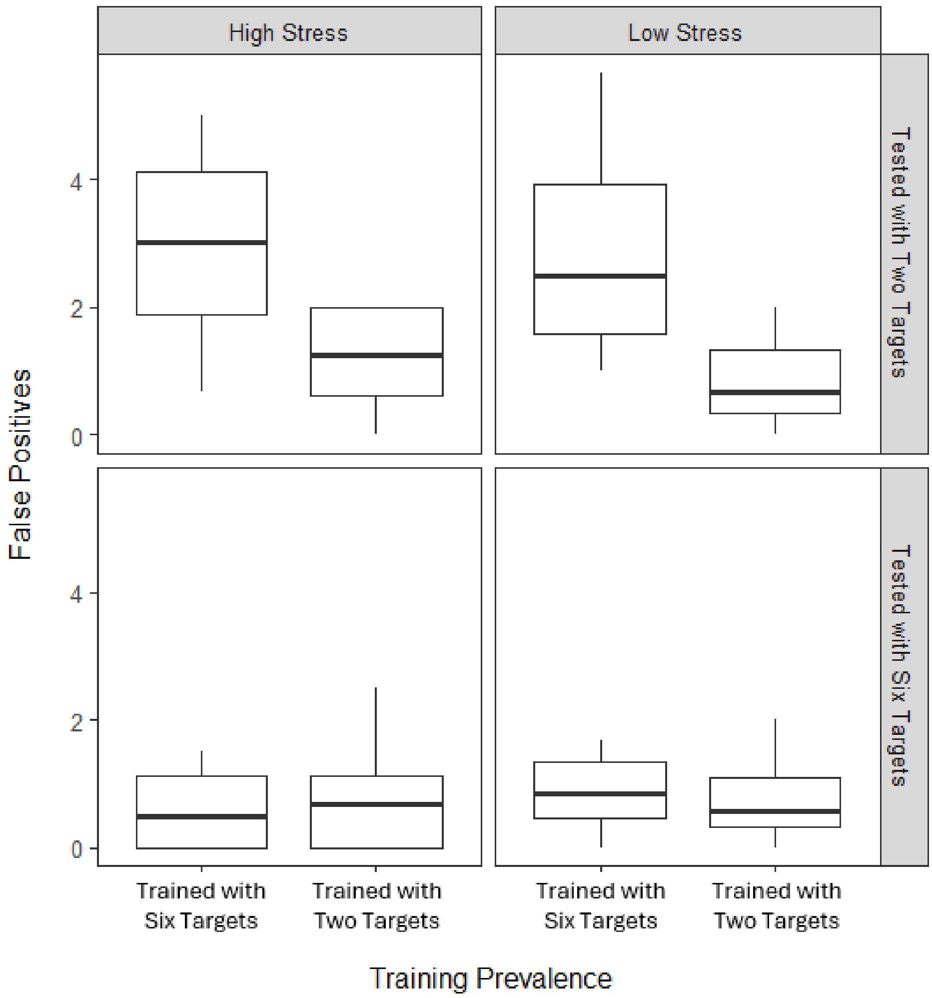


Figure 4.6. Boxplot including the median number (bold line) and interquartile range (box) of false positives in the test in both stress conditions, separated by whether training and test involved two or six targets.

Due to heterogeneous variances and non-normality, a robust between-subjects ANOVA was conducted collapsed across stress conditions. Figure 4.7 shows the distribution of average false positives across the groups. There were significant main effects of Training Prevalence (X^2 = 7.50, p = .016) and test prevalence (X^2 = 14.63, p = .002), and a significant interaction between them (X^2 = 6.83, p = .020). Bayes factors indicated very strong evidence for Training Prevalence (BF_{10} = 33.61), extreme evidence for Test Prevalence (BF_{10} = 644.45), and very strong evidence for their interaction (BF_{10} = 30.04). Robust independent t-tests showed no significant differences between the groups tested with six targets (F(1, 12.19) = 0.03, Bonferroni-corrected p = 1.00, ξ = .05), with anecdotal evidence for the null (BF_{10} = 0.37). However, a significant difference between the group tested with two targets (F(1, 9.43) = 8.26, Bonferroni-corrected p = .035, ξ = 0.77), with very strong evidence supporting the effect (BF_{10} = 40.02).

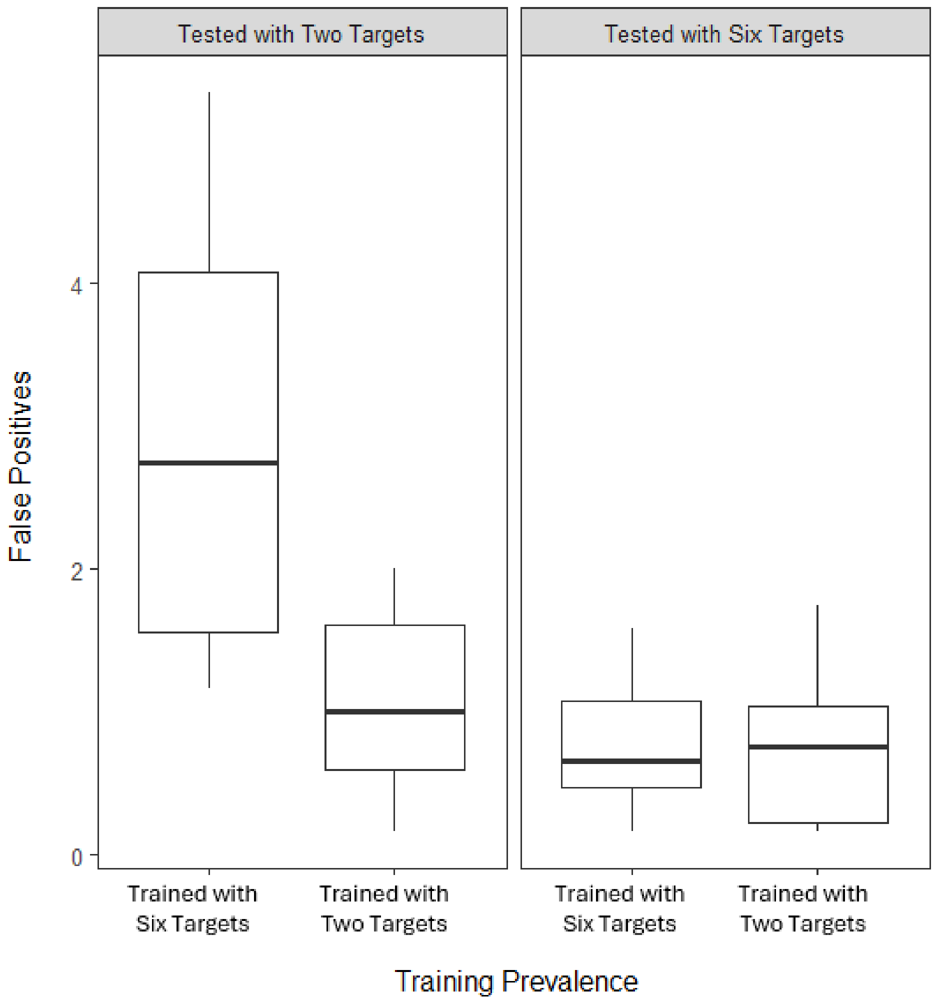


Figure 4.7. Boxplot showing the median number (bold line) and interquartile range (box) of false positives in the test by training and test prevalence.

4.4.5 Test Blocks: False Negatives

Figure 4.8 shows that there were slightly fewer false negatives when tested with two targets than when tested with six, but no effect of Training Prevalence or session type (high or low stress). A mixed ANOVA confirmed a significant main effect of Test Prevalence, ($F(1, 44) = 5.81$, $p = .020$, $\eta^2_G = .08$), with anecdotal evidence supporting the effect ($BF_{10} = 2.96$), but not other main effects or interaction effects (*p*s = .297-.998), with anecdotal to moderate evidence supporting the null ($BF_{10} = 0.27 – 0.40$).

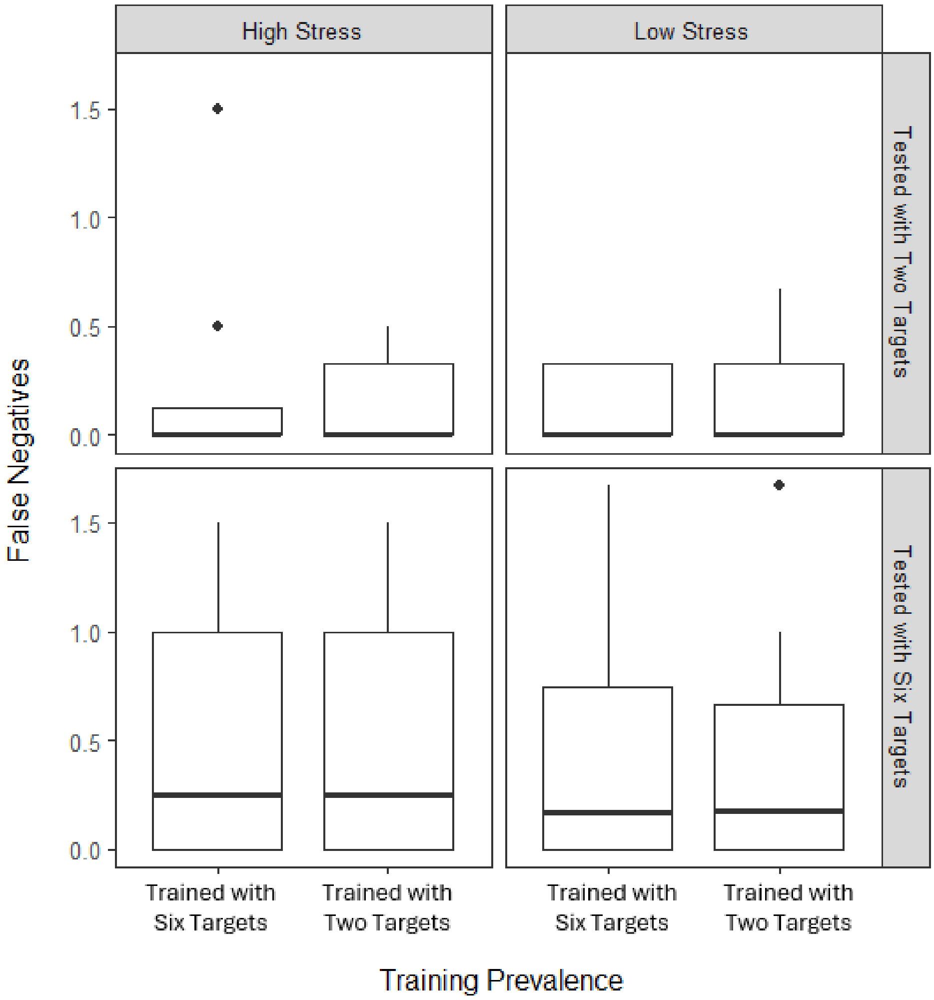


Figure 4.8. Boxplot including the median number (bold line) and interquartile range (box) of false negatives in the test across stress conditions, separated by whether training and test involved two or six targets. Individual points indicate outliers.

Due to heterogeneous variances and non-normality, a robust between-subjects ANOVA was conducted, collapsing across stress conditions. Figure 4.9 shows a boxplot of average false negatives across conditions. The main effects and interaction were not significant (Test Prevalence: X^2 = 3.73, p = .071; Training Prevalence: X^2 = 0.03, p = .865; interaction: X^2 = 0.07, p = .789). However, Bayes factor suggested moderate evidence for the effect of Test Prevalence (BF_{10} = 3.21), and moderate to anecdotal evidence supporting the null for Training Prevalence (BF_{10} = 0.32) or the interaction (BF_{10} = 0.38).

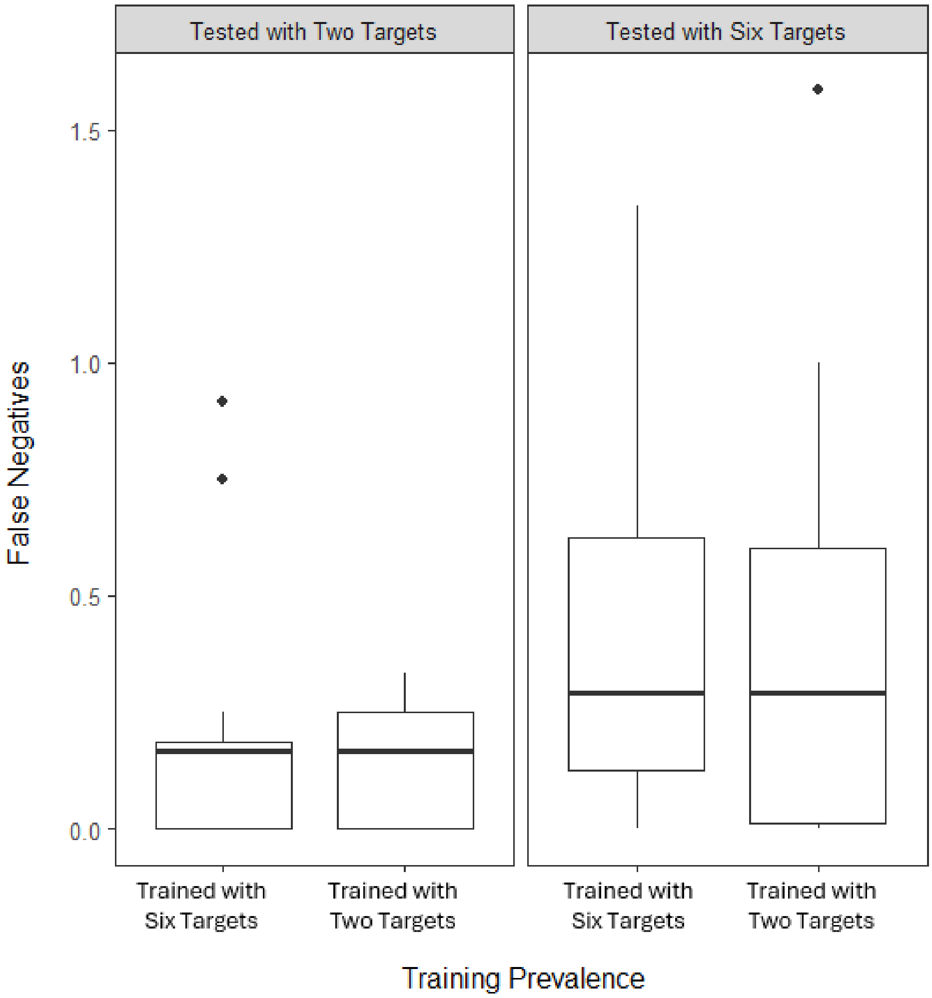


Figure 4.9. Boxplot showing the median number (bold line) and interquartile range (box) of false negatives depending on whether training and test involved two or six targets. Individual points indicate outliers.

4.5 Discussion

The present study investigated how self-generated and instructed expectations about target prevalence affects visual search in high-stress, high-stakes scenarios. In a simulated search and rescue mission, participants identified and removed explosive hazards amongst visually similar distractors, before rescuing a trapped person. In the high stress condition, an alarm, the risk of trial-ending explosions, and artificially reduced movement speed created time pressure during visual search. In the low stress condition, these stressors were absent. After the first two training blocks, where self-generated expectations about target prevalence were established, participants were informed that target prevalence would change or remain the same in the two test blocks. Visual search during the test was jointly influenced by training and test prevalence: When test prevalence was low, those instructed to expect fewer targets removed significantly more cylinders, including false positives (distractor cylinders), than those instructed to expect no change, with strong to very strong evidence supporting this difference. When test prevalence was high, those instructed to expect an increase in target prevalence and those instructed to expect no change removed a similar number of cylinders, including false positives, showing that self-generated expectations had little influence. The numbers of false negatives were largely similar across conditions, with any influence of test prevalence appearing minimal given the statistical evidence. High stress levels did not alter the pattern of findings. These findings are consistent with prior research demonstrating the dominance of self-generated expectations of target prevalence over Instructed Prevalence during high-stakes visual search (Ishibashi et al., 2012; Experiments 1a and 1b, Ishibashi & Kita, 2014; Lau & Huang, 2010).

The setup of the present study precluded the derivation of sensitivity (d') and criterion (C) measures as it bore similarities with multi-target searches (e.g., Cain et al., 2014).

Considering the similar number of false negatives across groups, differences in false positives provide a proxy measure of changes in criterion, with more false positives indicating a more liberal decision threshold. Therefore, learned decision thresholds likely failed to quickly adapt to the Instructed Prevalence when test prevalence was low, matching Lau and Huang's (2010) description of visual search performance as being governed mostly by prior experience. When test prevalence was high, the influence of self-generated prevalence was not evident. However, the low number of false negatives across groups complicates the interpretation of this null effect under high test prevalence. The influence of self-generated expectations could be present but obscured by floor effects (reflected by low false negatives overall). Alternatively, if self-generated expectations truly had minimal or no influence, this could instead suggest a conditional ability to inhibit their influence, such as when visual search was relatively easy under high target prevalence.

The influence of Training Prevalence on test performance with low target prevalence is consistent with the findings of Hon and Jabar (2018). They investigated the amount of exposure needed to learn about and apply prevalence information effectively during visual search. Using reaction time (RT) to index prevalence effects (e.g., longer RTs at low prevalence, they found that when prevalence changed halfway through a 600-trial block, RT adjusted to the new prevalence within 10 target trial (whether targets appeared close together or spaced apart, and whether prevalence increased or decreased), and stabilised thereafter. The authors proposed that prevalence information was incorporated into task representations within those initial trials and influenced subsequent performance (i.e., RT) consistently (Hon & Jabar, 2018). Similarly, in the present study, target prevalence information was established during training through direct experience and continued to bias search performance despite explicit instructions about changes in prevalence. In contrast, Instructed Prevalence, while accurate but not yet experienced directly, likely required time to recalibrate task representations to reliably debias search behaviour. Moreover, the lack of corrective feedback (i.e., whether all targets were removed) at the end of each trial during the test might have reinforced initial biases (Cox et al., 2021) and further delayed this process.

The fact that the influence of Training Prevalence on test performance was evident under high stress may suggest that any resulting depletion of cognitive resources did not disrupt self-generated prevalence expectations during the test (cf. Kemper et al., 2012). It is possible that, in visual search, self-generated expectations might operate differently from or with less reliance on the same attentional processes implicated in self-generated expectations during other visual tasks, such as shape discrimination (e.g., Gaschler et al., 2014; Kemper et al., 2012).

Low Prevalence Did Not Increased False Negatives

In single-target searches, the low prevalence effect refers to a strong tendency to miss targets in rarely-occurring target-present trials interleaved with target-absent trials (Wolfe et al., 2005, 2007). Interestingly, this low prevalence effect, or an analogous effect, was not observed in the present study: The number of false negatives during the test remained similar regardless of the Instructed Prevalence and Training Prevalence. In the present study, where all trials were target-present, low prevalence was defined instead by relatively fewer targets on a trial. As participants were never exposed to target-absent trials, the study was not directly analogous to standard visual search procedures where the low target prevalence effect was observed.

The present findings could also result from task framing. Although speed *and* accuracy were equally emphasised to participants, missing targets (i.e., explosive hazards) was more consequential than the limited time costs associated with accidentally removing a distractor, provided the trials were completed in time. It would be fruitful to investigate whether the dominance of Training Prevalence would persist when the consequences of missing a target and accepting a distractor are similar. Moreover, visual search performance and task performance did not fully overlap: Successfully rescuing the trapped person required the timely removal of all targets, whereas false positives carried no penalties that would affect the rescue beyond a slight delay, at least under low stress conditions. In other words, adopting a more cautious approach might compromise search performance but not task success. Whether these strategic considerations shaped their search behaviour, including their reliance on Training Prevalence (under Low Test Prevalence), is unknown.

The results of the present study demonstrate that in high-stakes contexts, such as firefighting and emergency response, visual search can be susceptible to previously learned prevalence information, even when new and accurate prevalence information is provided. Specifically, those trained under high prevalence made more false positives than those trained under low prevalence when instructed that prevalence would greatly decrease, likely because they were unable to quickly adjust their decision criterion. In real-world contexts, this might translate into time and effort being misdirected to innocuous stimuli misidentified as threats, potentially delaying response to immediate threats and increasing the likelihood of adverse outcomes in time-sensitive situations. The persistence of this effect under high stress, time-pressured conditions suggests that the influence of self-generated expectations might operate in a way that is relatively immune to cognitive resource depletion, rather than easily replaced with new information. These findings have important implications for visual search in first responders, suggesting that mission-critical information communicated in plain language might not enhance search performance if it contradicts previous experience. Future research should investigate the ways in which reliance on established target prevalence could be overcome. Understanding these factors would be critical in improving visual search during emergencies and could inform the development of more effective training protocols and communication strategies.

Chapter 5:

General Discussion

5.1 Overview

In high-stakes situations such as emergency incidents, individuals are often required to make rapid decisions that can significantly affect incident outcomes. These split-second decisions typically draw on extensive experience, which allows first responders to quickly recognise patterns and cues that help them identify from their repertoire a course of action that is typical for a situation (Klein et al., 2010). However, such incidents are inherently dynamic, and new information that is potentially critical to mission outcomes may be communicated at any point, requiring responders to reassess their understanding of the situation before confirming or adjusting their course of action. While one might expect that under high stakes, new information would be readily incorporated to increase the likelihood of positive outcomes, this is not always the case. Prior knowledge or experience, task demands, and the way information is framed could affect the uptake of new information in guiding goal-directed behaviour. Understanding the conditions under which such information is correctly utilised is essential, especially in high reliability domains where failure to adapt one's behaviour effectively can often result in serious and even fatal consequences. To begin to address this need, this thesis investigated factors that influence the integration of newly received information into goal-directed (search and rescue) behaviour. The research used undergraduate participants who played the role of firefighters in simulated fireground scenarios on desktop computers and within semi-immersive virtual reality (VR) environments.

5.2 Summary of New Results

Experiments 1 and 2 investigated how individuals responded to potentially inaccurate new information pertaining to goal-directed behaviour and how the framing of such information influenced its uptake during a desktop-simulated search and rescue (SAR) task. Participants first explored a virtual office building in well-lit conditions and then received a floorplan prior to a SAR mission in the same, now dark and burning building. When the floorplan was described as "might (or might not) be useful" for the SAR mission, participants' search patterns matched the floorplan even when information on the floorplan contradicted their prior knowledge of the building and the area labels in the building during search. It was only when the floorplan was additionally described as "could be outdated" that fewer participants relied on it during search.

Building on these findings, Experiment 3 investigated how the framing of new information that was directly relevant to task goals, but also contradicted previous experience, influenced goal-directed search. After being trained to rescue two managers in a building through six SAR training trials, the experimental groups received a report communicating one of the managers' absence in the building using either explicit or implicit framing. The control group received a similar report, also framed either explicitly or implicitly, about the absence of an unrelated manager not encountered during training. During the "real" SAR mission, those who were told explicitly about the manager's absence searched preferentially the areas in the building where the other manager was expected to be found. In contrast, participants who received the implicitly framed report about the manager's absence behaved like the control group, searching for both managers as they had during training.

Lastly, Experiment 4 investigated whether explicitly communicated probabilistic information that contradicted prior experience could influence visual search performance.

Participants completed a series of visual search tasks where they identified and removed explosive hazards (i.e., target cylinders) amongst similar-looking distractors cylinders. After they generated their expectations of target prevalence through the initial training blocks, they were instructed that target prevalence would change (i.e., increase or decrease) or remain the same in the subsequent test blocks. For participants who were tested with low target prevalence, those who were initially trained to expect high target prevalence removed more cylinders and made more false positives than those who were initially trained to expect low target prevalence. This effect was observed in spite of the clear instruction to expect fewer targets in the test blocks. For those who were tested with high target prevalence, previous training or expectations appeared to have little influence on the number of cylinders removed or false positives errors made. Across all conditions, false negatives largely similar between groups. The same findings were found whether visual search was conducted under high and low stress conditions.

Following this summary of the new results, the next section provides a synthesis of the key findings across experiments to highlight trends and explain divergences where results differ. This analysis aims to identify the underlying factors contributing to or hindering the integration of newly received information into goal-directed behaviour as observed in my experiments. Next, I discuss the theoretical and practical implications of my findings, situating them within existing frameworks and considering how they could inform real-world applications. This chapter concludes with an exploration of future directions prompted by the present findings and a conclusion to the thesis.

5.3 Synthesis of Findings

While the four experiments collectively examine how new information is integrated into goal-directed behaviour in high-stakes and high-stress contexts, they differed in task demands, the framing and format of information, the strength of prior knowledge, and the timing of decisions. This section describes how these factors might have influenced the integration of new information into participants' responses in such contexts.

5.3.1 Task Characteristics and Their Influence on Behavioural Adaptation

Despite sharing the same firefighting context, a notable difference among the experiments lies in the type of task participants performed and how the different task features shaped the decisions they had to make. Experiments 1 and 2 involved navigating through a virtual office building during a SAR mission to search for a trapped person, Experiment 3 involved a goal-directed search task in which participants decided whether to search for one or two trapped persons in the same building, and Experiment 4 involved a timed visual search task in which they identified and removed explosive hazards among similar-looking distractors.

In Experiments 1 and 2, participants had explored the environment beforehand and were then given a floorplan which was either more implicitly or explicitly implied to require further verification before use. Critically, at this point, participants were already informed that they would be looking for the Assistant Finance Manager (which in theory should be in the Finance Department) and could use the floorplan or their prior knowledge of the building layout to plan their rescue path prior to the SAR task. It would be reasonable to assume that they would only begin the SAR mission after they have decided whether to rely on the floorplan and/or planned their rescue path. While Experiment 3 involved the same initial exploration of the environment, the task focus shifted. Rather than navigating the building in search of one person, participants conducted a goal-directed search to rescue one of two managers whom they were trained to expect in the building. The implicit or explicit message that one of the managers was absent was intended to alter the search strategy they learned during training. Like the previous experiments, interpretation of the messaging and the decision to rescue one or two managers likely occurred before participants started the SAR mission, allowing them time to formulate their search strategy.

In contrast, Experiment 4 introduced the most substantial shift in task structure and demands. It required participants to make multiple rapid perceptual decisions during a visual search task, rather than making one decision (e.g., rescue Marketing Manager only in Experiment 3) or several major decisions before action (e.g., judging the floorplan as inaccurate, therefore deciding not to follow it, but rather rely on in-situ area labels during the SAR mission instead of prior knowledge, in Experiments 1 and 2). In Experiment 4, participants had to identify and remove explosive cylinders amongst similar-looking distractors within a fixed duration, with information about changes in target prevalence explicitly conveyed after initial training and before the test blocks. Unlike in Experiments 1-3, although participants also started the task at their own pace, adaptation to updated target prevalence had to occur rapidly during the task and under strict time constraints as they responded to each visual stimulus.

These variations in task structure likely shaped how information was used in ways not fully explained by factors such as pragmatic processing failures in high-stakes, high stress contexts. Experiments 1-3 provided a relatively stable decision window that allowed participants to process new information and incorporate them into their behavioural strategy. The new information in Experiment 4, however, had to be acted on quickly, thus leaving little room for full integration. Furthermore, the perceptual demands of the task itself likely increased the difficulty of completely adapting behaviour to the instructed changes in target prevalence.

5.3.2 Role of Prior Knowledge and Experience

Across these experiments, the role of prior knowledge on how participants responded to the new information likely depended on how strongly that knowledge was encoded, its relevance to the task goals, and the context in which it as acquired. In Experiments 1 and 2, prior knowledge of the building environment was acquired through initial self-paced exploration of the building. While participants performed above chance level at identifying the general direction of the key area (i.e., Finance Department) within the building during the floorplan test, this spatial knowledge might have been weakly encoded. They had no advance knowledge of the upcoming SAR mission and had no indication of which parts of the building would become relevant later in the experiment. In other words, their spatial knowledge was encoded in a relatively neutral context. In response to a new floorplan, the utility of which was lightly or heavily implied to be uncertain, most participants relied on it even though it contradicted with their prior knowledge. This suggests that the spatial knowledge, which was acquired incidentally and without clear ties to task goals, was either easily overridden by the new information or was more difficult to access and apply.

In contrast, task goals were clearly communicated to participants before the six SAR training trials in Experiment 3, where they repeatedly rescued the Finance and Marketing Managers. As this learning was reinforced with performance feedback across trials, the training process appears to have established a strong expectation that both managers would be in the building. While those who were explicitly informed of one of the manager's absence changed their search strategy accordingly, those who received the same message in implicit framing did not alter their search behaviour. The present data do not allow a definitive conclusion as to whether this reflected possible constraints in pragmatic processing (i.e., failing to infer the absence of the manager), or persistent influence from prior expectations (i.e., searching for both managers despite having inferred the absence of one manager). Nonetheless, the findings suggest that when prior knowledge is acquired in a task-relevant context, it might hinder adaptive behaviour unless it is directly contradicted by clear and explicitly framed information.

Prior knowledge in Experiment 4 (i.e., self-generated expectations of target prevalence) was acquired in a fashion similar to Experiment 3: through multiple exposures to the same task, leading to strongly encoded expectations. Unlike Experiment 3, however, consistent or complete behavioural adaptation was not observed for those who were clearly instructed to expect “far fewer” targets than experienced earlier in the experiment. In contrast, those who were instructed to expect “a lot more” targets successfully adapted to the task. This discrepancy in performance suggests that well established expectations about target prevalence acquired from repeated perceptual decisions may be somewhat more resistant to update when visual search became more challenging (i.e., when target prevalence was low). Taken together, the strength of prior knowledge could potentially constrain behavioural adaptation regardless of domain, unless its influence was overridden by explicit communication and/or other task-related factors that ease behavioural adaptation.

5.3.3 Framing and Format of Information

Across the experiments, new information was presented differently, which likely affected whether and how participants adapted their behaviour. When participants were informed that the floorplan “might (or might not) be useful” in Experiments 1 and 2, most followed the (inaccurate) floorplan despite the presence of in-situ area labels that could be used for navigation during the SAR mission. When additionally warned that the floorplan “could be outdated”, they were slightly more cautious. While the warning reduced their reliance on the inaccurate floorplan relative to those who received no warning, more than a third still used it. The floorplan, being a formal schematic, likely conferred an implicit authority that deterred additional scrutiny from the participants despite the more explicit warning that the floorplan might be inaccurate.

Unlike the floorplans in the previous experiment, Experiments 3 and 4 provided new information in written and verbal form. In Experiment 3, the stark difference in outcomes between those who received an explicit message versus those who received an implicit message illustrated how explicit messaging impacts participants' ability to adjust established strategies in a high-stakes task. However, the same clear messaging format in Experiment 4 was less effective in a perceptual task, suggesting again an interaction between how explicit the messaging was and task demands.

5.3.4 Timing and Decision Context

Another important distinction was the timing of decision-making, specifically, whether new information could be incorporated into the existing response strategy before being acted upon, or had to be used in real time. In Experiments 1 to 3, participants likely made the key decisions, such as route selection or search decisions, before starting the SAR mission. This planning occurred under minimal time pressure as participants could begin the mission at their own pace. In contrast, in Experiment 4 participants had to make 25 rapid perceptual decisions while navigating the building under 2.5 minutes. Although the instructions were clear and each trial began at their own pace, the demands of the task might have limited their ability to fully adjust their strategy, leaving behaviour to be influenced by prior expectations.

This contrast highlighted the role timing plays in promoting behavioural adaptation. When decisions could be made in advance, new information was more likely to guide behaviour. When decisions were made in real-time and under time pressure, prior knowledge or experience was more likely to continue influencing responses even when they were no longer task-relevant.

5.4 Theoretical Implications

As outlined in the synthesis presented above, the results of the experiments suggest that behavioural adaptation in high-stakes, high-stress scenarios is jointly influenced by message clarity, task structure, prior experience, and time pressure. This section builds on those findings by exploring how they could be interpreted through existing theories and the implications for our understanding of behavioural adaptation in such scenarios.

5.4.1 Challenges to Pragmatic Reasoning in Simulated Operational Settings

The findings from Experiments 1 to 3 suggest that under high-stress and high-stakes conditions, indirectly framed messages appeared less successful than messages framed with greater degrees of explicitness at eliciting behavioural adaptation, even when the information conveyed was critical to task performance. However, even when an additional semantic cue was included (Experiment 2), a notable proportion of participants still failed to process the implicature as intended, suggesting that partial explicitness in communication facilitates but does not ensure accurate pragmatic processing. It could be argued that the additional semantic cue itself contains an implicature (i.e., “could be outdated”), thereby potentially introducing additional interpretive load and making it more challenging to correctly interpret the overall message (i.e., “the floorplan might or might not be useful”, signalling the questionable utility of the floorplan). If so, one might expect this group of participants to perform similarly as those who did not receive the additional cue (i.e., with a large majority relying on the inaccurate floorplan). The fact that the group with the cue was less reliant on it suggests that “could be outdated" made its questionable utility more salient and effectively cautioned against using the floorplan, thereby guiding participants to interpret the overall message correctly.

Relevance Theory (Sperber & Wilson, 1986, 1987) helps explain this pattern by highlighting that pragmatic inference depends on communicative context, and unless the necessary background assumptions to support interpretation are available, implicatures are unlikely to be processed. The high stress conditions under which inferential effort had to be made might have hindered participants' ability to access the relevant contextual assumptions, thereby hindering pragmatic processing in the absence of salient semantic cues, despite the absence of time pressure. While one might expect the high-stakes nature of the task to encourage more deliberate processing due to the significant consequences of decisions (Kahneman, 2011), the present findings suggest that high stakes alone, or the contextual cues of a fire incident, are insufficient to guarantee the engagement of cognitive resources required to support pragmatic processing unless salient semantic cues are present. Even so, a notable minority still failed to interpret implicatures correctly. This aligns with the view that pragmatic reasoning is cognitively demanding, echoing previous literature demonstrating that cognitive constraints (e.g., visuospatial memory load) reduced pragmatic interpretations of scalar implicatures (De Neys & Schaeken, 2007; van Tiel, Pankratz, et al., 2019).

In addition, the absence of other disambiguating non-lexical cues such as emphasis and tone (Chevallier et al., 2008; Jang et al., 2013; Ryzhova & Demberg, 2023) might have also diminished the participants' ability to infer the intended meaning from the indirectly phrased messages. Instead of interpreting "might be useful" as signalling uncertainty about the information provided, the hierarchical nature of the firefighting context, where deference to authority is expected, likely encouraged participants to interpret it as a polite directive (Bonnefon et al., 2011; Bonnefon & Villejoubert, 2006; Brown & Levinson, 1987; Holtgraves, 2014). While pragmatic reasoning is known to be shaped by cognitive capacity and social contexts, the current findings extend this understanding by demonstrating that the operational context alone failed to activate the assumptions required for processing implicatures. Collectively, these insights highlight a boundary condition for pragmatic inference: under high-stress, simulated operational conditions, stress-induced cognitive constraints and hierarchical expectations may limit pragmatic processing, which is otherwise carried out with little effort in everyday life. Under such conditions, new but critical information, when implicitly framed, may fail to support behaviour adaptation if it is not interpreted as intended.

5.4.2 Perceptual Decision-Making

Experiment 4 presented an interesting contrast to the previous experiments: whereas significant behavioural adaptation was observed when new information was conveyed explicitly in Experiment 3, such adaptation was less robust during the high-stakes visual search task in Experiment 4, even though instructions were similarly explicit. This pattern of results points to an important distinction in how prior expectations interact with clearly communicated information in visual search tasks compared to more deliberative forms of decision-making. Consistent with past literature (Ishibashi et al., 2012; Experiments 1a and 1b, Ishibashi & Kita, 2014; Lau & Huang, 2010), visual search performance remained influenced by self-generated expectations acquired through prior training even though explicit instructions were provided to calibrate expectations. While motivational incentives (e.g., monetary rewards) has been shown to reduce certain search-related biases, such as those related to target prevalence or saliency (Hadjipanayi et al., 2023; Navalpakkam et al., 2010), the high-stakes firefighting context simulated in the present experiment appeared insufficient to fully override outdated expectations. This observation suggests that high stakes operational contexts alone are insufficient to motivate recalibration of the starting point of evidence accumulation towards a decision boundary when evaluating stimuli (Peltier & Becker, 2016; Wolfe & Van Wert, 2010). This is consistent with Hon and Jabar's (2018) finding that adjusting to a new target prevalence required sampling of approximately 10 targets (regardless of intervening distractors) before prevalence expectation updated to reliably guided visual search.

The fact that increased false negatives was not observed under low versus high target prevalence is notable. As discussed in Chapter 4 (Section 4.5), this deviation from the low prevalence effect (Wolfe et al., 2005, 2007) – or an analogous effect – might reflect task framing and the asymmetry of visual search performance and task performance: Making false positives incurred minimal cost, whereas making false negatives directly undermined the rescue. The overall findings might reflect this strategic prioritisation of task goals over strict search accuracy. Thus, when new information conflicts with prior experience, behavioural adaptation might only partially incorporate that information insofar as prioritised task goals are not compromised, rather than fully incorporating it for optimal performance (e.g., when target prevalence decreased). However, when task goals align with strict search accuracy (e.g., when target prevalence increased), behavioural adaptation fully incorporates new information. If this were to be interpreted through the lens of conservation of cognitive effort, it is possible that participants incorporated new information only to the extent required for task success whilst avoiding the additional cognitive costs of fully revising established expectations. This aligns with accounts suggesting that cognitive control is only engaged when doing so confers sufficient benefits relative to effort costs (Shenhav et al., 2013, 2017; Sidarus et al., 2019).

5.4.2.1 Task-Specific Factors and Prevalence Effects. The low number of false negatives across conditions in Experiment 4 might also reflect the specific design of the visual search task. In single-target searches, the low prevalence effect was observed in searches where rarely-occurring target-present trials interleaved between target-absent trials (Wolfe et al., 2005, 2007). In contrast, all trials were target-present in the current study and low prevalence was defined by relatively fewer targets per trial. As participants never encountered target-absent trials, Experiment 4 did not reproduce the standard search context that produces the low prevalence effect. Nevertheless, each trial consisted of five mini-trials, as the stimuli were distributed across five rooms in the virtual building, with at least two of these always being target-absent rooms. Thus, while participants expected at least one target somewhere per trial, they did not expect each room to contain a target. When target-absent rooms were not anomalous, it is therefore reasonable to expect miss errors from target-present rooms.

When target prevalence was high, between five to seven targets were randomly distributed in a maximum of three rooms, with a maximum of three targets in one room. This created more multi-target rooms which increased the susceptibility to Subsequent Search Misses (SSM), where identifying one target reduces the likelihood of identifying another. However, greater prevalence of multi-target relative to single-target mini-trials (Adamo et al., 2022; Cheng & Rich, 2018) and the use of perceptually identical targets (Gorbunova, 2017) might have reduced this susceptibility, contributing to low number of false negatives. When target prevalence was low, one to three targets were randomly distributed to a maximum of three rooms, thus creating more single-target rooms, which could have produced effects similar to the low prevalence effect through premature search termination. The lack of this effect might have stemmed from the generalisation of the strategy acquired during training to the experiment proper. Specifically, during training, participants were instructed to compare cylinder sizes to learn to identify targets. This strategy resembles similarity search where, instead of making binary absent/present decisions, the search involved locating the stimulus that most resembled the target (Taylor et al., 2022). Relative to absent/present search, similarity search reduces the low prevalence effect, possibly by maintaining a high search termination threshold or/and reducing the accumulation of evidence towards that threshold (Taylor et al., 2022). Likewise, in the present study, participants' training might have biased them to judge whether each cylinder was worth removing, potentially sustaining their search efforts and reducing false negatives. This strategic bias offers a theoretical explanation for the low levels of false negatives in Experiment 4.

5.4.3 When Does New Information Guide Behavioural Adaptation?

Across all experiments, successful behavioural adaptation depended on the framing and clarity of new information, and when that information needed to be acted upon. However, while new information was communicated with comparable clarity in Experiment 3 (Explicit-message condition) and Experiment 4, behavioural adaptation was consistently observed only in Experiment 3. This contrast might have reflected how task ambiguity was resolved and how this might have in turn influenced the way participants balanced their subjective evaluations of speed and accuracy within the constraints of each task. In Experiment 3, directly informing participants that a manager was not in the building removed uncertainty and eliminated the need for inference. The best course of action became clear: Searching the area where the absent person was expected would sacrifice speed in rescuing the remaining person, without offering any potential benefits regarding accuracy (i.e., reducing the risk of a missed rescue). In other words, participants did not have to make speed-accuracy trade-offs given that the optimal decision, whether subjectively or objectively weighed, was to search only the area where the remaining trapped person was expected to be.

In contrast, while the communication in Experiment 4 was unambiguous about changes (or lack thereof) in target prevalence, it likely had little effect on reducing the variation in how participants subjectively valued completing the visual search task quickly versus avoiding a missed explosive target. Changes in the decision criterion might have been subjected to the relative weights placed on speed versus accuracy during visual search, which varied between participants based on their subjective valuations and prior experience. For example, among participants tested with low target prevalence, there was greater variability in the total number of cylinders removed and the number of false positives among participants who were initially trained under high target prevalence compared to those trained under low target prevalence. This suggests that some participants who were trained under high target prevalence prioritised speed and adapted their decision criterion more completely by removing “far fewer cylinders than before” as instructed, thereby reducing the time taken to rescue the trapped person. Conversely, others might have prioritised accuracy in terms of avoiding missed targets and relied more heavily on their prior experience, resulting in a more liberal decision criterion but delays in rescuing the trapped person. Therefore, when making (repeated) perceptual decisions under time pressure, individual differences in valuations of task priorities might have weakened the impact of clear communication.

In summary, these findings support the view that behavioural adaptation in high-stakes, high-stress conditions is determined not only by the clarity of new information or the strength of prior knowledge, but also by their interaction with task structure, cognitive demands, and the timing of decisions. Effective integration of new information occurs when the new information has minimal inference load, and when the task affords time and cognitive capacity for strategy revision. The present findings contribute empirical validation of these boundary conditions.

5.5 Practical Implications

The findings of this thesis underscore the importance of clear and explicit communication in high-stakes, high stress operational contexts such as search and rescue operations. Although existing communication guidelines in such domains already emphasise directness and clarity, evidence suggests that individuals under cognitive load might nonetheless introduce unintended ambiguity when communicating. For instance, Kurinec et al. (2019) demonstrated that cognitive constraints reduced participants' ability to avoid ambiguity when giving instructions, even after they were encouraged to communicate more clearly by asking them to take the listener's perspective. A similar risk might arise in operational settings from the use of pragmatic implicatures that are commonly found in everyday communication. Moreover, because such implicatures are typically interpreted effortlessly in daily communication, some individuals might mistakenly assume their indirect communication to be sufficiently direct in cognitively demanding situations. The present findings highlight the risk of using such language as it presumes inferential effort or shared assumptions that might be unavailable or inaccessible to the recipient. The behavioural consequences of this limitation were evident in Experiments 1 and 2, where participants did not infer that they should verify the floorplan before use unless additionally cued, and Experiment 3, where they did not infer the absence of a to-be-rescued person. Thus, the assumption that information, once conveyed, is automatically interpreted and acted upon as intended is untenable, even when the intended meaning would normally be easily inferred in daily communication.

These insights have important practical implications for operational communication. Organisations should reinforce, and re-evaluate where necessary, the extent to which existing communication training addresses the limitations of everyday communication habits in high-stakes contexts. In particular, training programmes should highlight the potential pitfalls of using language that, while typically understood in everyday communication, might lead to miscommunication during high pressure incidents. Under stress or cognitive load, individuals might be limited in their ability to process implicatures that would otherwise be easily understood, potentially leading to misinterpretation of the communicator's intent and serious consequences. Given the risk of unintentionally introducing ambiguity during communication (Kurinec et al., 2019), training should therefore emphasise the use of explicit language rather than assuming shared inferences. This could include developing the ability to quickly recognise the use of implicit language and respond with immediate clarification. For example, if “the floorplan might or might not be useful” was uttered, it should be immediately recognised and followed up with directly actionable communication, such as, “so we need to verify it against the labels in the building when we’re inside,”, or, “so we need to verify it against what we know about the building”.

5.6 Limitations and Future Directions

The present experiments are, to my knowledge, the first to investigate how language use and prior knowledge affect behavioural adaptation in different simulated operational contexts. A key strength of this research lies in the use of virtual reality (VR), either implemented on desktop computers or within a semi-immersive VR environment, which not only greatly enhanced feelings of presence and immersion in the experiments but also allowed precise standardisation of the virtual environment. This platform provided a safe and systematic way to investigate how different factors influence behavioural adaptation in a high-stakes operational context, a feat that would be more challenging to replicate in real-world settings. However, this high level of experimental control also introduced limitations that should be considered when interpreting the findings and addressed in future research.

Specifically for Experiments 1 and 2, the primary outcome measured whether the first turn in the building was towards the general direction of the Finance Department as indicated on the floorplan. When participants turned in the opposite direction, this binary measure limits the ability to attribute their navigational decision to either verification of the floorplan in response to the implicature (e.g., recalling prior knowledge or using in-situ area labels) or a reaction to the suggestion that the floorplan might be inaccurate (i.e., turn in the opposite direction suggested by the floorplan). Moreover, potentially weak encoding of the spatial layout during the exploration phase makes it difficult to determine whether participants were able to activate prior knowledge to verify new information under high pressure conditions. Extensions of Experiments 1 and 2 should not only encourage more thorough spatial encoding during exploration or provide floorplans that violate more widely held background assumptions about building layouts, such as the typical locations of toilets in a building (Frankenstein et al., 2012; Gath-Morad et al., 2024), but also use a building that is more spatially complex (e.g., a multistorey building) to enable a more granular analysis of navigational choices. A more complex environment with multiple plausible routes would make it possible to draw sharper inferences for participants who do not use the floorplan: whether they are relying on prior knowledge, using in-situ area labels, or blindly reacting to the warnings of inaccuracies.

In addition, because the present findings showed that conversational implicatures were not always interpreted correctly under high-stakes conditions (which are often stressful), future work should include both high- and low-stakes versions of the task to determine whether the reducing pressure would facilitate implicature processing. For example, in future experiments the cover story could include such shared elements in both conditions: The Assistant Finance Manager, who has a mobility impairment, has been working overtime when the electricity cut off and the fire alarm was triggered. Participants in the high-stakes condition would be informed that the Assistant Finance Manager also reported a strong chemical smell, possibly from overheating servers, and that monitoring systems indicated a small fire in the server room. Those in the low-stakes condition would instead be informed that the small fire in the server room had been contained by the building's suppression system, but because it was very dark, the Assistant Finance Manager, who was in no immediate danger, needed assistance to move safely to the exit. Both groups of participants would then be given an inaccurate floorplan and told that the floorplan "might or might not be useful" for navigation in the building, allowing researchers to examine how perceived stakes affect both implicature interpretation and spatial behaviour.

5.6.1 Improving the Generalisability of the Present Findings

The use of undergraduate participants in the present experiments demonstrated how prior knowledge, new knowledge, and message framing interact to influence behavioural adaptation in high-stakes, high stress situations within an untrained population. This approach provides a baseline against which the effects of experience and training in high pressure domains (e.g., firefighting) could be compared. It could thereby enable future research to distinguish aspects of behavioural adaptation that are changeable through experience and training, versus those that may be more constrained by inherent limitations in adaptive capacity. However, most of the null results were supported by Bayes factors showing only anecdotal evidence for the null, likely reflecting insufficient power. These findings should therefore be interpreted with caution, and replication with larger samples is needed to confirm their robustness. To understand how expertise, experience, and new information interact with message framing to shape behavioural adaptation under pressure, it is also essential to replicate the present experiments with both novice and veteran firefighters.

On the other hand, firefighters have also been shown to rely on standard operating procedures when the situation called for deliberative processing (Butler et al., 2023), suggesting that the failure in pragmatic reasoning, which depends on such deliberative processes (Moeschler, 2023), could also emerge in trained populations. Moreover, according to the Recognition-Primed Decision model, firefighters rely strongly on pattern recognition to make rapid decisions based on familiar cues (Klein et al., 2010). This reliance might reduce behavioural adaptation when the integration of new and indirectly framed information is required. Although firefighters spent similar time processing information under high and low stress conditions, those under high stress made faster decisions and spent less time to gain situational awareness in a simulated fire incident (Keren et al., 2013). This pattern potentially reflected a greater reliance on prior knowledge to interpret the situation rapidly, which might reduce the likelihood of engaging in deliberative processes when new information is presented indirectly, such as through implicatures, thereby increasing their vulnerability to framing effects. However, there is also evidence showing that more experienced firefighters reviewed more information and took longer to reach a decision compared to their less experienced counterparts (Bayouth et al., 2013), suggesting that expertise might support more deliberate processing. Thus, whether the present findings would generalise to firefighters of varying experience remains an open question.

5.6.2 Salience and Use of Navigational Cues

The low visibility conditions in Experiments 1 and 2 might have also discouraged the use of visual information in the building (i.e., plaques with area labels). These plaques on the walls or doors were inconspicuous from a distance but became clearer as participants approached. Given the limited evidence that these plaques were used to guide navigation, I propose that once participants decided that they would rely on the floorplan, they might have formulated and followed a simple directional strategy (e.g., "keep to the right to the Finance Department"). It would be worthwhile to investigate whether the limited use of in-situ area labels to verify a potentially inaccurate floorplan stemmed from overconfidence in a formal schematic, which might have led them to overlook plaques that were only visible up close. Alternatively, not using in-situ area labels might also stem from reduced cognitive capacity caused by environmental stressors (e.g., alarm, fire), that limited their ability to exploit easily accessible visual information even when nearby. One option would be to manipulate the presence of environmental stressors, the visibility of the area labels (e.g., fluorescent plaques versus non-fluorescent plaques), and floorplan presentation (e.g., professionally drawn versus hand-drawn) to better understand their individual and joint effects on the interplay between prior knowledge, new information, and navigation behaviour.

5.6.3 Reduction or Redirection of Pragmatic Processes?

Another avenue for future research concerns the competing interpretations of the findings from Experiment 1. As described in Chapter 2 (Section 2.6), the apparent reduction in pragmatic reasoning when salient semantic cues (e.g., additional lexical information, such as describing the floorplan as "could be outdated") were absent could be understood through the lens of Relevance Theory (Sperber & Wilson, 1986, 1987) and Politeness Theory (Brown & Levinson, 1987). Relevance Theory suggests that the lack of non-lexical cues (e.g., emphasis and tone) hindered pragmatic reasoning, whereas Politeness Theory suggests that the participant did engage in pragmatic reasoning but interpreted the hedged phrase as a polite directive to use the floorplan. In the present context, these competing perspectives diverged in meaningful ways. While Relevance Theory implies a reduction in pragmatic processing under high-stakes, high-stress conditions, Politeness Theory implies that pragmatic processing was intact under such conditions but was guided by or redirected towards social cues such as perceived authority or politeness norms. To test these interpretations, future work could present the inaccurate floorplan and utilise a 2 (Emphasis: Emphasis, No Emphasis) by 2 (Authority: High Authority, Low Authority) factorial design, with the following messages proposed for each condition:

- Emphasis, High Authority: "The Sector Commander provided the following floorplan. He said the floorplan ***MIGHT*** be useful for the search and rescue mission."

- No Emphasis, High Authority: “The Sector Commander provided the following floorplan. He said the floorplan might be useful for the search and rescue mission.”
- Emphasis, Low Authority: “The Building Security Officer provided the following floorplan. He said the floorplan ***MIGHT*** be useful for the search and rescue mission.”
- No Emphasis, Low Authority: “The Building Security Officer provided the following floorplan. He said the floorplan might be useful for the search and rescue mission.”

If the lack of non-lexical cues is what limited pragmatic reasoning, then those who received a message without an emphasis on the word “might” should be more likely to follow the floorplan during the SAR mission than those who received a message with the emphasis. In this context, emphasis on the hedging term “might” could highlight the messenger’s lack of confidence in the utility of the floorplan, thus encouraging participants to interpret this emphasis as an intentional signal to exercise caution before using it. If instead pragmatic reasoning is redirected by social cues (e.g., perceived authority), then those who received a message from the Sector Commander should be more likely to follow the floorplan than those who received the same message from the Building Security Officer. It is also possible that the influence of emphasis and authority interact. For example, participants might be more sensitive to the emphasised hedged word when the message was from the Sector Commander than when it is from the Building Security Officer, potentially indicating greater attention and deference given to someone with higher authority. Conversely, emphasis might have little additional impact when the message is from the Building Security Officer.

5.6.4 Are Non-Lexical Cues Necessary to Trigger Pragmatic Processing?

Another point of contention concerns the relative influence of the absence of non-lexical cues versus cognitive constraints on pragmatic processing limitations. Although participants received the new messages during a relatively quiet decision window and without time pressure in Experiments 1 to 3, anticipatory stress might have hindered pragmatic reasoning due to impaired working memory and attentional control (Bambini et al., 2021; Cain et al., 2011; Hyun et al., 2019). In other words, it is plausible that even without acute stressors, the narrative context building up to the SAR mission might have been sufficient to trigger pragmatic processing, but anticipatory stress prevented its full execution. Support for this interpretation comes from literature showing that firefighters have anticipatory increases in cortisol levels before attending a demonstration of fire extinguishers, receiving training to use the self-contained breathing apparatus (SCBA), or completing a SAR exercise (Robinson et al., 2013). Moreover, external administration of cortisol has been associated with impaired working memory but not other executive functions, such as improved ability to suppress dominant responses (i.e., inhibition) but unchanged ability to adapt behaviour in response to changing task demands (i.e., cognitive flexibility) (Shields et al., 2015). Nevertheless, Shields et al. (2015) proposed that elevated cortisol levels elicited by naturalistic stress responses would interact with other biological processes, which might amplify the effects observed under isolated cortisol administration. Together, these results suggest that while the narrative context alone in Experiments 1 to 3 might have initiated pragmatic reasoning, anticipatory stress might have constrained the cognitive resources required to support its execution.

To test this suggestion, a 2 (Emphasis: Emphasis, No Emphasis) by 4 (Constraints: Anticipatory Stress, Concurrent Task, Sensory Stress, No Stress) factorial design could be employed to investigate whether the correct inference would be drawn from the implicitly framed message in Experiment 3, for example. In the Emphasis condition, the message might be formatted as such: "The Finance Department is holding their ***TEAMBUILDING*** event in a ***NEARBY HOTEL***"; in the No Emphasis condition, the same message would be presented without capitalising the words. As I proposed that the procedures used in Experiment 3 induced anticipatory stress, the Anticipatory Stress condition in future work could replicate these procedures. Specifically, the cover story and SAR mission environment should remain unchanged, with participants being informed that the upcoming trial is "real" and exposed to the same audio-visual stressors. The Concurrent Task condition serves as a comparison to determine whether the behavioural outcomes under divided attention resemble those observed under anticipatory stress. In this condition, the virtual environment during SAR includes the same audio-visual stressors, but the cover story is modified: Participants could be told that the upcoming trial is an additional practice trial in which none of the trapped persons is in actual danger. They could be informed that this additional practice trial is included to test a new augmented reality (AR) module integrated with the SCBA visor, which enables simulation of realistic visual effects during training. The concurrent task could also be framed as a password to activate the AR module. Specifically, participants would be required to memorise a complex 4 x 4 pattern matrix prior to receiving the implicitly-framed message, and recall it before they commence the trial (van Tiel, Pankratz, et al., 2019). Similarly, the Sensory Stress condition removes the anticipatory stress of a "real" SAR mission by using the same cover story and SAR environment as the Concurrent Task condition, but without the concurrent memory task. While this necessarily weakens the perceived stakes of the task, it is still embedded within a gamified firefighting training and equipment testing context with post-trial feedback, which likely sustains some level of personal stakes. Lastly, the No Stress control condition introduces the upcoming trial as merely another training session and devoid of audio-visual stressors. These two conditions would allow an assessment of whether or not the firefighting context is itself sufficient to trigger pragmatic reasoning even without non-lexical cues in the message communicated.

I hypothesise that a main effect of Emphasis will be observed, with those in the Emphasis condition more likely to engage in pragmatic reasoning and correctly infer the absence of the manager than those in the No Emphasis condition. However, this effect might be less pronounced in the Sensory Stress condition. If anticipatory stress introduces cognitive constraints comparable to those induced by a dual-tasking scenario, participants in the Anticipatory Stress and Concurrent Task conditions will be less likely to make the correct pragmatic inference compared to those in the Sensory Stress condition. Those in the No Stress condition would be the most likely to make the correct pragmatic inference. If the presence of non-lexical cues could compensate for reduced cognitive capacity, participants in the Emphasis condition should be more likely to make the correct inference than those in the No Emphasis condition within the Anticipatory Stress and Concurrent Task conditions. Conversely, if non-lexical cues do not offset cognitive constraints, there would be little difference in the likelihood of making correct inferences between the Emphasis and No Emphasis groups within those conditions. These outcomes would be able to clarify whether pragmatic processing in high-stakes scenarios requires non-lexical cues or whether it can be triggered by contextual factors alone, thus advancing our understanding of how pragmatic processes operate in applied, high-stakes settings.

5.6.5 Persistence of Self-Generated Expectations: Strategic Choice or Cognitive Limitation?

When direct and explicit communication is provided, the incomplete behavioural adaptation observed during visual search in Experiment 4 poses an interesting question: When do self-generated expectations stop their dominant influence in guiding behaviour? Existing visual search literature proposed that the persistent influence of self-generated expectations of target prevalence stems from the inherent properties of visuals search (Hon & Jabar, 2018; Ishibashi et al., 2012; Lau & Huang, 2010). However, the asymmetry in search performance between participants who were told that target prevalence would decrease and those who were told that it would increase suggests that such expectations might not be equally resistant to updating in all cases. This difference might instead reflect individual differences in task priorities or strategic considerations, rather than a persistent influence of prior expectations in high-stakes visual search. Future work could seek to standardise task priorities by emphasising either speed or accuracy during high-stakes visual search, thus reducing the variation in subjective valuations of these competing demands. Moreover, another important consideration in interpreting the findings from Experiment 4 (and indeed, Experiments 1-3) is the absence of consequential penalties for incorrect decisions, which might have promoted a satisficing strategy that limited the extent of behavioural adaptation.

Experiment 4 could be replicated and extended to investigate whether varying the cost of mistakes will alter the influence of self-generated expectations of target prevalence versus instructed prevalence in search performance. For instance, the same experiment could be conducted under three different conditions: one where false positive and false negatives are penalised equally (e.g., by subtracting five seconds per error from the countdown timer and/or triggering an aversive sound), one where only false positives are penalised, and one where only false negatives are penalised. If participants adapted their behaviour strategically, I would expect to observe differential patterns of adaptation based on different penalty structures. For example, when both error types are penalised, we might observe the most thorough adaptation to the instructed prevalence regardless of whether participants were instructed to expect an increase, decrease, or no changes in target prevalence. Conversely, penalising only false positives might result in a more conservative approach. This could lead to more complete adaptation for those who were instructed to expect reduced target prevalence (as avoiding false positives aligns with the instruction), but more false negatives among those who were instructed to expect increased target prevalence (as they might be overly cautious when uncertain about the identity of targets). The opposite trend might be observed when only false negatives are penalised. Likewise, a similar manipulation could be applied to Experiment 3 to emphasise the consequence of incorrect decisions. The proposed extensions would be invaluable in clarifying whether overreliance on self-generated expectations or prior experience of target prevalence during high-stakes visual search represents a hardwired limitation in perceptual decision-making, cognitive inflexibility induced by time-pressure, or a strategic trade-off influenced by motivation and perceived costs of errors.

5.7 Concluding Remarks

The research reported in this thesis, situated at the intersection of theory and application, was motivated by a central practical question: When do individuals adapt their goal-directed behaviour in high-stakes, high-stress contexts in response to newly received information that contradicts their prior knowledge or experience? To answer this question, I developed experimental platforms incorporating virtual environments that replicate firefighting conditions, used across four experiments on desktop computers and within a semi-immersive virtual reality setting. Over the course of this work, it became clear that in emergency incidents, the provision of goal-critical information does not necessarily guarantee behavioural adaptation towards achieving stated goals. Behavioural adaptation appeared to be most likely when information was conveyed explicitly, had minimal interpretive load, and could be integrated to update action plans. However, behavioural adaptation was less likely when self-generated expectations were more ingrained or when decisions had to be made in real time under time pressure, even when critical information was conveyed in explicit language. Of course, these conclusions require further corroboration through replication with career firefighters, but they nonetheless identify candidate factors that might affect the uptake of new information in high-stakes contexts. Understanding how these factors interact under cognitive constraints can help inform training protocols that emphasise explicit communication and mitigate the influence of ingrained expectations on goal-directed behaviour.

References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2013). Self-induced attentional blink: A cause of errors in multiple-target search. *Psychological Science*, *24*(12), 2569–2574. https://doi.org/10.1177/0956797613497970

Adamo, S. H., Gereke, B. J., Shomstein, S., & Schmidt, J. (2021). From “satisfaction of search” to “subsequent search misses”: A review of multiple-target search errors across radiology and cognitive science. *Cogn Res Princ Implic*, *6*(1), 1–19.

Adamo, S. H., Patel, A., & Ortiz, M. (2022). Are you really satisfied? How overall prevalence and relative-prevalence of single-to-dual target trials affects multiple-target search misses. *Journal of Vision*, *22*(14), 4373. https://doi.org/10.1167/jov.22.14.4373

Alexander, J. K., Hillier, A., Smith, R. M., Tivarus, M. E., & Beversdorf, D. Q. (2007). Beta-adrenergic modulation of cognitive flexibility during stress. *Journal of Cognitive Neuroscience*, *19*(3), 468–478. https://doi.org/10.1162/jocn.2007.19.3.468

Anderson, B. A. (2024). An examination of the motivation to manage distraction. *Cognition*, *250*, 105862. https://doi.org/10.1016/j.cognition.2024.105862

Anderson, B. A., & Lee, D. S. (2023). Visual search as effortful work. *Journal of Experimental Psychology: General*, *152*(6), 1580–1597. https://doi.org/10.1037/xge0001334

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, 521–543. https://doi.org/10.1177/0149206314560412

Arcara, G., & Bambini, V. (2016). A test for the Assessment of Pragmatic Abilities and Cognitive Substrates (APACS): Normative data and psychometric properties. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00070

Attar, N., Schneps, M. H., & Pomplun, M. (2016). Working memory load predicts visual search efficiency: Evidence from a novel pupillary response paradigm. *Memory & Cognition*, *44*(7), 1038–1049. https://doi.org/10.3758/s13421-016-0617-8

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baker, C., Pawling, R., & Fairclough, S. (2020). Assessment of threat and negativity bias in virtual reality. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-74421-1

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419. https://doi.org/10.1016/S0028-3908(98)00033-1

Bambini, V., Van Looy, L., Demiddele, K., & Schaeken, W. (2021). What is the contribution of executive functions to communicative-pragmatic skills? Insights from aging and different types of pragmatic inference. *Cognitive Processing*, *22*(3), 435–452. https://doi.org/10.1007/s10339-021-01021-w

Bayouth, S. T., Keren, N., Franke, W. D., & others. (2013). Examining firefighter decision-making: How experience influences speed in process and choice. *International Fire Service Journal of Leadership Management*, *7*.

Beitner, J., Helbing, J., David, E. J., & Võ, M. L.-H. (2024). Using a flashlight-contingent window paradigm to investigate visual search and object memory in virtual reality and on computer screens. *Scientific Reports*, *14*(1), 8596. https://doi.org/10.1038/s41598-024-58941-8

Bennett, C. R., Bex, P. J., & Merabet, L. B. (2021). Assessing visual search performance using a novel dynamic naturalistic scene. *Journal of Vision*, *21*(1), 5. https://doi.org/10.1167/jov.21.1.5

Berbaum, K. S., Franken, E. A., Dorfman, D. D., Rooholamini, S. A., Coffman, C. E., Cornell, S. H., Cragg, A. H., Galvin, J. R., Honda, H., Kao, S. C. S., Kimball, D. A., Ryals, T. J., Sickels, W. J., & Smith, T. P. (1991). Time course of satisfaction of search. *Investigative Radiology*, *26*(7), 640–648. https://doi.org/10.1097/00004424-199107000-00003

Berger, A., Mitschke, V., Dignath, D., Eder, A., & van Steenbergen, H. (2020). The face of control: Corrugator supercilii tracks aversive conflict signals in the service of adaptive cognitive control. *Psychophysiology*, *57*(4), e13524. https://doi.org/10.1111/psyp.13524

Biggs, A. T., Adamo, S. H., & Mitroff, S. R. (2014). Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. *Acta Psychologica*, *152*, 158–165. https://doi.org/10.1016/j.actpsy.2014.08.005

Bissett, P. G., Jones, H. M., Hagen, M. P., Bui, T. T., Li, J. K., Rios, J. A. H., Mumford, J. A., Shine, J. M., & Poldrack, R. A. (2023). A dual-task approach to inform the taxonomy of inhibition-related processes. *Journal of Experimental Psychology. Human Perception and Performance*, *49*(3), 277–289. https://doi.org/10.1037/xhp0001073

Bogdanov, M., Nitschke, J. P., LoParco, S., Bartz, J. A., & Otto, A. R. (2021). Acute psychosocial stress increases cognitive-effort avoidance. *Psychological Science*, *32*(9), 1463–1475. https://doi.org/10.1177/09567976211005465

Bonnefon, J.-F., Feeney, A., & De Neys, W. (2011). The risk of polite misunderstandings. *Current Directions in Psychological Science*, *20*(5), 321–324. https://doi.org/10.1177/0963721411418472

Bonnefon, J.-F., & Villejoubert, G. (2006). Tactful or doubtful?: Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, *17*(9), 747–751. https://doi.org/10.1111/j.1467-9280.2006.01776.x

Botch, T. L., Garcia, B. D., Choi, Y. B., Feffer, N., & Robertson, C. E. (2023). Active visual search in naturalistic environments reflects individual differences in classic visual search performance. *Scientific Reports*, *13*(1), 631. https://doi.org/10.1038/s41598-023-27896-7

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. https://doi.org/10.1016/j.jml.2004.05.006

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463. https://doi.org/10.1016/j.cognition.2005.07.003

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behaviour with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, *52*(2), 455–463. https://doi.org/10.3758/s13428-019-01242-0

Brosschot, J. F., Gerin, W., & Thayer, J. F. (2006). The perseverative cognition hypothesis: A review of worry, prolonged stress-related physiological activation, and health. *Journal of Psychosomatic Research*, *60*(2), 113–124. https://doi.org/10.1016/j.jpsychores.2005.06.074

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge University Press.

Bruning, A. L., Mallya, M. M., & Lewis-Peacock, J. A. (2023). Rumination burdens the updating of working memory. *Attention, Perception, & Psychophysics*, *85*(5), 1452–1460. https://doi.org/10.3758/s13414-022-02649-2

Butler, P. C., Bowers, A., Smith, A. P., Cohen-Hatton, S. R., & Honey, R. C. (2021). Decision making within and outside standard operating procedures: Paradoxical use of operational discretion in firefighters. *Hum. Factors*, 187208211041860.

Butler, P. C., Bowers, A., Smith, A. P., Cohen-Hatton, S. R., & Honey, R. C. (2023). Decision making within and outside standard operating procedures: Paradoxical use of operational discretion in firefighters. *Human Factors*, *65*(7), 1422–1434. https://doi.org/10.1177/00187208211041860

Cain, M. S., Biggs, A. T., Darling, E. F., & Mitroff, S. R. (2014). A little bit of history repeating: Splitting up multiple-target visual searches decreases second-target miss errors. *Journal of Experimental Psychology: Applied*, *20*(2), 112–125. https://doi.org/10.1037/xap0000014

Cain, M. S., Dunsmoor, J. E., LaBar, K. S., & Mitroff, S. R. (2011). Anticipatory anxiety hinders detection of a second target in dual-target search. *Psychological Science*, *22*(7), 866–871. https://doi.org/10.1177/0956797611412393

Cain, M. S., & Mitroff, S. R. (2013). Memory for found targets interferes with subsequent performance in multiple-target visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1398–1408. https://doi.org/10.1037/a0030726

Campbell, J., & Ehlert, U. (2012). Acute psychosocial stress: Does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, *37*(8), 1111–1134. https://doi.org/10.1016/j.psyneuen.2011.12.010

Cheng, P. (Xin), & Rich, A. N. (2018). More is better: Relative prevalence of multiple targets affects search accuracy. *Journal of Vision*, *18*(4), 2. https://doi.org/10.1167/18.4.2

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology*, *61*(11), 1741–1760. https://doi.org/10.1080/17470210701712960

Cohen, B. H., Davidson, R. J., Senulis, J. A., Saron, C. D., & Weisman, D. R. (1992). Muscle tension patterns during auditory attention. *Biological Psychology*, *33*(2), 133–156. https://doi.org/10.1016/0301-0511(92)90028-S

Cox, P. H., Kravitz, D. J., & Mitroff, S. R. (2021). Great expectations: Minor differences in initial instructions have a major impact on visual search in the absence of feedback. *Cogn Res Princ Implic*, *6*(19), 1–14.

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, 135–142. https://doi.org/10.1145/166117.166134

Curci, A., Lanciano, T., Soleti, E., & Rimé, B. (2013). Negative emotional experiences arouse rumination and affect working memory capacity. *Emotion*, *13*(5), 867–880. https://doi.org/10.1037/a0032492

Cyberith. (2020, February 27). The Cyberith SDK for VR Locomotion & Haptic Feeback: Unity & Unreal Engine. *Cyberith Virtualizer VR Treadmills*. https://medium.com/cyberith-virtualizer-vr-treadmills/the-cyberith-sdk-for-vr-locomotion-haptic-feeback-unity-unreal-engine-96370a40d957

Czarnek, G., Strojny, P., Strojny, A., & Richter, M. (2020). Assessing engagement during rescue operation simulated in virtual reality: A psychophysiological study. *International Journal of Human–Computer Interaction*, *36*(5), 464–476. https://doi.org/10.1080/10447318.2019.1655905

Dammen, L. van, Finseth, T. T., McCurdy, B. H., Barnett, N. P., Conrady, R. A., Leach, A. G., Deick, A. F., Van Steenis, A. L., Gardner, R., Smith, B. L., Kay, A., & Shirtcliff, E. A. (2022). Evoking stress reactivity in virtual reality: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *138*, 104709. https://doi.org/10.1016/j.neubiorev.2022.104709

de Morree, H. M., & Marcora, S. M. (2012). Frowning muscle activity and perception of effort during constant-workload cycling. *European Journal of Applied Physiology*, *112*(5), 1967–1972. https://doi.org/10.1007/s00421-011-2138-2

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*(5), 428–433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177/0963721419855658

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133. https://doi.org/10.1027/1618-3169.54.2.128

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology*, *27*(5), 481–494. https://doi.org/10.1111/j.1469-8986.1990.tb01962.x

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., Goldenstein, S. K., Vogler, C., Venkataraman, S., & Metaxas, D. N. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviat Space Environ Med.*, *76*(6 Suppl).

Doroudian, S., Wu, Z., Wang, W., Galati, A., & Lu, A. (2022). A study of real-time information on User behaviors during search and rescue (SAR) training of firefighters. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 387–394. https://doi.org/10.1109/VRW55335.2022.00085

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3).

Edwards, A., & Kelly, D. (2017). Engaged or frustrated? Disambiguating emotional state in search. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 125–134. https://doi.org/10.1145/3077136.3080818

Effective VR Treadmill—Cyberith Virtualizer R&D Kit. (n.d.). Retrieved 25 March 2025, from https://www.cyberith.com/virtualizer-rd-kit/

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Personality Processes and Individual Differences*, *58*(2), 342–353.

Elor, A., Powell, M., Mahmoodi, E., Hawthorne, N., Teodorescu, M., & Kurniawan, S. (2020). On shooting stars: Comparing CAVE and HMD immersive virtual reality exergaming for adults with mixed ability. *ACM Transactions on Computing for Healthcare*, *1*(4), 1–22. https://doi.org/10.1145/3396249

emteq labs. (2022, November 18). *Unity SDK Manual (V2) · Emteq Labs Support Docs*. https://support.emteqlabs.com/emteq-sdk/unity.html

emteq labs, & Mavridou, I. (2025, February 28). *Sample Analysis Scripts · Emteq Labs Support Docs*. Emteq Labs Support Docs. https://support.emteqlabs.com/02_emteqpro/04_data/sample-analysis.html

Engert, V., Efanov, S. I., Duchesne, A., Vogel, S., Corbo, V., & Pruessner, J. C. (2013). Differentiating anticipatory from reactive cortisol responses to psychosocial stress. *Psychoneuroendocrinology*, *38*(8), 1328–1337. https://doi.org/10.1016/j.psyneuen.2012.11.018

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors*, *59*(5), 734–764. https://doi.org/10.1177/0018720817690639

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Evans, K. K., Birdwell, R. L., & Wolfe, J. M. (2013). If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening. *PLoS ONE*, *8*(5), e64366. https://doi.org/10.1371/journal.pone.0064366

Evans, K. K., Tambouret, R. H., Evered, A., Wilbur, D. C., & Wolfe, J. M. (2011). Prevalence of abnormalities influences cytologists' error rates in screening for cervical cancer. *Archives of Pathology & Laboratory Medicine*, *135*(12), 1557–1560. https://doi.org/10.5858/arpa.2010-0739-OA

Feder, S., Püschel, A., Şimşek, M., Odenwald, S., Bendixen, A., & Einhäuser, W. (2024). Visual search for hazardous items: Using virtual reality (VR) in laypersons to test wearable displays for firefighters. *Frontiers in Virtual Reality*, *5*. https://doi.org/10.3389/frvir.2024.1252351

Feng, J., Kim, J., Luu, W., & Palmisano, S. (2019). Method for estimating display lag in the Oculus Rift S and CV1. *SIGGRAPH Asia 2019 Posters*, 1–2. https://doi.org/10.1145/3355056.3364590

Finseth, T., Barnett, N., Shirtcliff, E. A., Dorneich, M. C., & Keren, N. (2018). Stress inducing demands in virtual environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 2066–2070. https://doi.org/10.1177/1541931218621466

Finseth, T., Dorneich, M. C., Keren, N., Franke, W. D., & Vardeman, S. B. (2022). Manipulating stress responses during spaceflight training with virtual stressors. *Applied Sciences*, *12*(5), Article 5. https://doi.org/10.3390/app12052289

Frankenstein, J., Brüssow, S., Ruzzoli, F., & Hölscher, C. (2012). The language of landmarks: The role of background knowledge in indoor wayfinding. *Cognitive Processing*, *13*(S1), 165–170. https://doi.org/10.1007/s10339-012-0482-8

Gaschler, R., Schwager, S., Umbach, V. J., Frensch, P. A., & Schubert, T. (2014). Expectation mismatch: Differences between self-generated and cue-induced expectations. *Neurosci. Biobehav. Rev.*, *46 Pt 1*, 139–157.

Gath-Morad, M., Grübel, J., Steemers, K., Sailer, K., Ben-Alon, L., Hölscher, C., & Aguilar, L. (2024). The role of strategic visibility in shaping wayfinding behaviour in multilevel buildings. *Scientific Reports*, *14*, 3735. https://doi.org/10.1038/s41598-024-53420-6

Geißler, C. F., Friehs, M. A., Frings, C., & Domes, G. (2023). Time-dependent effects of acute stress on working memory performance: A systematic review and hypothesis. *Psychoneuroendocrinology*, *148*, 105998. https://doi.org/10.1016/j.psyneuen.2022.105998

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295x.103.4.650

Gnacek, M., Broulidakis, J., Mavridou, I., Fatoorechi, M., Seiss, E., Kostoulas, T., Balaguer-Ballester, E., Kiprijanovska, I., Rosten, C., & Nduka, C. (2022). emteqPRO—Fully integrated biometric sensing array for non-Invasive biomedical research in virtual reality. *Frontiers in Virtual Reality*, *3*. https://www.frontiersin.org/articles/10.3389/frvir.2022.781218

Goldfarb, E. V., Froböse, M. I., Cools, R., & Phelps, E. A. (2017). Stress and cognitive flexibility: Cortisol increases are associated with enhanced updating but impaired switching. *J. Cogn. Neurosci.*, *29*(1), 14–24.

Gorbunova, E. S. (2017). Perceptual similarity in visual search for multiple targets. *Acta Psychologica*, *173*, 46–54. https://doi.org/10.1016/j.actpsy.2016.11.010

Grant, D. A., & Berg, E. A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Grice, H. P. (1989). Logic and conversation. In *Studies in the Way of Words* (pp. 22–40). Harvard University Press. https://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf

Hadjipanayi, V., Ludwig, C. J. H., & Kent, C. (2023). Graded prioritisation of targets in search: Reward diminishes the low prevalence effect. *Cognitive Research: Principles and Implications*, *8*(1), 52. https://doi.org/10.1186/s41235-023-00507-9

Halaki, M., Ginn, K., Halaki, M., & Ginn, K. (2012). Normalization of EMG signals: To normalize or not to normalize and what to normalize to? In *Computational Intelligence in Electromyography Analysis—A Perspective on Current Applications and Future Challenges*. IntechOpen. https://doi.org/10.5772/49957

Hampshire Fire and Rescue Service. (2013). *Fatal fire investigation: Report of the Hampshire Fire and Rescue Service investigation into the deaths of Firefighters Alan Bannon and James Shears in Flat 72, Shirley Towers, Church Street, Southampton, SO15 5PE, on Tuesday 6 April 2010* (p. 184). Hampshire Fire and Rescue Service. https://highrisefire.co.uk/docs/shirleytowers.pdf

Hayes, S., Hirsch, C., & Mathews, A. (2008). Restriction of working memory capacity during worry. *Journal of Abnormal Psychology*, *117*(3), 712–717. https://doi.org/10.1037/a0012908

Hazlett, R. L. (2006). Measuring emotional valence during interactive experiences: Boys at video game play. *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Conference on Human Factors in Computing Systems. https://doi-org.abc.cardiff.ac.uk/10.1145/1124772.1124925

Head of Operational Procedures. (2012). *Search and rescue procedures within structures*. London Fire Brigade. https://assets.grenfelltowerinquiry.org.uk/LFB00084003_Exhibit%20DL_75%20-%20LFB%20Policy%20803%20entitled%20%27Search%20and%20Rescue%20Procedures%20within%20Structures%27%20issued%2027%20July%202012.pdf

Helminen, E. C., Morton, M. L., Wang, Q., & Felver, J. C. (2021). Stress reactivity to the Trier Social Stress Test in traditional and virtual environments: A meta-analytic comparison. *Psychosomatic Medicine*, *83*(3), 200. https://doi.org/10.1097/PSY.0000000000000918

Hettiarachchi, I. T., Hanoun, S., Nahavandi, D., & Nahavandi, S. (2019). Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. *PLoS ONE*, *14*(5), e0217288. https://doi.org/10.1371/journal.pone.0217288

Hölscher, C., Tenbrink, T., & Wiener, J. M. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition*, *121*(2), 228–247. https://doi.org/10.1016/j.cognition.2011.06.005

Holtgraves, T. (2014). Interpreting uncertainty terms. *Journal of Personality and Social Psychology*, *107*(2), 219–228. https://doi.org/10.1037/a0036930

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, *136*(2), 189–202. https://doi.org/10.1016/j.actpsy.2010.04.011

Hon, N., & Jabar, S. B. (2018). Learning in the target prevalence effect. *Perception*, *47*(7), 789–798. https://doi.org/10.1177/0301006618773081

Huang, D.-H., Chou, S.-W., Chen, Y.-L., & Chiou, W.-K. (2014). Frowning and jaw clenching muscle activity reflects the perception of effort during incremental workload cycling. *Journal of Sports Science & Medicine*, *13*(4), 921–928.

Hyun, J., Sliwinski, M. J., & Smyth, J. M. (2019). Waking up on the wrong side of the bed: The effects of stress anticipation on working memory in daily life. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *74*(1), 38–46. https://doi.org/10.1093/geronb/gby042

IglooVision.com. (2020). *Igloo vs CAVE*. https://www.igloovision.com/Documents/Igloo%20vs%20CAVE%20-%20June%202020.pdf

Ishibashi, K., & Kita, S. (2014). Probability cueing influences miss rate and decision criterion in visual searches. *I-Perception*, *5*(3), 170–175. https://doi.org/10.1068/i0649rep

Ishibashi, K., Kita, S., & Wolfe, J. M. (2012). The effects of local prevalence and explicit expectations on search termination times. *Attention, Perception, & Psychophysics*, *74*(1), 115–123. https://doi.org/10.3758/s13414-011-0225-4

Jang, G., Yoon, S., Lee, S.-E., Park, H., Kim, J., Ko, J. H., & Park, H.-J. (2013). Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *NeuroImage*, *81*, 61–72. https://doi.org/10.1016/j.neuroimage.2013.05.027

Janssens, L., & Schaeken, W. (2013). 'But' how do we reason with it: An experimental investigation of the implicature stemming from 'but'. *Journal of Pragmatics*, *57*, 194–209. https://doi.org/10.1016/j.pragma.2013.08.021

Jiang, J., Wagner, A. D., & Egner, T. (2018). Integrated externally and internally generated task predictions jointly guide cognitive control in prefrontal cortex. *Elife*, *7*, 1–23.

Jönsson, P., Wallergård, M., Osterberg, K., Hansen, A. M., Johansson, G., & Karlson, B. (2010). Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: A pilot study. *Psychoneuroendocrinology*, *35*(9), 1397–1403. https://doi.org/10.1016/j.psyneuen.2010.04.003

Kahneman, D. (2011). *Thinking, fast and slow* (1st ed.). Farra, Straus, & Giroux.

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 1151–1160. https://doi.org/10.1145/2638728.2641695

Kelly, O., Matheson, K., Martinez, A., Merali, Z., & Anisman, H. (2007). Psychosocial stress evoked by a virtual audience: Relation to neuroendocrine activity. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *10*(5), 655–662. https://doi.org/10.1089/cpb.2007.9973

Kemper, M., Gaschler, R., & Schubert, T. (2017). Stronger effects of self-generated than cue-induced expectations when verifying predictions in data graphs. *J. Cogn. Psychol.*, *29*(5), 626–641.

Kemper, M., Umbach, V., Schwager, S., Gaschler, R., Frensch, P., & Stürmer, B. (2012). What I say is what I get: Stronger effects of self-generated vs. cue-induced expectations in event-related potentials. *Frontiers in Psychology*, *3*, 1–14.

Keren, N., Bayouth, S. T., Franke, W. D., & Godby, K. M. (2013). Examining the effect of level of stress on firefighters' time-to-decision in virtual reality. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 299–303. https://doi.org/10.1177/1541931213571066

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1–2), 76–81. https://doi.org/10.1159/000119004

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and*

Decision Making, *4*(3), 186–209. https://doi.org/10.1518/155534310X12844000801203

Kramer, A. C., Neubauer, A. B., Scott, S. B., Schmiedek, F., Sliwinski, M. J., & Smyth, J. M. (2021). Stressor anticipation and subsequent affective well-being: A link potentially explained by perseverative cognitions. *Emotion*, *22*(8), 1787–1800. https://doi.org/10.1037/emo0000954

Kroll, S. L., Mayo, L. M., Asratian, A., Yngve, A., Perini, I., & Heilig, M. (2021). Negative self-evaluation induced by acute stress indexed using facial EMG. *Psychoneuroendocrinology*, *133*, 105402. https://doi.org/10.1016/j.psyneuen.2021.105402

Kurinec, C. A., Wise, A. V. T., Cavazos, C. A., Reyes, E. M., & Weaver, C. A. (2019). Clarity under cognitive constraint: Can a simple directive encourage busy speakers to avoid ambiguity? *Language and Cognition*, *11*(4), 621–644. https://doi.org/10.1017/langcog.2019.40

Kwok, K. K. K., Ng, A. K. T., & Lau, H. Y. K. (2018). Effect of navigation speed and VR devices on cybersickness. *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 91–92. https://doi.org/10.1109/ISMAR-Adjunct.2018.00041

Laessoe, U., Abrahamsen, S., Zepernick, S., Raunsbaek, A., & Stensen, C. (2023). Motion sickness and cybersickness – Sensory mismatch. *Physiology & Behavior*, *258*, 114015. https://doi.org/10.1016/j.physbeh.2022.114015

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776–785. https://doi.org/10.1111/1469-8986.00078

Lau, J. S.-H., & Huang, L. (2010). The prevalence effect is determined by past experience, not future prospects. *Vision Research*, *50*(15), 1469–1474. https://doi.org/10.1016/j.visres.2010.04.020

Lawrence, R. K., Cochrane, B. A., Eidels, A., Howard, Z., Lui, L., & Pratt, J. (2023). Emphasizing responder speed or accuracy modulates but does not abolish the distractor-induced quitting effect in visual search. *Cognitive Research: Principles and Implications*, *8*(1), 63. https://doi.org/10.1186/s41235-023-00516-8

Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual Review of Psychology*, *44*, 1–21. https://doi.org/10.1146/annurev.ps.44.020193.000245

Lazarus, R. S., & Folkman, S. (1987). Transactional theory and research on emotions and coping. *European Journal of Personality*, *1*(3), 141–169. https://doi.org/10.1002/per.2410010304

LeBlanc, V. R. (2009). The effects of acute stress on performance: Implications for health professions education. *Academic Medicine*, *84*(Supplement), S25–S33. https://doi.org/10.1097/ACM.0b013e3181b37b8f

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalised conversational implicature*. MIT press.

MacLeod, C. M. (1991). Half a century of research on the Stroop Effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Maechler, M., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Koller, M., & Conceicao, E. L. T. (2006). *robustbase: Basic Robust Statistics* (p. 0.99-4–1) [Dataset]. https://doi.org/10.32614/CRAN.package.robustbase

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Martirosov, S., Bureš, M., & Zítka, T. (2022). Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality. *Virtual Reality*, *26*(1), 15–32. https://doi.org/10.1007/s10055-021-00507-4

Mayo, L. M., & Heilig, M. (2019). In the face of stress: Interpreting individual differences in stress-induced facial expressions. *Neurobiology of Stress*, *10*, 100166. https://doi.org/10.1016/j.ynstr.2019.100166

McCarley, J. S. (2009). Effects of speed–accuracy instructions on oculomotor scanning and target recognition in a simulated baggage X-ray screening task. *Ergonomics*, *52*(3), 325–333. https://doi.org/10.1080/00140130802376059

Meng, F., & Zhang, W. (2014). Way-finding during a fire emergency: An experimental study in a virtual environment. *Ergonomics*, *57*(6), 816–827. https://doi.org/10.1080/00140139.2014.904006

Mineka, S., & Hendersen, R. W. (1985). Controllability and predictability in acquired motivation. *Annual Review of Psychology*, *36*, 495–529. https://doi.org/10.1146/annurev.ps.36.020185.002431

Mitroff, S. R., Biggs, A. T., Adamo, S. H., Dowd, E. W., Winkle, J., & Clark, K. (2015). What can 1 billion trials tell us about visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 1–5. https://doi.org/10.1037/xhp0000012

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moeschler, J. (2023). What makes inferences reliable? The unpredictable relationship between pragmatic inference and truth. *Journal of Pragmatics*, *218*, 153–164. https://doi.org/10.1016/j.pragma.2023.10.010

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024). *BayesFactor: Computation of Bayes Factors for Common Designs* (Version 0.9.12-4.7) [Computer software]. https://cran.r-project.org/web/packages/BayesFactor/index.html

Nakashima, R., Watanabe, C., Maeda, E., Yoshikawa, T., Matsuda, I., Miki, S., & Yokosawa, K. (2015). The effect of expert knowledge on medical search: Medical experts have specialized abilities for detecting serious lesions. *Psychological Research*, *79*(5), 729–738. https://doi.org/10.1007/s00426-014-0616-y

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences*, *107*(11), 5232–5237. https://doi.org/10.1073/pnas.0911972107

Nys, B. L., Wong, W., & Schaeken, W. (2024). Some scales require cognitive effort: A systematic review on the role of working memory in scalar implicature derivation. *Cognition*, *242*, 105623. https://doi.org/10.1016/j.cognition.2023.105623

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cogn. Psychol.*, *66*(2), 157–211.

Olk, B., Dinu, A., Zielinski, D. J., & Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *Royal Society Open Science*, *5*(5), 172331. https://doi.org/10.1098/rsos.172331

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Ottaviani, C., Shapiro, D., & Couyoumdjian, A. (2013). Flexibility as the key for somatic health: From mind wandering to perseverative cognition. *Biological Psychology*, *94*(1), 38–43. https://doi.org/10.1016/j.biopsycho.2013.05.003

Padgitt, A. J., & Hund, A. M. (2012). How good are these directions? Determining direction quality and wayfinding efficiency. *Journal of Environmental Psychology*, *32*(2), 164–172. https://doi.org/10.1016/j.jenvp.2012.01.007

Palmisano, S., Allison, R. S., & Kim, J. (2020). Cybersickness in head-mounted displays is caused by differences in the user's virtual and physical head pose. *Frontiers in Virtual Reality*, *1*. https://www.frontiersin.org/articles/10.3389/frvir.2020.587698

Palmisano, S., Allison, R. S., Teixeira, J., & Kim, J. (2022). Differences in virtual and physical head orientation predict sickness during active head-mounted display-based virtual reality. *Virtual Reality*. https://doi.org/10.1007/s10055-022-00732-5

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Passchier, J., & Helm-Hylkema, H. (1981). The effect of stress imagery on arousal and its implications for biofeedback of the frontalis muscles. *Biofeedback and Self-Regulation*, *6*(3), 295–303. https://doi.org/10.1007/BF01000655

Peltier, C., & Becker, M. W. (2016). Decision processes in visual search as a function of target prevalence. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1466–1476. https://doi.org/10.1037/xhp0000248

Plessow, F., Fischer, R., Kirschbaum, C., & Goschke, T. (2011). Inflexibly focused under stress: Acute psychosocial stress increases shielding of action goals at the expense of reduced cognitive flexibility with increasing time lag to the stressor. *J. Cogn. Neurosci.*, *23*(11), 3218–3227.

Plessow, F., Kiesel, A., & Kirschbaum, C. (2012). The stressed prefrontal cortex and goal-directed behaviour: Acute psychosocial stress impairs the flexible implementation of task goals. *Exp. Brain Res.*, *216*(3), 397–408.

Posada-Quintero, H. F., Reljin, N., Mills, C., Mills, I., Florian, J. P., VanHeest, J. L., & Chon, K. H. (2018). Time-varying analysis of electrodermal activity during exercise. *PLOS ONE*, *13*(6), e0198328. https://doi.org/10.1371/journal.pone.0198328

Pupil Labs. (n.d.). *Core—Home—Pupil Labs Docs*. https://docs.pupil-labs.com/core/

Qin, S., Hermans, E. J., van Marle, H. J. F., Luo, J., & Fernández, G. (2009). Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biological Psychiatry*, *66*(1), 25–32. https://doi.org/10.1016/j.biopsych.2009.03.006

R Core Team. (2024). *R: A Language and Environment for Statistical Computing* (Version 4.4.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Radun, J., Maula, H., Rajala, V., Scheinin, M., & Hongisto, V. (2022). Acute stress effects of impulsive noise during mental work. *Journal of Environmental Psychology*, *81*, 101819. https://doi.org/10.1016/j.jenvp.2022.101819

Reuderink, B., Nijholt, A., & Poel, M. (2009). Affective Pacman: A frustrating game for brain-computer interface experiments. In A. Nijholt, D. Reidsma, & H. Hondorp (Eds.), *Intelligent Technologies for Interactive Entertainment* (Vol. 9, pp. 221–227). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02315-6_23

Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision*, *8*(15), 15. https://doi.org/10.1167/8.15.15

Rieger, T., Heilmann, L., & Manzey, D. (2021). Visual search behavior and performance in luggage screening: Effects of time pressure, automation aid, and target expectancy. *Cognitive Research: Principles and Implications*, *6*(1), 12. https://doi.org/10.1186/s41235-021-00280-7

Rieger, T., & Manzey, D. (2024). Understanding the impact of time pressure and automation support in a visual search task. *Human Factors*, *66*(3), 770–786. https://doi.org/10.1177/00187208221111236

Robinson, S. J., Leach, J., Owen-Lynch, P. J., & Sünram-Lea, S. I. (2013). Stress reactivity and cognitive performance in a simulated firefighting emergency. *Aviation, Space, and Environmental Medicine*, *84*(6), 592–599. https://doi.org/10.3357/ASEM.3391.2013

Rodrigues, J., Studer, E., Streuber, S., & Sandi, C. (2021). IMVEST, an immersive multimodal virtual environment stress test for humans that adjusts challenge to individual's performance. *Neurobiology of Stress*, *15*, 100382. https://doi.org/10.1016/j.ynstr.2021.100382

Rubin, M. (2021). When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese*, *199*(3–4), 10969–11000. https://doi.org/10.1007/s11229-021-03276-4

Ryzhova, M., & Demberg, V. (2023). Processing cost effects of atypicality inferences in a dual-task setup. *Journal of Pragmatics*, *211*, 47–80. https://doi.org/10.1016/j.pragma.2023.04.005

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience*, *14*. https://www.frontiersin.org/articles/10.3389/fnhum.2020.00096

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology*, *35*(7), 977–986. https://doi.org/10.1016/j.psyneuen.2009.12.010

Schwager, S., Gaschler, R., Rünger, D., & Frensch, P. A. (2017). Tied to expectations: Predicting features speeds processing even under adverse circumstances. *Mem. Cognit.*, *45*(4), 611–624.

Schwark, J. D., MacDonald, J., Sandry, S., & Dolgov, I. (2013). Prevalence-based decisions undermine visual search. *Visual Cognition*, *21*(5), 541–568. https://doi.org/10.1080/13506285.2013.811135

Selly, L. (2022). *V1.3 Installation Instructions—IglooUnityToolkit—Confluence*. https://360igloo.atlassian.net/wiki/spaces/IG/pages/3272867841/v1.3+Installation+Instructions

Shea, C. H., Wright ,David L., Wulf ,Gabriele, & and Whitacre, C. (2000). Physical and observational practice afford unique learning opportunities. *Journal of Motor Behavior*, *32*(1), 27–36. https://doi.org/10.1080/00222890009601357

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annu. Rev. Neurosci.*, *40*, 99–124.

Shi, Y., Kang, J., Xia, P., Tyagi, O., Mehta, R. K., & Du, J. (2021). Spatial knowledge and firefighters' wayfinding performance: A virtual reality search and rescue experiment. *Safety Science*, *139*, 105231. https://doi.org/10.1016/j.ssci.2021.105231

Shiban, Y., Diemer, J., Brandl, S., Zack, R., Mühlberger, A., & Wüst, S. (2016). Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains. *International Journal of Psychophysiology*, *110*, 47–55. https://doi.org/10.1016/j.ijpsycho.2016.10.008

Shields, G. S., Bonner, J. C., & Moons, W. G. (2015). Does cortisol influence core executive functions? A meta-analysis of acute cortisol administration effects on working memory, inhibition, and set-shifting. *Psychoneuroendocrinology*, *58*, 91–103. https://doi.org/10.1016/j.psyneuen.2015.04.017

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neurosci. Biobehav. Rev.*, *68*, 651–668.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Sidarus, N., Palminteri, S., & Chambon, V. (2019). Cost-benefit trade-offs in decision-making and learning. *PLoS Computational Biology*, *15*(9), e1007326. https://doi.org/10.1371/journal.pcbi.1007326

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. https://doi.org/10.1037/h0020586

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Blackwell.

Sperber, D., & Wilson, D. (1987). Précis of relevance: Communication and cognition. *Behavioral and Brain Sciences*, *10*(04), 697. https://doi.org/10.1017/S0140525X00055345

Srivastava, P., Rimzhim, A., Vijay, P., Singh, S., & Chandra, S. (2019). Desktop VR Is better than non-ambulatory HMD VR for spatial learning. *Frontiers in Robotics and AI*, *6*. https://doi.org/10.3389/frobt.2019.00050

Staal, M. A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework* (No. NASA/TM-2004-212824). National Aeronautical Space Agency. https://ntrs.nasa.gov/citations/20060017835

Starcke, K., Wolf, O. T., Markowitsch, H. J., & Brand, M. (2008). Anticipatory stress influences decision making under explicit risk conditions. *Behavioral Neuroscience*, *122*(6), 1352–1360. https://doi.org/10.1037/a0013281

Stauffert, J.-P., Niebling, F., & Latoschik, M. E. (2018). Effects of latency jitter on simulator sickness in a search task. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 121–127. https://doi.org/10.1109/VR.2018.8446195

SteamVR Unity Plugin. (2023). [C#]. Valve Software. https://github.com/ValveSoftware/steamvr_unity_plugin (Original work published 2017)

Steinhauser, M., Maier, M., & Hübner, R. (2007). Cognitive control under stress: How stress affects strategies of task-set reconfiguration. *Psychological Science*, *18*(6), 540–545. https://doi.org/10.1111/j.1467-9280.2007.01935.x

Stothart, C., Clement, A., & Brockmole, J. R. (2018). Satisfaction in motion: Subsequent search misses are more likely in moving search displays. *Psychonomic Bulletin & Review*, *25*(1), 409–415. https://doi.org/10.3758/s13423-017-1300-1

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sumwalt, R. L. (1993). *The sterile cockpit* (No. 4; ASRS Directline). Aviation Safety Reporting System (ASRS). https://asrs.arc.nasa.gov/publications/directline/dl4_sterile.htm

Taylor, J. E. T., Hilchey, M. D., Weidler, B. J., & Pratt, J. (2022). Eliminating the low-prevalence effect in visual search with a remarkably simple strategy. *Psychological Science*, *33*(5), 716–724. https://doi.org/10.1177/09567976211048485

Treisman, A. M. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 652–676. https://doi.org/10.1037/0096-1523.17.3.652

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tuddenham, W. J. (1962). Visual search, image organization, and reader error in roentgen diagnosis. Studies of the psycho-physiology of roentgen image perception. *Radiology*, *78*(5), 694–704. https://doi.org/10.1148/78.5.694

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. https://doi.org/10.1016/0749-596x(89)90040-5

Umbach, V., Schwager, S., Frensch, P., & Gaschler, R. (2012). Does explicit expectation really affect preparation? *Frontiers in Psychology*, *3*, 1–12.

Van Boxtel, A., & Jessurun, M. (1993). Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task. *Psychophysiology*, *30*(6), 589–604. https://doi.org/10.1111/j.1469-8986.1993.tb02085.x

Van Den Oever, F., Gorobets, V., Saetrevik, B., Fjeld, M., & Kunz, A. (2022). Comparing visual search between physical environments and VR. *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 411–416. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00089

van der Vijgh, B., Beun, R. J., van Rood, M., & Werkhoven, P. (2014). GASICA: Generic automated stress induction and control application design of an application for controlling the stress state. *Frontiers in Neuroscience*, *8*, 400. https://doi.org/10.3389/fnins.2014.00400

van Tiel, B., Marty, P., Pankratz, E., & Sun, C. (2019). Scalar inferences and cognitive load. *Proceedings of Sinn Und Bedeutung*, *23*(2), Article 2. https://doi.org/10.18148/sub/2019.v23i2.622

van Tiel, B., Pankratz, E., & Sun, C. (2019). Scales and scalarity: Processing scalar inferences. *Journal of Memory and Language*, *105*, 93–107. https://doi.org/10.1016/j.jml.2018.12.002

Williams-Bell, F. M., Kapralos, B., Hogue, A., Murphy, B. M., & Weckman, E. J. (2015). Using serious games and virtual simulation for training in the fire service: A review. *Fire Technology*, *51*(3), 553–584. https://doi.org/10.1007/s10694-014-0398-1

Wolf, K., Mass, R., Ingenbleek, T., Kiefer, F., Naber, D., & Wiedemann, K. (2005). The facial pattern of disgust, appetence, excited joy and relaxed joy: An improved facial EMG study. *Scandinavian Journal of Psychology*, *46*(5), 403–409. https://doi.org/10.1111/j.1467-9450.2005.00471.x

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychon. Bull. Rev.*, *28*(4), 1060–1092.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision*, *13*(3), 33. https://doi.org/10.1167/13.3.33

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 419–433. https://doi.org/10.1037/0096-1523.15.3.419

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*(7041), 439–440. https://doi.org/10.1038/435439a

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology. General*, *136*(4), 623–638. https://doi.org/10.1037/0096-3445.136.4.623

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121–124. https://doi.org/10.1016/j.cub.2009.11.066

Yu, R. (2016). Stress potentiates decision biases: A stress induced deliberation-to-intuition (SIDI) model. *Neurobiology of Stress*, *3*, 83–95. https://doi.org/10.1016/j.ynstr.2015.12.006

Zhang, H., & Houpt, J. W. (2020). Exaggerated prevalence effect with the explicit prevalence information: The description-experience gap in visual search. *Attention, Perception, & Psychophysics*, *82*(7), 3340–3356. https://doi.org/10.3758/s13414-020-02045-8

Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., & Domes, G. (2019). Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*, *101*, 186–192. https://doi.org/10.1016/j.psyneuen.2018.11.010

Zoccola, P. M., Dickerson, S. S., & Zaldivar, F. P. (2008). Rumination and cortisol responses to laboratory stressors. *Psychosomatic Medicine*, *70*(6), 661–667. https://doi.org/10.1097/PSY.0b013e31817bbc77

Appendices

Appendix A: Verbatim Instructions in SAR Game (Experiments 1 and 2)

Key-press instructions are right-aligned for ease of reading.

Changes or additions to the instructions for Experiment 2 are presented in a boxed section.

<u>Familiarisation</u>

Walking test

Welcome to the Game Tutorial!

Let's start by learning how to control your character's movement. Once you're confident in controlling your character, you'll undergo a brief walking test before we proceed further.

Press [Enter] to proceed

--

How to Walk

1. Move your mouse to look around.
2. Press [W] to move forward.
3. For the most part, movement is controlled by holding down the [W] button whilst moving your mouse. This allows you to move in the direction you are looking.
4. Press [A], [S], [D] to move leftward, backward, and rightward when necessary.

These instructions will appear below during the tutorial.

--

Before you start the walking test, take your time to practice walking in the park with these controls until you feel comfortable. A good way to practice is to walk towards your preferred landmarks in the scene.

Press [Enter] to proceed

--

(after successfully completing the walking test)

You have passed the walking test and are now ready to proceed to the next phase of the tutorial. You will be asked to complete some tasks to help you learn the layout of an office building.

Exploring the office building: Item retrieval tasks

You will be given a series of tasks to complete. The task instructions will be updated here. You will also be prompted to press the [R] key to interact with task-relevant elements. Take your time to learn the layout of the building as you complete all tasks.

--

Get the broken laptop from the Quality Assurance Specialist’s office in the Software Hub.

Press [R] to get the broken laptop

Leave the broken laptop on the red tray in the IT Department in the Service Point.

Press [R] to leave the broken laptop here

Get the folders on the Assistant Finance Manager's desk.

Press [R] to get the folders

Pass the folders to the Assistant Finance Manager in the Copy Room.

Press [R] to pass the folders

Get the dirty plates from the Conference Room via the back corridor.

Press [R] to get the dirty plates

Place the dirty plates in the kitchen sink via the Common Room.

Press [R] to leave the dirty plates here

Get the pizza box from the Assistant Marketing Manager's desk.

Press [R] to get the pizza box

Get the remote control from the Assistant Marketing Manager's desk.

Press [R] to get the remote control

Place the remote control on the desk where the dirty plates were in the Conference Room via the back corridor.

Press [R] to leave the remote control here

Place the pizza box in the red bin in the Common Room.

Press [R] to leave the pizza box here

Get the new laptop from the IT Department.

Press [R] to get the new laptop

Place the new laptop on the Quality Assurance Specialist's desk.

Press [R] to leave the new laptop here

Get the boxes from the back corridor.

Press [R] to get the boxes

Place one box in the Finance Department Copy Room.

Press [R] to leave a box here

Place the other box in the Marketing Department Copy Room.

Press [R] to leave the other box here

Get the garbage from the Women's toilet.

Press [R] to get the garbage

Get the garbage from the Men's toilet.

Press [R] to get the garbage

Remove the garbage through the fire exit.

Press [R] to exit the building

Floorplan construction (knowledge of building)

You will be asked to reconstruct from memory the layout of the building you have just navigated.

You will be asked to label the different areas of the building on a blank floorplan.

Please try to label the areas as accurately as you can. However, if you don’t know what an area is called, please select the ‘Not Sure’ option instead of guessing.

Press [Enter] to proceed

--

Label the different areas of the map by dragging the options below and placing them in the orange boxes on the map.

Press [B] to start

Click ‘Submit’ once you are ready to submit your answers.

Search and Rescue Cover Story

There has been a catastrophic electrical fault within the primary power supply system of the building, posing a dire threat to the occupants' safety.

Press [Enter] to proceed

According to reports, the Assistant Finance Manager is unaccounted for after evacuation.

Press [Enter] to proceed

As a firefighter, you must look for the person and get them to safety. There is also a portable stove that needs to be removed form the kitchen to prevent an explosion.

Press [Enter] to proceed

You must complete both tasks in the shortest time possible.

Press [Enter] to proceed

--

If you get lost in the building, you can press [Backspace] to return to the starting point.

Press [Enter] to proceed

This [Backspace] prompt will remain visible at the bottom left corner of the screen.

Press [Enter] to proceed

The four LED lights below indicate how much air is left in your Self-Contained Breathing Apparatus (SCBA).

Press [Enter] to proceed

The more LED lights are turned off, the less air you have left.

Press [Enter] to proceed

If the screen starts fogging, you should press [Spacebar] when prompted to regulate your breathing and reduce fogging.

Press [Enter] to proceed

If only the red LED light is left blinking, you have very little air left.

Press [Enter] to proceed

If you have used up all your air supply before you complete the tasks, you perish.

Press [Enter] to proceed

--

To recap, a catastrophic electrical fault with the primary power supply system has plunged the building into darkness.

The Assistant Finance Manager is in the building and could be in danger.

You need to bring them to safety using the fire exit and remove the portable stove from the kitchen as quickly as possible to prevent an explosion.

Press [Enter] to proceed

At the end of this trial, you will receive feedback about your performance:

- Whether you have removed the portable stove
- Whether you have rescued the Assistant Finance Manager

You have only one chance to complete both tasks.

--

If you need to review the full instructions, please press [Backspace]

If you are ready to start, please press [Enter]

(to confirm choice if pressed [Enter])

Press [Enter] again to confirm that you are ready to proceed. Otherwise, press [Backspace] to review the full instructions.

(to confirm choice if pressed [Backspace])

Press [Backspace] again to confirm that want to review the full instructions. Otherwise, press [Enter] to proceed.

--

<u>Search and Rescue mission</u>

(Right before the mission commenced)

New floorplan

You will be shown a floorplan that might be useful for the search and rescue mission.

Experiment 2 *(IF: Inaccurate-outdated condition)* We have found a floorplan of the same building that could be outdated, which might or might not be useful for the search and rescue mission. *(IF: Inaccurate-same condition)* We have found a floorplan of the same building, which might or might not be useful for the search and rescue mission.

You have a maximum of 2 minutes to study the floorplan.

When the time is up, the floorplan will disappear automatically, and you will enter the burning building.

If you don't need the full 2 minutes to study the floorplan, you can opt to stop studying the floorplan earlier and enter the building.

Press [B] to start

(Floorplan presentation)

Click 'Proceed' to stop studying the floorplan

(After exiting the building through the Fire Exit.)

Feedback

(IF: Retrieved stove, did not rescue person)

You failed to locate the person in the building! You must look for them and bring them to safety.

(IF: Did not retrieve stove, rescued person)

You managed to locate the person. However, you failed to remove the stove. You must remove it to prevent an explosion.

(IF: Retrieved stove, rescued person)

Congratulations! You managed to locate the person and also removed the stove.

--

How were you coping with the search and rescue task just now?

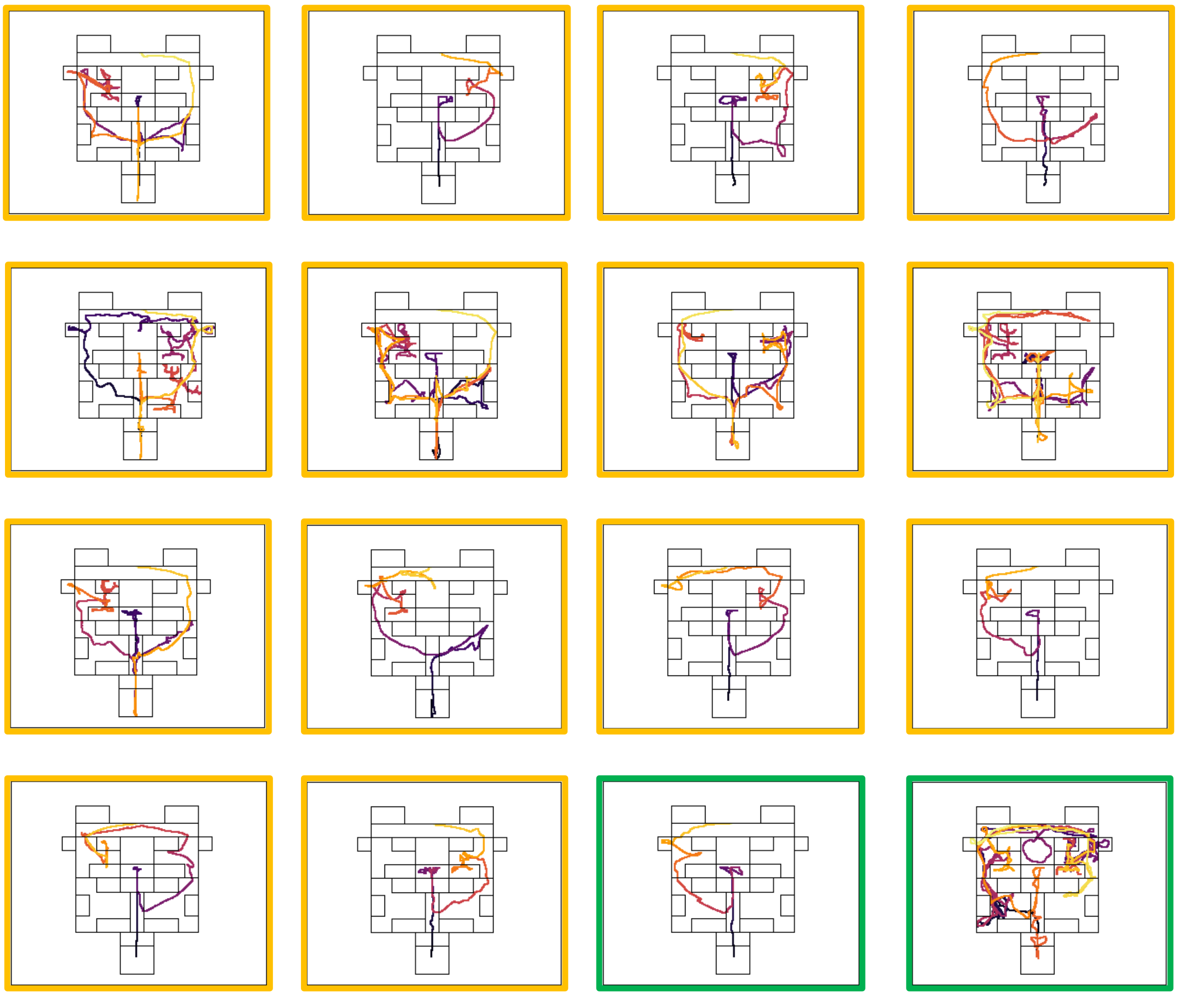
◯ ◯ ◯ ◯ ◯ ◯ ◯ ◯ ◯ ◯

1	10
I felt no pressure	I was unable to cope with the pressure

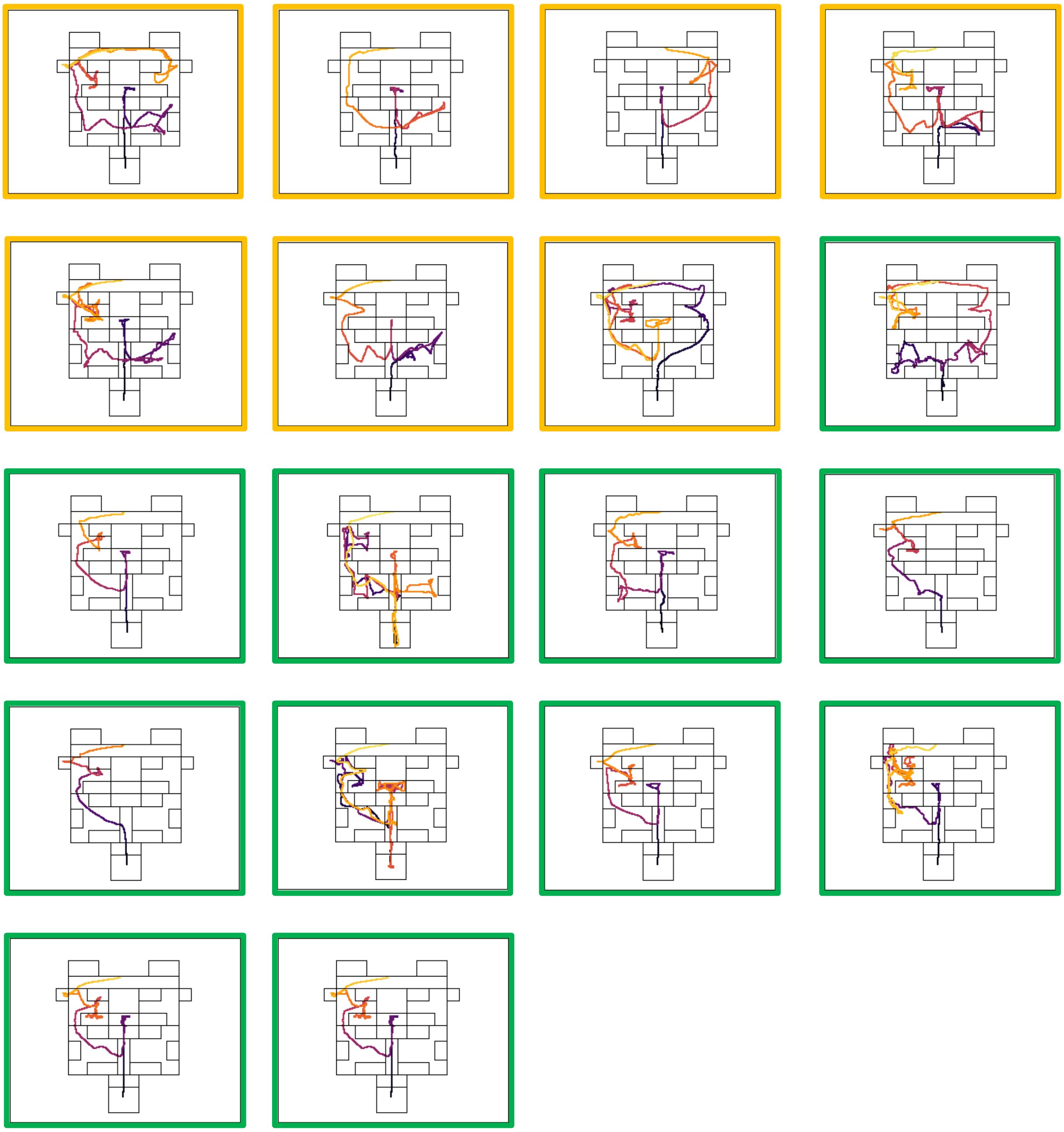
Appendix B: Participants' Search Patterns in Experiment 2

These search patterns are organised by the Inaccurate-same and Inaccurate-outdated conditions. Floorplan B was flipped horizontally so that the Finance Department (where the trapped Assistant Finance Manager was) is on the *left* side of the building in the diagrams below. Participants who followed the floorplan (yellow) are differentiated from those who did not (green) by the colour of the border.

Inaccurate-same group



Inaccurate-outdated group



Appendix C: Verbatim Instructions in SAR Game (Experiment 3)

Key-press instructions are right-aligned for ease of reading.

Familiarisation

Walking Test *is identical to that in Experiments 1 and 2. Please see Appendix A.*

Search and Rescue Cover Story

Lemonade Tech Ltd has very kindly offered their office space and two of their employees for our firefighting exercises today.

Press [Enter] to proceed

Before we start these exercises, you will be given a series of tasks to complete to help you familiarise yourself with the building.

Press [Enter] to proceed

Exploring the office building: Item retrieval tasks

The task instructions will be updated here. You will also be prompted to press the [R] key to interact with task-relevant elements.

Press [Enter] to proceed

Take your time to learn the layout of the building as you complete all tasks.

Press [Enter] to proceed

Get the broken laptop from the Quality Assurance Specialist's office in the Software Hub.

Press [R] to get the broken laptop.

Leave the broken laptop on the red tray in the IT Department in the Service Point.

Press [R] to leave the broken laptop here.

Get the folders from the desk outside the Finance Manager’s office.

Press [R] to get the folders.

Pass the folders to the Finance Manager in the Finance Department Copy Room.

Press [R] to pass the folders to the Finance Manager.

Get the diary planner on the Conference Room table via the back corridor.

Press [R] to get the diary planner.

Pass the diary planner to the Marketing Manager in the Marketing Department Copy Room via the back corridor.

Press [R] to pass the diary planner to the Marketing Manager.

Get the pizza box from the desk outside the Marketing Manager's office.

Press [R] to get the pizza box.

Get the remote control from the desk outside the Marketing Manager's office.

Press [R] to get the remote control.

Place the remote control on the desk next to the dirty plates in the Conference Room via the back corridor.

Press [R] to leave the remote control here.

Get the dirty plates on the Conference Room table.

Press [R] to get the dirty plates.

Place the dirty plates in the kitchen sink via the Common Room.

Press [R] to leave the dirty plates here.

Place the pizza box in the red bin in the Common Room.

Press [R] to leave the pizza box here.

Get the new laptop from the IT Department.

Press [R] to get the new laptop.

Place the new laptop on the Quality Assurance Specialist's desk.

Press [R] to leave the new laptop here.

Get the boxes from the back corridor via the Marketing Department.

Press [R] to get the boxes.

Place one box in the Finance Department Copy Room via the back corridor.

Press [R] to leave a box here.

Place the other box in the Marketing Department Copy Room via the back corridor.

Press [R] to leave the other box here.

Get the garbage from the Women's toilet.

Press [R] to get the garbage.

Get the garbage from the Men's toilet.

Press [R] to get the garbage.

Remove the garbage through the fire exit.

Press [R] to exit the building

The fire exit is jammed and cannot be opened. We are awaiting parts and repair, which will take several weeks. Please use the main entrance until further notice. Apologies for the inconvenience and thank you for your patience.

Press [R] to stop reading and continue.

Remove the garbage through the front entrance.

Press [R] to exit the building

Floorplan construction (knowledge of building)

You will be asked to reconstruct from memory the layout of the building you have just navigated.

You will be asked to label the different areas of the building on a blank floorplan.

Please try to label the areas as accurately as you can. However, if you don’t know what an area is called, please select the ‘Not Sure’

Press [Enter] to proceed

--

Label the different areas of the map by dragging the options below and placing them in the orange boxes on the map.

Press [B] to start

Click ‘Submit’ once you are ready to submit your answers.

Search and Rescue Training

Now that you have familiarised yourself with the building, we are almost ready for our training exercises. We have used fake smoke to simulate a burning building.

Press [Enter] to proceed

The Finance Manager and Marketing Manager of Lemonade Tech, whom you have met earlier, will be roleplaying as victims that are trapped in a burning building.

Press [Enter] to proceed

There is also a portable stove that needs to be removed from the Kitchen to prevent an explosion.

Press [Enter] to proceed

You will be trained to rescue both of them and also retrieve the portable stove as quickly as possible.

Press [Enter] to proceed

Before we start the training exercises, let's go through some important information.

Press [Enter] to proceed

--

The four LED lights below indicate how much air is left in your Self-Contained Breathing Apparatus (SCBA).

Press [Enter] to proceed

The more LED lights are turned off, the less air you have.

Press [Enter] to proceed

If the screen starts fogging, you should press [Spacebar] when prompted to regulate your breathing and reduce fogging.

Press [Enter] to proceed

If only the red LED light is left blinking, you have very little air left.

Press [Enter] to proceed

In a real fire, you would perish if you used up all your air supply before completing the task.

Press [Enter] to proceed

To recap, you are training for rescue speed and efficiency for today's firefighting exercises. The Finance Manager and Marketing Manager are in the building and could be in danger. You need to bring both of them to safety and remove the portable stove from the kitchen as quickly as you can to prevent an explosion.

Press [Enter] to proceed

At the end of each exercise, you will receive feedback about your performance:

- Whether you have rescued the Finance Manager
- Whether you have rescued the Marketing Manager
- Whether you have retrieved the portable stove
- Whether you took the shortest route and were quick enough to complete the tasks above.

There are 6 exercises in total. Try to perform to the best of your ability in each exercise.

--

If you need to review the full instructions, please press [Backspace]

If you are ready to start, please press [Enter]

(to confirm choice if pressed [Enter])

Press [Enter] again to confirm that you are ready to proceed. Otherwise, press [Backspace] to review the full instructions.

(to confirm choice if pressed [Backspace])

Press [Backspace] again to confirm that want to review the full instructions. Otherwise, press [Enter] to proceed.

--

How were you coping with the search and rescue task just now?

○ ○ ○ ○ ○ ○ ○ ○ ○ ○

1 10

I felt no pressure	I was unable to cope with the pressure

(The question above and feedback appeared after each SAR training trial. Please see Appendix D for feedback variations.)

Search and Rescue Mission

That was a nice training session. You've done a good job overall today.

.

..

…

.

..

…

As a firefighter, you are called to a blazing office building in the city.

Upon arrival, you realise that the building in question is Lemonade Tech Ltd., where you had your firefighting training two days ago.

It turns out that there has been a catastrophic electrical fault within the primary power supply system of the building, posing a dire threat to the occupants' safety.

Press [Enter] to proceed

According to reports, the Finance Manager and Marketing Manager are unaccounted for after evacuation. You must look for them and get them to safety.

The portable stove that is still in the kitchen also needs to be removed to prevent an explosion.

Press [Enter] to proceed

--

Remember, this is not a training exercise but a real incident, so you must complete the following tasks in the shortest time possible:

- Rescue the Finance Manager
- Rescue the Marketing Manager, and
- Retrieve the portable stove.

When you are ready to enter the building, press [Enter]

.

..

…

.

..

…

.

..

…

We've just received a report. It is from the Receptionist of Lemonade Tech who was evacuated from the building some moments ago.

(Participants were either informed about the Marketing Manager/Dept. or the Finance Manager/Dept, and did not have to rescue them. The following will refer to the Marketing Manager/Dept.)

(IF: Explicit-message group)

The Marketing Manager is attending their teambuilding event off-site and not here.

(IF: Implicit-message group)

The Marketing Department is holding their teambuilding event in a nearby hotel.

(IF: Control group – Participants received either the Explicit or Implicit message.)

(Explicit) The IT Manager is attending their teambuilding event off-site and not here.

(Implicit) The IT Department is holding their teambuilding event in a nearby hotel.

Press [Enter] to proceed

--

(IF: Rescued the Marketing Manager/Finance Manager and retrieved stove)

Congratulations! In this search and rescue mission you…

- Rescued the Marketing Manager
- Retrieved the portable stove

(IF: Rescued the Marketing Manager/Finance Manager but did not retrieve stove)

Oh no! In this search and rescue mission you…

- Rescued the Marketing Manager
- Failed to retrieve the portable stove

(IF: Did not rescue the Marketing Manager/Finance Manager but retrieved stove)

Oh no! In this search and rescue mission you…

- Failed to rescue the Marketing Manager
- Retrieved the portable stove

(IF: Did not rescue the Marketing/Finance Manager nor retrieved stove)

Oh no! In this search and rescue mission you…

- Failed to rescue the Marketing Manager
- Failed to retrieve the portable stove

Press [Enter] to proceed

--

How were you coping with the search and rescue task just now?

○ ○ ○ ○ ○ ○ ○ ○ ○ ○

1 10

I felt no pressure	I was unable to cope with the pressure

Appendix D: Feedback for Search and Rescue Training Trials (Experiment 3)

Determination of the 'shortest time' to complete a trial

- The time taken to compete each trial was only recorded for successful trials: Rescued both the Finance and Marketing Managers and retrieved the portable stove using the shortest route (Figure 4 in main text).
- The completion time for the first successful trial was recorded as a 'best time'. The time for each subsequent instance of successful trials was compared against the existing 'best time'.
- If the duration was within 3s of the existing 'best time', participants would be considered to have completed the trial quickly enough, with the shorter duration of the two replacing the 'best time'.
- If the duration was more than 3s longer than the 'best time', participants would be considered to not have completed the trial quickly enough and the 'best time' would remain unchanged.

Feedback :

The feedback participants received at the end of each of the six SAR training trails was divided into two sections (that were presented together).

<u>The first section depended on whether they:</u>

- Rescued the Assistant Marketing Manager
- Rescued the Assistant Finance Manager
- Retrieved the portable stove

For any task that participants failed to complete, a red cross will appear next to the feedback, e.g., " Failed to retrieve the portable stove".

For any task that participants completed successfully, a green cross will appear next to the feedback, e.g., " Rescued the Assistant Marketing Manager."

The second (lower) section depended on whether they:

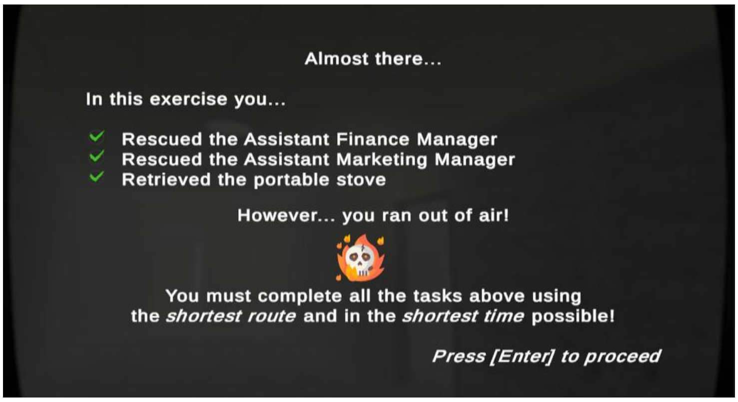
- Exceeded the time limit (4.5 minutes) for each trial
- Searched the relevant areas
- Took the shortest route
- Took the shortest time to complete each trial

The feedback will appear below the first section.

Time limit

If participants exceeded the time limit (450s), regardless of whether they met the other goals in the second section, the feedback would depend on whether they have completed all three tasks in the first section.

If all three tasks were completed, a failure sound effect would accompany:



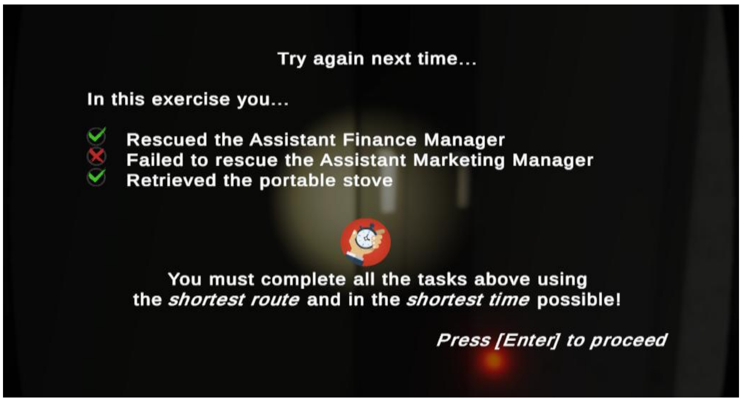
If they failed to complete all three tasks, a failure sound effect would accompany:

Searched relevant areas

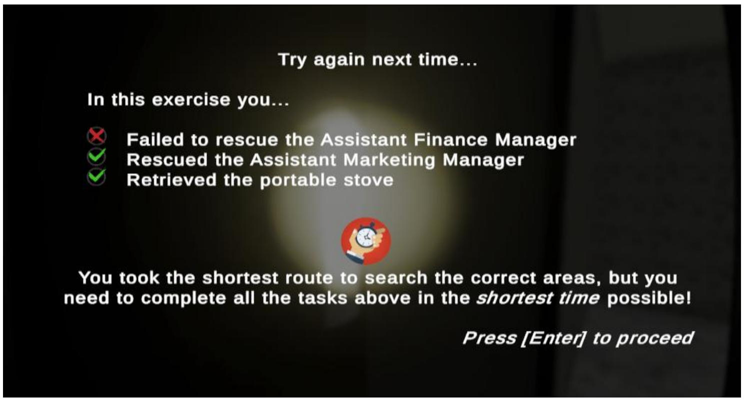
These areas, coded A0, B0, C0, D0, Z1, Z2, K0, and K1, are areas participants have to pass through in order to complete all three tasks.

If any of these area codes are missing from the participant's path data, it would mean that they have not searched all relevant areas.

However, to avoid participants searching every area in the building in subsequent trials, participants were reminded to use the shortest route, and a clock-ticking sound effect would accompany:

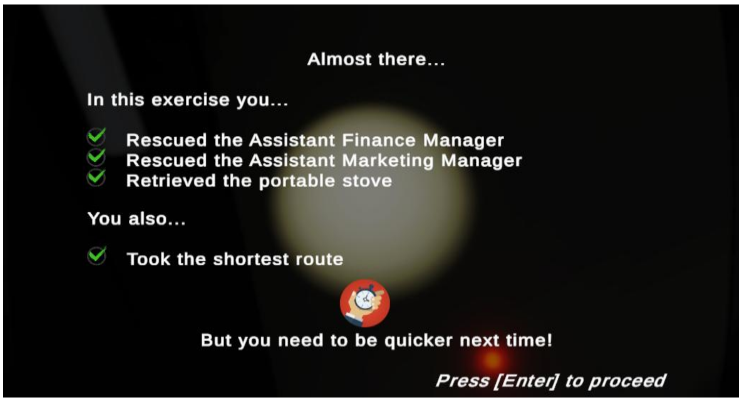


If participants searched all relevant areas but failed to complete the three tasks, a clock-ticking sound effect would accompany:

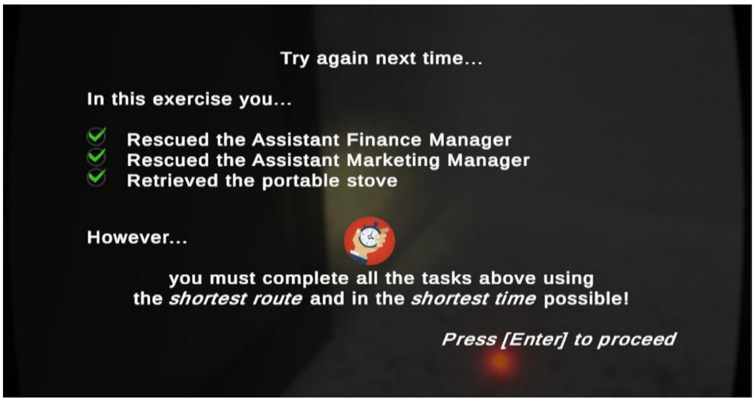


Shortest route

This would only be computed if participants have completed the three tasks. If they completed all three tasks using the shortest route, but not in the shortest time, a clock-ticking sound effect would accompany:



If participants completed all three tasks but not using the shortest route, a clock-ticking sound effect would accompany:



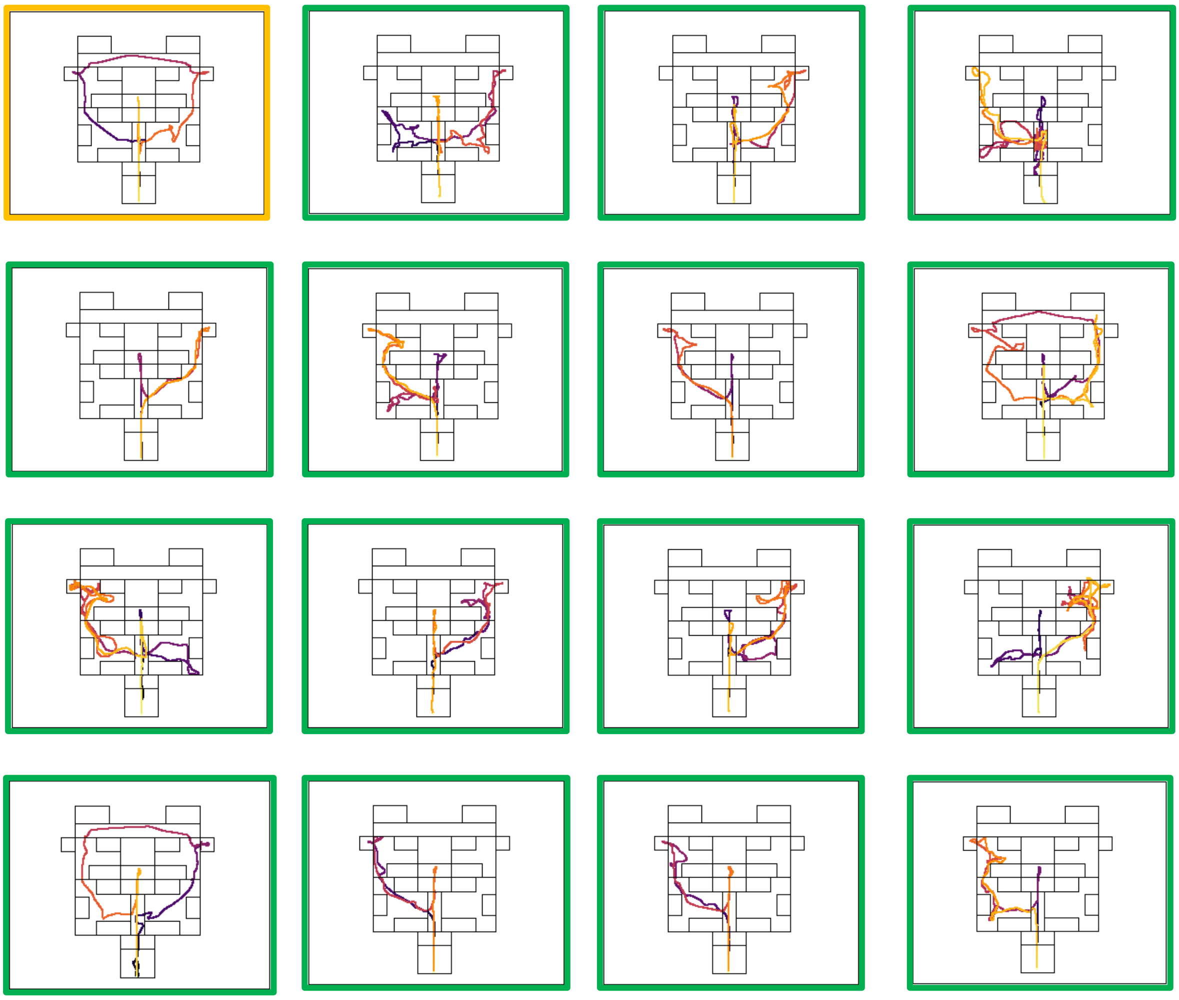
Shortest time

If it was their first time having completed all three tasks and they also took the shortest route, or if they completed the three tasks using the shortest route and in the shortest time, a congratulatory sound effect would accompany:

Appendix E: Participants' Search Patterns in Experiment 3

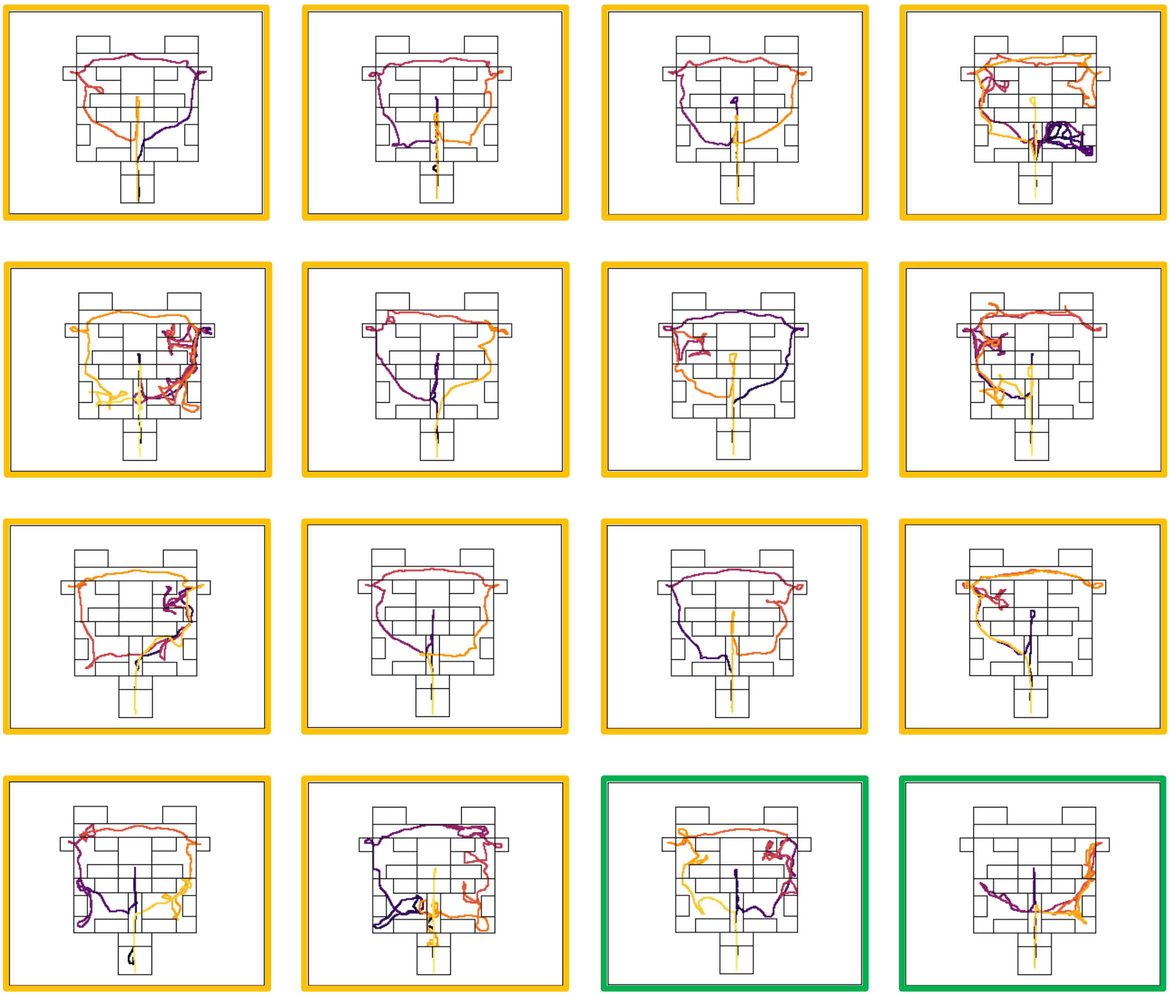
These search patterns are organised by the Explicit-message group, Implicit-message group, and Control group. Floorplan B was flipped horizontally so that the Finance Department is on the *left* side of the building and the Marketing Department is on the *right* side of the building in the diagrams below. Participants who searched both copy rooms (orange) are differentiated from those who searched only one (green) by the colour of the border.

Explicit-message group



Implicit-message group

Control group



Appendix F: Information Framing and Eyewitness Trustworthiness (Additional Experiment)

Experiments 1 and 2 (Chapter 2) investigated the effects of framing in conveying the potential inaccuracy of a floorplan to be used during a search and rescue (SAR) mission in a virtual building. When the floorplan was described as "might (or might not) be useful" for the SAR mission, participants relied on it to search the building even though it conflicted with their prior knowledge of the building and the area labels on the wall and door plaques. Although reliance decreased when the floorplan was additionally described as "might be outdated", a considerable minority continued to use it despite the explicit warning of its questionable utility.

The present experiment examined whether the framing of an eye witness' potential reliability would affect reliance on their report about the presence of a person during search. Prior to a SAR training stage and the "real" SAR missions, participants explored the same virtual office building to be used during the SAR stages in well-lit conditions. During this exploration, they came across a notice that the Fire Exit was jammed and would take several weeks to repair. During SAR training, they entered the now smoke-filled building to rescue two people (Finance Manager and Marketing Manager) trapped in their respective Copy rooms, and retrieve a portable stove. Any attempts at exiting the building through the Fire Exit during the six training trials would trigger the same notice. Before the SAR mission, participants were informed that an eyewitness saw one of the managers (e.g., Finance Manager) had escaped the building via the Fire Exit. Crucially, the reliability of the eyewitness was communicated explicit or implicitly, or not communicated at all. Group Explicit-hint were informed that the witness was "some distance away so he might be mistaken", group Implicit-hint were informed that he was "watching from a distance", and group Control were informed that he was "watching the incident unfold".

The impact of message framing was assessed by whether it influenced participants' behaviour in response to the inconsistency between prior knowledge and the witness report during the SAR mission. If messaging about the witness' unreliability affected participants' reliance on their report, they would be more likely to recognise the inconsistency and rescue both managers before exiting the building. However, if messaging had no effect or was absent, they might rely on the report and exit the building (or attempt to do so) after rescuing only one Manager instead of two.

Methods

Participants and Materials

Seventy participants (58 females, 12 males; mean age = 20.27 years, range: 18.08-37.33 years) with normal or corrected-to-normal vision were recruited from the student population of the School of Psychology, Cardiff University, and received course credit for their participation. The materials and procedures were adapted from Experiment 3 (Section 2.5.1), with changes described below.

Procedure

Participants received the same search and rescue (SAR) cover story as in Experiment 3 (Section 2.5.1.2). The Familiarisation stage, which consisted of an exploration of the virtual office building in well-lit conditions and a floorplan labelling test, were identical to Experiment 3. Similarly, the six SAR Training trials followed the same procedure as before. The notice about the Fire Exit is repeated here for emphasis due to its relevance to the experiment. When participants attempted to use the Fire Exit during exploration, the following notice appeared: "The fire exit is jammed and cannot be opened. We are awaiting parts and repair, which will take several weeks. Please use the Front Entrance until further notice. Apologies for the inconvenience caused and thank you for your patience”. The date at the bottom of the message was programmed to be the day before the experiment to signal that the event described was recent. The same notice would appear under the same conditions during SAR training.

After building exploration and before the floorplan labelling test, participants were asked two attention check questions to assess whether they had read the notice carefully. The questions were prefaced by: “Your trainer is surprised to see you exit the building from the Front Entrance. Out of curiosity, she asks you a couple of questions:” The first question was, “You have just exited via the Front Entrance. Why?” Participants selected one of the following response options: “The Fire Exit is jammed”, “The Front Entrance is nearer”, and “I have no idea why I used the Front Entrance”. The second question was, “When will the Fire Exit be working again?”. Participants chose from: “Later today”, “A couple of days from now”, “Several weeks from now”, or “I have no idea when that will be.”

Search and Rescue Mission. After being instructed to rescue the Finance and Marketing managers and retrieve the portable stove, more messages appeared after a pause: “We’ve just received a report. It is from a former employee of Lemonade tech ...” The remaining sentence differed based on whether participants received an explicit, implicit, or no message (i.e., Control group) about the reliability of the former employee’s report. The Explicit-message group received, “…some distance away so he might be mistaken”, the Implicit-message group received, “…who is watching from a distance”, and the Control group received, “…who is watching the incident unfold”. Following that, the last message was the similar for all participants: “He claimed that he saw the Marketing Manager escape from the building through the fire exit.”, with half of them receiving the same message but referring to the Finance Manager. Participants then pressed the Enter key to proceed to the untimed SAR mission.

Response Measures for the SAR Mission. The framing of the message or hint might influence how easily the inconsistency between the witness report (i.e., a Manager escaped via the Fire Exit) and prior knowledge (i.e., the Fire Exit would remain jammed for several weeks) could be detected. To assess the effect of message framing, the proportion of participants who relied on the report were compared across groups. A participant was considered to have relied on the report if they attempted to exit the building through the (jammed) Fire Exit or exited via the Front Entrance after rescuing only one Manager. Conversely, if participants rescued both managers before exiting the building, it would suggest that they evaluated the report against prior knowledge of the jammed Fire Exit and concluded that nobody could have escaped that way, leading them to infer that both Managers were still inside and needed to be rescued.

When the potential unreliability of the witness' report was explicitly suggested (i.e., "might be mistaken") in group Explicit-message, fewer participants were expected to rely on the report. In contrast, more participants in group Control might rely on the report as they were not informed or hinted of the unreliability of the report. If participants in group Implicit-message inferred that the witness' report was unreliable, search behaviour would resemble group Explicit-message. Otherwise, search behaviour would resemble group Control.

All statistical tests were performed using R, version 4.4.0 (R Core Team, 2024). Analyses were collapsed across the fully counterbalanced factor of virtual environment identity of the building layouts given that there were no significant differences in labelling performance between the environments (Mann-Whitney U test, $Z = 1.87$, $p = .061$) or correct identification of the general direction of the Finance and Marketing Departments in the building (Chi-squared test, $X^2(1) = 0.51$, $p = .473$). When parametric assumptions were violated, non-parametric tests were conducted.

Results

Of the complete sample of 70 participants, 21 failed to answer both attention check questions correctly. Therefore, analyses were conducted on both the complete sample and the subset of 49 who answered both questions correctly.

Familiarisation and Search and Rescue Training. Floorplan labelling was expressed as a percentage of correct labels out of six in Fixed Areas and out of four in Flanking Areas. One-sample Wilcoxon signed-rank tests showed that for the complete sample of 70, participants performed above chance (i.e., mu = 34.62% and 21.70% for Fixed and Flanking Areas, respectively) when labelling Fixed Areas (median = 100%, IQR = 29.17%, $V = 2436$, $p < .001$) and Flanking Areas (median = 50%, IQR = 50%, $V = 2395$, $p < .001$). Of these participants, 64.29% correctly labelled the Finance and Marketing Departments on their respective sides of the building, which was above chance level (binomial test, $p = .023$, 95% CI [51.93%, 75.39%]). The confidence ratings for groups Explicit-message (median = 7.00, IQR = 1.25), Implicit-message (median = 7.00, IQR = 2.00), and Control (median = 7.00, IQR = 1.25) did not differ significantly according a Kruskal-Wallis test, $H(2) = 0.78$, $p = .678$, $\eta^2 = .02$. For the subset of 49, labelling performance was identical for Fixed Areas (median = 100%, IQR = 16.67%, $V = 2436$, $p < .001$) and Flanking Areas (median = 50%, IQR = 50%, $V = 2395$, $p < .001$). Of this subset, 69.39% correctly labelled both departments on their respective sides of the building, which was also above chance level (binomial test, $p = .009$, 95% CI [54.58%, 81.75%]). The confidence ratings for groups Explicit-message (median = 7.00, IQR = 1.00), Implicit-message (median = 8.00, IQR = 1.00), and Control (median = 7.00, IQR = 1.25) did not differ significantly according a Kruskal-Wallis test, $H(2) = 3.75$, $p = .145$, $\eta^2 = .04$.

Figure 1 shows the perceived difficulty to cope scores for all trials pooled across the three groups for the complete sample. The cope scores did not differ during SAR training (Explicit-message: median = 3.67, IQR = 2.88; Implicit-message: median = 3.33, IQR = 2.25; Control: median = 4.75, IQR = 2.75), $H(2) = 1.83$, $p = .400$, $\eta^2 = .00$. During SAR test, median scores were 6.50 (IQR = 2.50) for group Explicit-message, 6.00 (IQR = 3.00) for group Implicit-message, and 8.00 (IQR = 2.50) for group Control. A significant difference was found across groups, $H(2) = 7.61$, $p = .022$, $\eta^2 = .08$, with post hoc Dunn's test confirming a significant difference between groups Implicit-message and Control ($p = .020$) but not between the other pairs (ps = .158-1.00). Regardless of the significant difference, the cope scores were pooled across the three groups to illustrate overall changes across trials and simplify presentation. Figure 1 shows that median scores gradually reduced from the first training trial to the sixth training trial, but increased again during SAR mission. Friedman's test found a significant of training trial, $X^2(6) = 214$, $p < .001$, *Kendall's W* = 0.51. Pairwise comparisons showed that differences between consecutive training trials were sometimes significant and sometimes not (ps < .001-1.00). Crucially, ratings for the first trial were significantly higher than for the sixth trial ($p < .001$), and ratings for the sixth trial were also significantly lower than for the SAR mission ($p < .001$). Findings for the subset of 49 closely mirrored those of the complete sample; therefore, only the complete sample results are presented in Figure 1 for brevity. Analyses repeated with this subsample produced a similar pattern of results, with the exception that the differences in perceived difficulty to cope scores were no longer significant different across groups, $H(2) = 3.75$, $p = .145$, η2 = .04. Otherwise, there were no meaningful differences in statistical outcomes or interpretations.

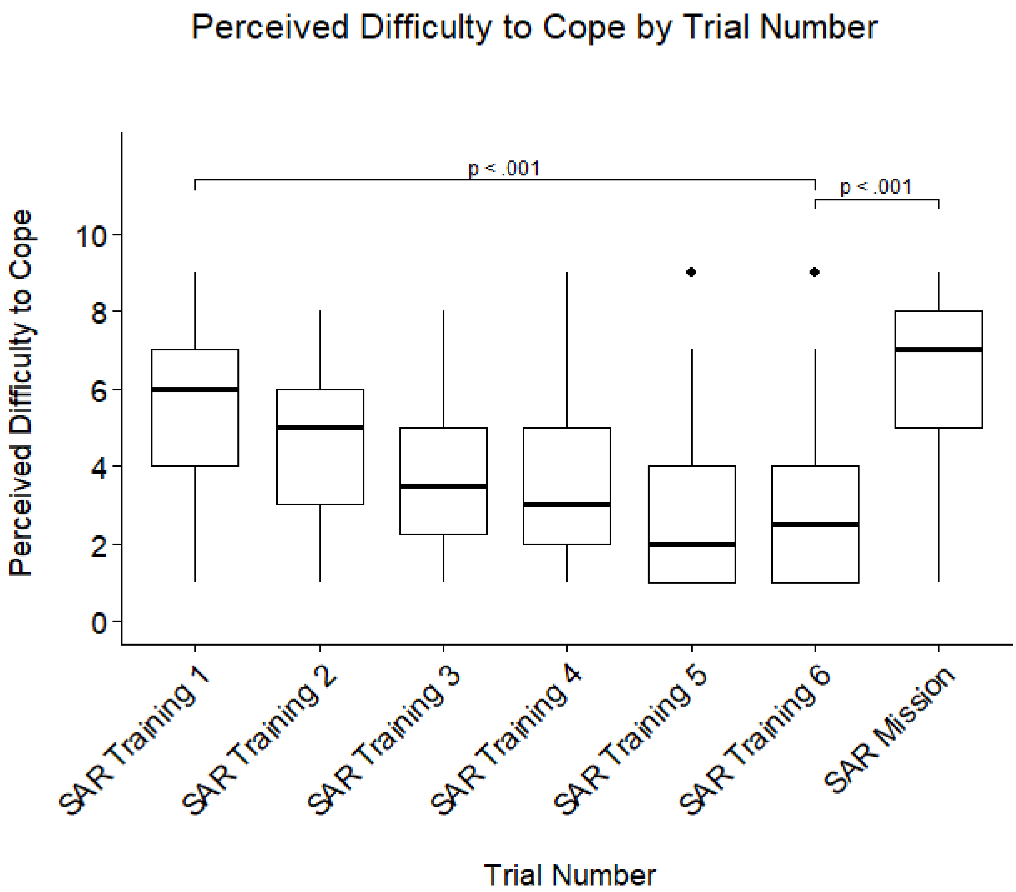


Figure 1. Median perceived difficulty to cope with firefighter search and rescue (SAR) training trials (1-6) and the "real" SAR mission (bold horizontal lines; ±IQR). Outliers were denoted by circles. During training, participants rescued the Finance and Marketing Managers in a virtual building, and retrieved a stove. The final SAR mission was presented as a "real" incident in which their mission was the same, but two groups were given supplementary information that the witness report about the Finance/Marketing Manager having escaped through the Fire Exit earlier was inaccurate, framed in either an explicit or implicit manner. A third control group received a similar report but without information suggesting its potential inaccuracy.

Search and Rescue Mission. For the complete sample, most participants rescued both Managers (82.86%). and retrieved the portable stove (97.14), binomial tests, *p*s < .001. For the subset of 49, most participants rescued both Managers (85.71%) and all of them retrieved the stove (100%), binomial tests, *p*s < .001. Table 1 shows the number of participants (percentages) who rescued both Managers and retrieved the stove in the three groups for the complete sample and the subset of 49. Across Explicit-message, Implicit-message, and Control, the majority of participants rescued both Managers and retrieved the portable stove. Fisher exact tests and a Chi-square test confirmed that these proportions did not differ significantly across groups in the complete sample and the subset of 49, *p*s = .528-1.00.

Table 1: Number (and percentages) of participants who rescued both Managers and retrieved the portable stove

	Complete sample (n = 70)		Subset (n = 49)	
	n	Person / Stove	n	Person / Stove
Group				
Explicit-message	24	20 (83.33%) / 23 (95.83%)	17	15 (88.23%) / 17 (100%)
Implicit-message	26	20 (76.92%) / 25 (96.15%)	16	13 (81.25%) / 16 (100%)
Control group	20	18 (90.00%) / 20 (100%)	16	14 (87.50%) / 16 (100%)

Overall, 28 (40%) of the complete sample and 20 (40.82%) of the subset of 49 relied on the witness report. This includes those who attempted to exit via the Fire Exit or the Front Entrance after rescuing only one Manager. Binomial tests showed that proportion is below chance level, *ps* = .120-.253. Table 2 shows the number of participants (and percentages) who relied on the witness report for the complete sample and the subset of 49. Across Explicit-message, Implicit-message, and Control, the proportions of participants who relied on the witness report was not significantly different in the complete sample, $X^2(2) = 0.12$, $p = .941$, and in the subset of 49, $X^2(2) = 0.33$, $p = .849$.

Table 2: Number (and percentages) of participants who relied on the witness report

	Complete sample (n = 70)		Subset (n = 49)	
	n	Relied on report	n	Relied on report
Group				
Explicit-message	24	9 (37.50%)	17	6 (35.29%)
Implicit-message	26	11 (42.31%)	16	7 (43.75%)
Control group	20	8 (40.00%)	16	7 (43.75%)

Discussion

This experiment examined whether the implicit or explicit communication about the potential unreliability of an eyewitness influenced participants' response to the inconsistency between their prior knowledge and the eyewitness report in a virtual search and rescue (SAR) mission. While exploring a virtual building, participants read a notice that the Fire Exit was jammed and awaiting repairs for several weeks. In the same building, they then practiced rescuing the Finance and Marketing Managers and retrieving the portable stove during the six SAR training trials. Before they participated in a "real" SAR mission that occurred "two days after" the training, they were told that according to an eyewitness, one of the managers had escaped from the building via the Fire Exit. Those in the experimental groups were informed either explicitly or implicitly that the eyewitness might be mistaken, whereas those in the control group received no such suggestion. A sizable minority of participants relied on the eyewitness report during the SAR mission and this proportion did not differ significantly between the three groups, regardless of their attention check performance. However, binomial tests indicated that their behaviour did not significantly differ from chance, suggesting that neither the explicit, implicit, nor control message had a consistent effect on search behaviour.

Conceptually, this experiment bore similarities with Experiments 1 and 2 (Chapter 2) in that these experiments investigated the framing of information regarding the potential disutility of newly received and seemingly important information for the SAR mission. Where it differed from these experiments were the additional SAR training trials and more importantly, the nature of the new information provided. In Experiments 1 and 2, participants who were informed that the floorplan “might (or might not) be useful” consistently relied on the floorplan despite its inaccuracy. In contrast, the Control group in the present experiment (who were not explicitly or implicitly informed of the eyewitness’ potential unreliability) did not behave consistently during the SAR mission. This contrast seemed to suggest the possibility that participants might be more inclined to take certain types of information (e.g., official-looking documents like floorplans) at face value compared to eyewitness accounts.

In comparison with Experiment 3 (Chapter 2), the present experiment also shared the following similarities: inclusion of six SAR training trials and more importantly, the provision of information directly relevant to the SAR goal (i.e., the likely presence of individuals in the building). They key difference was that the information provided in Experiment 3 was accurate whereas the eyewitness report in the present experiment was inaccurate. In Experiment 3, participants who were explicitly informed of a manager’s absence in the building adapted their search behaviour accordingly, but not those received an implicit or an unrelated message. In the present study, however, participants who were explicitly warned that the eyewitness “might be mistaken” behaved similarly as those who were told that the eyewitness was “some distance away” or had merely “witnessed the incident unfold”. While speculative, this contrast between both experiments suggests that detecting inaccuracies in newly received information might be less straightforward, even when participants were explicitly warned about the source’s potential unreliability.

Appendix G: Instructions for Eye-Tracking Calibration (Experiment 4)

1. Please orient yourself towards the orange marker and keep your head still throughout the calibration process.
2. Roll your eyes repeatedly until you hear “stop!”
3. Six targets will then appear at different places on the screen consecutively, starting from the centre.
4. Please look at the target for the duration that it is on screen. It is important for the calibration process to be done properly for your data to be usable.
5. At the end of the calibration, there will be an option to re-do it if you feel that you hadn't done it properly.
6. Press the [big left button] to start.