Complex Decision-Making Under Threat

Aaron Laycock

Doctor of Philosophy

University of York

Psychology

October 2025

Abstract

Research suggests that threat may disrupt complex decision-making. Nevertheless, findings are mixed, likely due to indirect, incidental experimental manipulations of threat, an overreliance on summative performance measures, and underexplored motivational influences. This thesis confronts these methodological and conceptual gaps. For my first empirical chapter, I developed and validated a virtual world to manipulate threat during a complex decision-making task. In this paradigm, threat was directly linked to decision-making (and not merely a distraction from the decision-making task). Participants performed worse under threat, taking longer to improve from baseline and scoring lower through the final trials. Moreover, computational modelling revealed that participants in the threat condition were more responsive to short-term rewards and less likely to perseverate on a given choice. In my next empirical chapter, we used an adapted probabilistic reversal learning task to examine decision-making under two specific forms of uncertainty: volatility and stochastic variability. Threat led to higher learning rates and inverse temperature estimates, indicating increased responsiveness to feedback and a greater tendency to exploit current beliefs. These two sets of studies suggest that contrary to the notion of threat rigidity, being under threat increases reward responsivity in ways that affect performance differently as a function of context. The final empirical chapter investigates the effect of evaluative threat on motivational factors relevant to complex decision-making. Specifically, I examined the effects of real-world military training on the motivation to engage with complexity. Results suggest that the evaluative nature of training might interfere with one's willingness to deal with complexity. However, formative training can mitigate this risk, which may promote motivation to engage with uncertainty. Altogether, these findings help explain current contradictions in the literature. They furthermore offer a foundation for future work supporting individuals who must make difficult decisions within hazardous environments.

List of Contents

Acknowledgements

I write these acknowledgements on the 20th of September 2025 from a small room in Frognerseteren. I still have a little work to do, but I am making progress. For the first time in a long time, it's quiet, and I am alone. This makes recognising the impact of those around me on the demanding process of constructing and presenting this thesis easy.

I must first thank my supervisors. Professor Cade McCall has given me unconditional support, patience, and direction through every stage. His proprietary blend of intellectual humility, landmine wit, and infectious curiosity has been comforting and inspiring. Professor Harriet Over has given me faith and guidance throughout. When Harriet talks, I listen. I feel privileged to have spent time with these extremely unique, thoughtful, and accomplished humans.

In support have been Professor Beth Jefferies, Dr Catherine Preston, and Professor Katie Slocombe. They were always diligent, altruistic, and committed to helping guide me through this process. Through all the criticism and praise, I have never felt that I was not being listened to. Anybody who knows me will understand how remarkable this truly is, and the ridiculous amount of energy it must have required. For this sacrifice, above all, I am grateful.

My partner Tara has been with me every step of the way. In many ways, she has invested more in all this than I, especially given my unusual sleeping habits when I am in a fist-fight with a problem. Of which I have had countless. It's been a long process. Seven years before the PhD started, Tara encouraged me to start the first of many courses required in preparation for this PhD. Now, ten years after that day, her commitment is as unyielding as ever.

I can't lie. Honestly, I am still not sure my mother, Bimbo, Nana, Amarian, and sisters, Jay and Dee know what I have been up to. It's never mattered. If at some point in the future any of them come across this thesis and see this section, understand that having you all around ensured perspective, helped me remain grounded, and protected me from the absurdity of taking myself and this (at times) brutal process a little too seriously.

I must also acknowledge the role my friends have played. Outside of University, Nemo, Marty, Yatesy, and O'tool daily remind me of the importance of friendship. From University, Delali, Bronte, Emma, Anna, Jo, Lewis, and Wendy to name only a few, have provided a playground for exploration, and a constant source of feedback (positively biased). The importance of which should not be underestimated.

In an early draft of my proposal to enter the PhD program (one that did not survive the red pen), I commented: "My principal motivation for wishing to join this PhD program is to develop the ability to dissociate good ideas from bad ones, towards understanding…". I now realise this was a lofty goal. The last few years have been an uncomfortable transition from hazy certainty toward lucid ambiguity. Yet, there is no doubt that I have made tremendous progress in moving towards my aim.

This experience is a treasure.

This work was supported by an Economic and Social Research Council Advanced Quantitative Methods Award [2602944]

Author's declarations

I, Aaron Laycock, declare that this thesis is a presentation of original work, and I am the sole author, under the supervision of Professor Cade McCall (Primary Supervisor), and Dr Harriet Over. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references. The following chapter from this thesis has been published in a peer-review journal:

Chapter Two from this thesis has been published in a peer-review journal: Laycock, A., Schofield, G., & McCall, C. (2024). The effects of threat on complex decision-making: evidence from a virtual environment. Scientific Reports, 14(1), 22637.

Chapter Four is under peer review for publication.

Chapter One: Introduction

Introduction

In everyday life, we often make decisions under uncertainty. This uncertainty may emerge from something as simple as unfamiliarity with the context (Fantino & Stolarz-Fantino, 2005) or perhaps previously good choices might on occasion lead to bad outcomes (Levy & Schiller, 2021; Simoens et al., 2024). These factors all set the stage for complex decision-making (Funke, 2012): decisions made based on uncertain information.

To deal with uncertainty optimally, an individual must make their best guess by sampling information from the environment (Globig et al., 2021), learning from experience, and making choices based on probabilistic inferences (Gonzalez, 2022; Rakow & Newell, 2010). Real-world accounts from first responders and military personnel suggest that complex decision-making is especially challenging in dangerous environments (Harris et al., 2017; Lieberman et al., 2005; McCall & Laycock, 2024). Particularly when self-preservation is at stake under threat and the potential for absolute loss can outweigh the significance of incremental gains (Fanselow, 2018).

In a recent project, our lab used phenomenological interviews to probe subjective reports from individuals who have been asked to make real-time choices in dangerous and unpredictable environments (McCall & Laycock, 2024). These reports describe how threat presents in many forms. In addition to threat posed by environmental dangers, threat can be evaluative, such as loss of reputation. The process of dealing with threat in these reports is marked by uncertainty. This uncertainty was defined as an inability to assign certainty to an outcome, as decision-makers seek to reach workable solutions to dynamic and ill-defined problems (Dörner & Funke, 2017), with potentially life-threatening consequences of action. Interestingly, threat was not seen by these individuals solely as a distraction to be ignored, but as integral to the decision-making process.

Early work into how threat affects cognitive performance began after World War II, sparked by reports that skilled peacetime pilots frequently made critical mental errors and crashed during combat operations (Arnsten, 2009; Broadbent, 1971). While the ethical and practical challenges of manipulating threat experimentally have made progress difficult in this area, evidence suggests that threat impacts brain neurochemistry and function (Arnsten, 2009; Ben-Zur & Zeidner, 2009), and has been shown to influence attention (Azriel & Bar-Haim, 2020), memory (Vytal et al., 2012), learning (Fanselow, 2018; Starcke & Brand, 2012), and behaviour (Simonovic et al., 2018) in ways that are likely to disrupt complex decision-making.

Nevertheless, questions remain regarding the nature of complex decision-making under threat. This chapter will first give a brief overview of complex decision-making. I will then discuss factors that underlie complex decision-making performance, reviewing an array of psychological mechanisms and individual motivational differences. Next, I will discuss threat and how it has been studied in laboratory and naturalistic settings. Finally, I will review existing research specifically on the effects of threat on complex decision-making.

Studying complex decision-making

Operationalising complex decision-making for the purposes of research is a considerable challenge. At a basic level, complex decision-making is decision-making under uncertainty. One does not have a complete set of information and, as a consequence, cannot ascribe certainty to any given outcome (Soltani & Izquierdo, 2019). This point contrasts with a classical view of decision-making, which can be thought of as trying to finish a Jigsaw puzzle. As each piece has a place, the job is to match the right pieces with the appropriate space, making best use of the completed example on the box cover as a concrete illustration of ground truth.

Complex decision-making, on the other hand, involves uncertain but (often) probabilistic outcomes. When psychologists use the word uncertainty, an example of coin flips and known probabilities often follows. However, with complexity (Rickles et al., 2007), this is not the case. Probabilities are not necessarily reliable and are built up over time by sampling options. Critically, this is distinct from classic decision-making, which may involve a great deal of information, but all the information is available, and outcomes are absolute. With complexity, the likelihood of outcomes may change over time and across contexts, such that optimal performance requires an individual to endure problems, not simply make a selection with the highest utility (as this is unknown) and stick with it. Here, metaphorically, the aim is still to complete the jigsaw, but without the privilege of the box to guide inference. Indeed, even if the box were available as a template, it would be of limited use, given the issue of degeneracy, a term used to describe many to one (or one to many) associations (Edelman & Gally, 2001). In other words, choices made in complex environments can lead individuals down differing paths to the same (or different) destinations.

A variety of research on complex decision-making examines real-world decisions. Using the naturalistic decision-making approach (NDM), researchers interview individuals about high stakes decisions made under uncertainty, discussing past or hypothetical decision-making experiences (Gore et al., 2015; Klein, 2008; Lipshitz et al., 2001). This approach, therefore, examines complexity as it is described by individuals with previous experience, e.g., fire service commanders (Klein, 2008; Klein et al., 2010). While the approach provides rich, detailed accounts of real-world decision-making (Schmitt & Klein, 1999), it is built upon the premise that decision-making policies that have once been successful will continue to be.

This assumption may downplay the dynamic effects of complexity and context (Fantino & Stolarz-Fantino, 2005) on performance.

For example, an experienced firefighter may have plenty of experience dealing with a fire in an apartment block. It's feasible that these experiences can help shape, to some degree, a broad response to the situation. That is, some factors will help optimise choices when dealing with fires in an apartment that are independent of context, e.g., dealing with multiple flights of stairs with heavy equipment. Yet, much of the cognitive workload in any complex situation may be event-specific; just because a fire damaged floor held your team's weight (with equipment) last time does not mean it will this time.

Moreover, interviewees can only tell us about cognitive process of which they are consciously aware and which they remember (Eccles & and Arsal, 2017; Miller, 1981), and some reports may just be too sensitive to disclose (e.g., professional incompetence), and cannot be verified independently in an economical, practical, or ethical way (Brener et al., 2003). So, while the NDM approach provides insights into the experience of real-world decision-makers and highlights the critical role of memory and attention in negotiating complex tasks (discussed later), the dependence on retrospective self-report limits our conclusions about the dynamic cognitive processes that emerge in the moment (Keren et al., 2013).

An alternative approach to overcome the challenges of retrospective research methods is to synthetically create the environmental context under investigation (Sergiou et al., 2024). Laboratory based paradigms attempt to address this challenge by creating tasks that capture the core aspects of uncertainty that make a task complex (Dörner & Funke, 2017; Greiff et al., 2013). For example, "the fungus eater" (Toda, 1962) was an interactive task that utilised the computer technology of the day to simulate task complexity. The fungus eater project was swiftly abandoned due to challenges (in part) posed by the scalability of these early technologies to approach what may be considered complex in the real world (Brehmer, 1992).

However, by the 1970s, "microworlds" (Turkle, 1984) were increasingly employed to investigate complex decision-making performance. In these paradigms, participants are asked to solve problems presented via a computerised display, which involves using multiple streams of information to reach a desired state/outcome over a prolonged period (Brehmer, 2005). The use of microworlds was made in response to a need to develop research methods that broadened the computational demand placed on participants and better reflected real-world complexity, yet maintained a degree of experimental rigour (Omodei & Wearing, 1995).

These elaborate computer-simulated microworlds have taken many forms over the years, employing a range of cover stories (i.e., narratives to frame that task in real-world terms).

For example, in the *"MORO"* world (Dörner et al., 1986), participants are asked to control a range of variables over time to enhance the living circumstances of a small fictitious tribe. Alternatively, in the *"Fire chief"* microworld (Omodei & Wearing, 1995), participants play the role of a local fire chief, responsible for allocating resources (e.g., workforce, water, and methods of fire prevention) to control the spread of a forest fire efficiently. This is achieved by making choices over time to moderate the outcome of several externally controlled factors (e.g., wind direction and speed). Although microworlds vary greatly in terms of the cover story and the number/type of variables linked to producing a successful outcome, the central aim is generally always to use a simulated digital environment to observe the process of decision-making over time, under uncertain experimental constraints.

The central aims of microworlds are also shared with more conventional multi-armed bandit tasks (Schulz et al., 2020). Multi-armed bandit tasks restrict the degrees of freedom presented to participants experimentally and focus on targeting quantifiable behavioural processes. These behavioural tasks may be viewed as less computationally demanding in terms of scope, yet critically, they may still be considered complex because of the uncertain outcomes of multiple options. For example, participants may be asked to choose a card from one of four decks. Each deck will be associated with an unknown payout representing various levels of uncertainty, or frequency/scale of costs and rewards. This procedure will be repeated over many iterations. Observing the decisions individuals make on such paradigms can reveal a great deal with regard to the processes individuals employ when making decisions over time (Steingroever, Wetzels, Horstmann, et al., 2013).

Perhaps the best known and influential multi-armed bandit paradigm is the Iowa Gambling Task (IGT; Bechara et al., 1994). The IGT was developed to assess decision-making deficits associated with ventromedial prefrontal damage. However, the IGT has increasingly been used in non-clinical populations (Bull et al., 2015) to operationalise complex decision-making (Buelow & Suhr, 2009). This trend has in part been driven by observations that performance on the IGT varies greatly among populations with no diagnosed deficit in decision-making ability (Steingroever, Wetzels, Horstmann, et al., 2013). Thus, researchers can use the IGT to identify individual differences and elucidate possible underlying mechanisms responsible for sub-optimal performance on the IGT related to, for example, self-reported decision-making style (Franken & Muris, 2005), personality characteristics (Buelow & Cayton, 2020), tolerance for uncertainty (Kornilov et al., 2015).

The reward structure of the IGT, although not random, is initially unknown to participants. Over the course of trials (generally one hundred), participants are asked to select cards from an array of four options (Decks: A, B, C, and D). Each selection results in wins and losses of facsimile dollars. Participants are asked to keep making selections with the aim of selecting from advantageous decks that result in a net positive reward. No indication is given as to how long the participants must carry on making selections (e.g., over an infinite horizon).

Decks A and B are disadvantageous. Selecting from these decks may initially appear optimal yet result in a net loss over time. Conversely, decks C and D are advantageous. Choosing from these decks results in a net positive reward over time. Thus, optimal performance requires an individual to track the long-term reward of selections.

Performance on the IGT has traditionally been measured by calculating the sum of advantageous choices after subtracting disadvantageous ones (Bechara et al., 1994). The process of learning can also be observed by analysing performance across blocks, generally bins of twenty trials (Bull et al., 2015; Kornilov et al., 2015). One issue with these traditional metrics is that they are a coarse measure of performance (Ligneul, 2019). That is, although informative, such methods may not fully capture the degree to which individuals differ in decision-making strategy (Sullivan-Toole et al., 2022). Given the nature of complexity (i.e., the inability to ascribe certainty to an outcome), optimum behaviours may, in the short term, result in suboptimal outcomes (performance degeneracy). Complex decision-making is probabilistic, and thus, measures of optimal performance should reflect the fact that decision-making is a process evolving over time and in response to feedback from the environment. To judge the quality of decisions, we must look beyond simple measures of outcome.

To address this issue of performance degeneracy, researchers have shifted the focus from the outcome of a decision to the underlying processes. As previously discussed, alternative metrics can be calculated from data gathered from multi-arm bandit tasks that may offer a greater understanding of how participants negotiate complex tasks (Fellows & Farah, 2005). For example, alternative measures can reveal the balance between exploring a problem space by individuals in a quest for more information and exploiting an already held body of knowledge (Sayfulina et al., 2020). Here, measures of choice perseveration, such as run length, e.g., consecutive deck selections (Zeif & Yechiam, 2020), deck switching, e.g., frequency of switching as calculated by the standard deviation of selections (Mussey et al., 2015), and decision time, e.g., average time invested in each trial (Oberdörfer et al., 2021), have proven useful.

Along similar lines, a range of computational models are increasingly being used in the analysis of multi-armed bandit data (Steingroever, Wetzels, & Wagenmakers, 2013). These methods allow for comparisons between groups (Moreno-Padilla et al., 2022) on a range of parameters that reflect learning and decision-making mechanisms such as reward sensitivity, loss aversion, and learning rate (Ligneul, 2019; Mkrtchian et al., 2021).

Cognitive mechanisms underlying complex decision-making

Fundamentally, the act of decision-making is the process of an agent (human or otherwise) selecting options among possible alternatives (Gilhooly, 1988). Yet the way decision-making is conceptualised can have a profound influence on one's approaches to research and the theories that consequently are developed (Lerner et al., 2015). The early decision-making literature assumed that human agents engage in the process of decision-making rationally (Geng et al., 2022). This rational model of decision-making works from the premise that an individual is driven to maximise outcome satisfaction and, critically, has all the required resources (e.g., information and computational capacity) to do so.

However, this is very rarely the case outside of the research laboratory. An alternative conceptualisation of real-world decision-making is to think of rationality as “bounded” by conditional factors (i.e., computational capacity or environmental constraints) that can vary and are often outside of a decision-maker's influence (Simon, 1955). In response, individuals may be driven not to maximise reward as a rational model assumes but rather to "satisfice", a term used to describe any option that meets some appropriate "just good enough" threshold (Tyszka et al., 1989).

Attention and Memory

Here, attention likely plays an important role in complex decision-making by allocating limited cognitive resources to make sense of uncertainty in information-rich environments. Indeed, attention predicts choice preference and, not surprisingly, unattended stimuli have little influence on decision-making tasks (Leong et al., 2017). It is also clear that higher-order operations such as complex decision-making rely on the interaction between attention and working memory (Oberauer, 2019). Working memory is itself a limited resource associated with holding small amounts of information in the service of higher order operations (Baddeley, 2007). Its role in complex decision-making is evidenced by a positive relationship between working memory capacity and a tendency towards selecting advantageous decisions in the IGT (Bagneux et al., 2013). Together, attention and working memory likely support complex decision-making performance by both performing informational triage and also enduringly updating estimations of what information is goal-relevant and what is not (Rolls, 2007).

Longer term memory is also important for complex decision-making. At a basic level, episodic and semantic memory are critical to the ability to imagine the future (Gilbert & Wilson, 2007). As such, they are likely to facilitate the ability to engage in prospective cognition about the potential outcomes of complex decisions (Pezzulo & Rigoli, 2011). Real-world analyses of decision-making support this notion. A common theme in naturalistic decision-making research is the Recognition-Primed Decision Model (Klein, 2008). The model was not built on a robust body of empirical evidence from lab-based experiments but rather from the accounts of lived experience, initially that of fire service commanders' responses to hypothetical scenarios (Klein et al., 2010). The argument here is that knowledgeable decision-makers look to their previous experiences to extract policies that have been successful. Further, the model highlights that individuals will adopt the first feasible option that meets the practical requirements to achieve success in any given situation rather than investing effort in attaining an optimal outcome, i.e., maximising (Shortland et al., 2020). This model describes how individuals use “patterns” to negotiate uncertainty. The functional term “patterns” refers to markers of the most predictive cues and meaningful inferences extracted from previous exposure. The patterns allow individuals to negotiate novel situations by providing a verified scaffold upon which to base future action. Over time, individuals accrue an array of typical courses of action for any given situation that can be brought to bear over the course of an unfolding event. In this sense, learning over time is central to complex decision-making.

Learning

Learning has also been a key theme within laboratory-based research on complex decision-making. Here, the idea is that complex decision-making performance is at least partly a function of reinforcement learning (Levy & Schiller, 2023; Sutton & Barto, 2018). An individual learns to optimise utility by actively sampling information from the environment and interrogating it closely, making choices that are sensitive to both immediate feedback (reward signal) and projected long term choice satisfaction (expected utility).

Early research using the IGT focused on the possibility of implicit learning in complex tasks. The Somatic Marker Hypothesis (SmH; Damasio, 1994; Damasio et al., 1996) suggested that in uncertain or ambiguous situations, choices are influenced by physiological markers, or "somatic states", that provide a representation of prospective value. These states emerge from associations made during experience. The SmH's fundamental assertion is that the influence of these physiological markers on decision-making comes before conscious awareness of any choice preference. Support for the SmH primarily came following a series of studies using the IGT that appeared to demonstrate that participants experienced increased arousal (Bechara et al., 1994) following a disadvantageous card selection, critically, before they had become consciously aware that they were selecting from a "bad" deck. However, Maia and McClelland (2004) challenged this assumption by demonstrating that individuals have far more explicit knowledge than proposed by Bechara et al., (1994). The debate is yet to be settled, with further investigation into the data collected by Maia and McClelland (2004) doing little to address the central issues of disagreement (Chiu et al., 2022). However, the robust effects of the early work supporting SmH have seldom been replicated (Fernie & Tunney, 2013; Steingroever, Wetzels, Horstmann, et al., 2013).

Reinforcement learning

Regardless of the role of somatic markers, reinforcement learning remains a key theme in research on complex decision-making. Several features of reinforcement learning are important here. Optimal complex decision-making relies on an optimal balance between exploration and exploitation. Inherent in decisions made under uncertainty is a dilemma: do you stick with (exploit) a selection that has offered previous reward, or seek to explore other options in search of better outcomes? No doubt at some point we have all been familiar with this trade-off (e.g., dating, employment, or food choice). To negotiate this challenge and boost long-term reward, an individual must collect information from the environment, risking the possibility of losing out in the short term (Schulz & Gershman, 2019). Understanding how individuals act in the early exploration phases of the decision-making process when individuals have no (or limited) information on which to base choices may be particularly important to understanding downstream performance on complex tasks (Mehlhorn et al., 2015; Sayfulina et al., 2020).

Exploration and exploitation

Learning and adapting behaviour as a function of experience is at the core of the exploration exploitation trade-off (Cohen et al., 2007). The use of multi-armed bandit tasks, computational modelling and non-invasive imaging methods have been employed to study this trade-off more closely and identify underlying neural mechanisms. Results here demonstrate that brain regions of interest considered fundamental to value-based decision-making, like the ventromedial prefrontal cortex (Schneider & Koenigs, 2017) are also employed at key stages during trade-off (Daw et al., 2006). Further, the use of computational methods has given researchers the ability to model the exploration and exploitation trade-off algorithmically and predict behaviour on ill-defined tasks (Schulz & Gershman, 2019). As such, the exploration exploitation trade-off is found to be integral to complex decision-making. This can be shown in work identifying the importance of sequential choice exploration on performance (Ligneul, 2019), as well as research that demonstrated that expert decision makers actively engage in the trade-off (Laureiro-Martínez et al., 2015).

Context dependent learning

The role of context in complex decision-making is also highlighted by the importance of context dependent learning. One aspect of complex decision-making performance is the ability of an individual to learn from environmental cues and adapt choice behaviour as a function of feedback. Tracking regularities in a complex environment may be very difficult, given that any associations made are likely to be transient, varying as a function of shifts in task and context (Gonzalez, 2022). Learning across contexts requires an individual to recognise failure successfully and then adapt behaviour to minimise future error. First, an agent must identify contextual change. To do so, an individual must identify subtle variations in environmental cues. An influential body of research has looked to view this decision-making process through the lens of experience. Here, the term “decisions from experience” (DoE; Gonzalez et al., 2005; Hertwig et al., 2004) denotes situations where individuals are required to develop a body of knowledge (outcomes and probability) via exploration of the experiment. Instance-Based Learning Theory (IBLT; Gonzalez et al., 2003) was developed to help understand the link between experiential inference and choice behaviour.

As far as IBLT is concerned, performing well on dynamic decision-making tasks is about making associations between a decision made (value-based, appropriate actions), the utility gained (index of satisfaction), and a set of environmental cues within a specific context (Gonzalez, 2022). IBLT argues that at the core of complex decision-making performance is the ability to learn from experience via the accrual of knowledge in the form of instances. The situation, decision, and utility (SDU) slots are compiled and bonded as a functional instance, and it's these instances that are the procedural building blocks an individual uses to navigate novel and dynamic situations (Gonzalez et al., 2003). Support for the IBLT can be found in the ability of computational models developed on the principles of IBLT (i.e., whose parameters operationalise different aspects of instance learning) to predict behaviour across a diverse range of tasks that require decisions to be made from experience (Gonzalez, 2013; Gonzalez & Dutt, 2016; Mehlhorn et al., 2015). The success of IBLT here has led to real-world application in applied settings, e.g., cyber defence (Ben-Asher & Gonzalez, 2015; Gonzalez et al., 2020), which may be considered testimony to the robustness of the theory.

Cognitive flexibility

However, the identification of environmental cues may be considered necessary but not sufficient regarding context-dependent learning. An agent must also recognise prediction error and critically adapt behaviour accordingly. Such operations can be defined by cognitive flexibility, e.g., adapting preference with environmental change and are studied typically using reversal learning tasks (Highgate & Schenk, 2021). Reversal learning is a central feature of cognitive flexibility (Izquierdo et al., 2017) and may be considered critical when an individual is required to make choices based on uncertain information. Regarding complex decision-making performance, it is necessary to perpetually update estimations of what information is predictive of success and what is not.

This challenge is compounded by the fact that informational uncertainty can be expected and/or unexpected (Soltani & Izquierdo, 2019). For example, the decision to call the police after a loved one has not returned home is uncertain. This choice will be impacted differently if they are always late (expected uncertainty) or are home at 6 pm promptly without fail (unexpected uncertainty). Regardless of this increased difficulty in knowing when the call should be made, in both cases, not engaging in reversal learning is a bad strategy, e.g., never ring because they are always late or ring at one minute past 6 pm because they are never late. To complicate matters further in complex situations, stochastic variability (i.e., good choices sometimes lead to bad outcomes) and volatility (i.e., previously rewarded good choices can simply stop being so) are often also in play (Levy & Schiller, 2021; Simoens et al., 2024). Thus, the ability to engage in reversal learning efficiently is key in promoting behavioural flexibility and optimising decision-making ability under uncertain constraints (D'Cruz et al., 2011).

Reward processing

Finally, reward processing is also a key factor that may influence complex decision-making. Despite IBLT identifying that the utility (i.e., satisfaction) of a decision must be tracked, the model does not account for the fact that reward or loss sensitivity might vary between individuals or between situations. The use of complex multi-armed bandit tasks like the IGT demonstrates that individuals often make selections with different preferences in mind. For example, on the IGT, optimum performance is associated with identifying (and selecting) the option that results in the highest net score over time. However, it is sometimes the case that individuals are motivated by a need to limit the frequency of losses (independent of amount) or respond to immediate feedback with little concern for the accumulation of reward (Steingroever, Wetzels, Horstmann, et al., 2013). Acknowledging these differences in reward processing is essential to understanding the drivers of complex decision making.

This may be achieved with the use of computational modelling methods that can be used to identify systematic variation in the ways individuals process reward and identify how these differences shape decision-making behaviour (Yechiam et al., 2005). Returning to the example of the IGT, a computational model such as the Outcome Representation Learning (ORL) model provides a series of parameters that index variations in reward processing, e.g., frequency, reward, and loss aversion (Haines et al., 2018; Sullivan-Toole et al., 2022). Predictions generated from the ORL model accurately predict behaviours on the IGT across multiple versions, which vary in terms of payoff structure (Haines et al., 2018). Prediction accuracy and parameter recovery of the ORL model outperform alternative reinforcement models such as the Prospect Valence Learning and Value-Plus-Perseverance models, which critically attempt to explain behaviour on the IGT with a reduced number of parameters (Haines et al., 2018). Further, as the ORL model utilises a Hierarchical Bayesian approach (Kruschke, 2014; Shiffrin et al., 2008), well documented issues with computational modelling in dealing with nested data, e.g., inability to extract robust individual differences (Estes & Maddox, 2005) can be overcome. This advantage has provided the ability to identify individual differences (Shiffrin et al., 2008) in reward processing across populations (Moreno-Padilla et al., 2022) and demonstrates that complex decision-making is not just about learning to maximise an arbitrary reward, what constitutes a reward is also a vital component.

As we have demonstrated, a wide range of psychological mechanisms play a role in driving complex decision-making performance. This brief review has highlighted the role of memory, attention, exploration/exploitation, learning, and reward processing. The list may not be exhaustive, but it identifies processes underlying complex decision-making that may be shaped by threat. Moreover, this review suggests that these processes can be measured through computational models of complex decision-making tasks. However, before we look at the potential influence of threat on these processes, one final factor that must be briefly reviewed is the influence of individual differences in motivational drive on complex decision-making.

Motivational mechanisms underlying complex decision-making

The above discussion suggests that optimal complex decision-making depends upon a variety of cognitive abilities. Research also suggests that motivation is important. Two individual differences that have been connected to one’s willingness to grapple with complexity are intolerance of uncertainty (Carleton et al., 2007) and the need for closure (Roets & Van Hiel, 2011).

Intolerance of uncertainty is a trait defined by negative beliefs about uncertainty and its consequences. Individuals who report low tolerance for uncertainty view uncertain future events as threatening (Chen & Lovibond, 2016). Individual differences in intolerance of uncertainty have been suggested to be predictive of a wide and diverse range of emotional, cognitive, and behavioural responses to real-world phenomena (Bavolar et al., 2021). Intolerance of uncertainty predicts risk-averse behaviour (Carleton et al., 2016), reward sensitivity (Williams et al., 2024), repetitive thinking (McEvoy & Erceg-Hurn, 2016), reduced confidence in decisions, being behaviourally inhibited in unpredictable situations (Jensen et al., 2014) and more frequent selection of an immediate, less valuable reward in decision-making tasks (Luhmann et al., 2011). As such, intolerance of uncertainty is related to a broad range of behaviours not conducive to optimal negotiation of complexity (Birrell et al., 2011; Kornilov et al., 2015).

The need for closure (Roets & Van Hiel, 2011) has also been linked with many aspects of decision-making. The need for closure is conceptualised as one's motivational drive to find a concrete answer to an ambiguous situation (Webster & Kruglanski, 1994). Therefore, in contrast to intolerance of uncertainty, need for closure is focused more directly on goal-directed motivation (Berenbaum et al., 2008; Kruglanski & Webster, 1996; Roets et al., 2015). Higher scores are associated with a need for decisiveness (Kruglanski, 2013) and unyielding adherence to pre-existing knowledge structures and expectations (Roets et al., 2015). Theoretically, based on this finding, individuals low in the need for closure may be better suited to negotiating complexity as they may be more inclined to embrace information that is not explicitly linked with a predetermined end state and engage in non-routine decision-making practice. However, this view has been challenged, with some research suggesting the opposite. That is, high need for closure drives motivation to find solutions to ill-defined problems and should therefore be predictive of improved performance as a function of increased effort invested in finding closure (Sankaran et al., 2017).

Regardless, the above findings suggest that individual differences in the motivation to engage with uncertainty may influence complex decision-making. If a greater motivation to engage with uncertainty indeed increases complex decision-making performance, then further questions arise regarding the effect of threat on these motivations.

Threat

While the idea that danger promotes a universal, fight or flight response can be a useful framework to categorise and interpret distinct groupings of behaviour (Milosevic & McCabe, 2015), it is (at best) an oversimplification. As highlighted by Joseph LeDoux, “Fear is often said to be universal. But instead, what is universal is danger. The human experience of being in danger is personal and unique” (Mobbs et al., 2019). Humans (Sporrer et al., 2023), and even animals (Evans et al., 2019) respond to threat in sophisticated ways, integrating information sourced rapidly from the environment to adapt action to meet contextual demands. In this section, I will discuss my working definition of the term threat, before briefly reviewing the methodological and ethical challenges in empirically investigating how being in a threatened state can influence complex decision-making.

I define the term threat as a perceived likelihood of impending physical or psychological harm. Being in a threatened state is a psychophysiological phenomenon associated with heightened levels of physiological arousal (Steimer, 2002) and psychologically subjective aversion (Han et al., 2015). A threatening environment is simply one in which the state of threat is promoted. Threat may be experienced as an evaluative interpersonal phenomenon (e.g., the projected loss of reputation or status) or as a result of environmental hazards (McCall & Laycock, 2024; Park et al., 2023). Regardless of how threat is experienced, it is generally associated with a negatively valenced response to a situation or stimuli that is paired with heightened physiological arousal (Hildebrandt et al., 2016) with the potential of a cascade of endocrine activation (Ulrich-Lai & Herman, 2009).

The above definition overlaps with terms used to describe affective states of fear and anxiety. Fear is often described as a transient response to a known threat, and anxiety a longer lasting (tonic) consequence to an uncertain one (Davis et al., 2010; Rosen & Schulkin, 1998). However, given the increasing difficulty in reaching ontological consensus across the affective sciences (Grogans et al., 2023; Mobbs et al., 2019) and challenges in dissociating the proposed states physiologically (Daniel-Watanabe & Fletcher, 2022), I refrain, where practical, from using rigid distinctions and assume that no single state will dominate across the full arc of any given emotive experience.

This definition of threat is also closely related to that of stress, an emotionally arousing experience that is interpreted as a physical or psychological risk to homeostasis (Levine, 2005). The term stress is often used more broadly than threat (Bienertova-Vasku et al., 2020; Putwain, 2007) and is not necessarily negatively valenced. Both terms (threat and stress) are routinely used interchangeably. This is despite evidence that the way stress is internalised by an individual (e.g., as a threat versus a challenge) can lead to different behavioural outcomes on decision-making tasks (Harris et al., 2017; Kassam et al., 2009). Further, stress can be manipulated in the lab in ways that promote increased arousal, like high-intensity physical exercise, yet are agnostic to aspects of psychological aversion associated with threat (Molins, Serrano, et al., 2021). In this review, I refer to research on stress but only when the type of stress used involved an aversive state.

Historically, threat has proved very difficult to manipulate for the purpose of research. Principally, this is due to the ethical dilemma of placing participants in harm's way (Baddeley, 1972; Idzikowski & Baddeley, 1983; Tashjian et al., 2022). Early psychologists argued that threat and the human drive for self-preservation are primary drivers of psychological function (Sidis, 1922). However, it was not until the 1960s that robust empirical research on the influence of threat on cognition started to emerge. Although creative, this work failed to acknowledge the ethical ramifications of research that *literally* recreated dangerous events experimentally. In one study that typifies this murky ethical past, participants travelling in an aircraft were persuaded by researchers that the plane was going to crash. Mistakes made during the completion of paperwork following this traumatic event led the authors to argue that being in a threatened state negatively impacts behaviour and cognition, as evidenced by the degree to which errors were made on the completed documents (Berkun et al., 1962). Thankfully, by the early 1970s, participant safety was being taken more seriously (Baddeley, 1972).

In more recent years, threat has frequently been operationalised in the laboratory. Here, a threatening stimulus is often stripped of all context, with the aim of deconstructing responses to emotive stimuli into their component parts. A typical example of this approach is to subject participants to valenced images, such as snakes vs flowers, while observing performance on the dependent variable of interest, such as visual search performance (Tipples et al., 2002). This method has produced robust results, for example, that images of fearful faces result in heightened neuronal activation in brain regions associated with processing threat, e.g., the amygdala (Breiter et al., 1996; Sladky et al., 2013). Yet, without manipulating the context in which these images are viewed, there is little evidence that these stimuli promote the subjective experience of being in a threatened state (Grillon & Charney, 2011). These methods fail to recreate the multidimensional responses that are typical of a real-world exposure to threat, such as aversive subjective experience (Lindquist & Barrett, 2008), behavioural adaptation (Grillon & Charney, 2011), or endocrine activation

(Kirschbaum et al., 1993). Thus, decontextualised approaches may lack ecological validity. Moreover, it's increasingly becoming clear that the nature of each stressor must be carefully considered when identifying any effects on behaviour (Giovanniello et al., 2023).

To address these issues of ecological validity (Kihlstrom, 2021) a very limited body of research has looked to collect data from participants in the midst of actual threats, such as in hospital waiting rooms prior to surgery (Frey et al., 2014), or by suggesting to participants that reduced oxygen levels are expected during an experiment (Kamphuis et al., 2011). In one study that illustrates this approach, participants were subjected to a barrage of cognitive tests (including a modified version of the IGT and working memory tasks) pre and post a 100m bungee Jump (Castellà et al., 2020). Such methods are used to promote a threatened state in subjects and facilitate an immersive experience. Yet, they offer very little experimental control. Each participant is subjected to an experience that may vary depending on factors outside of a researcher's influence. This makes it extremely difficult to use such methods and maintain experimental rigour. Further, asking participants to conduct experimental tasks during such events is not without risk. A participant being asked to conduct a cognitive behavioural task before (Idzikowski & Baddeley, 1987) or during a parachute jump, for instance, is potentially unsafe, given the range of safety procedures that cannot be neglected. Therefore, such methods of manipulating threat are rarely used.

Alternatively, a state of threat can be manipulated within the safety of the laboratory to examine threat and its psychological consequences. Like real-world manipulations of threat these methods tend to be incidental in the sense that the threatening stimuli are not directly related to the task at hand, but emerge in parallel or as a distraction to it (Västfjäll et al., 2016). In other words, incidental methods aim to facilitate an environment that fosters a state of threat within an individual in ways not theoretically or contextually associated with the behavioural task being performed. A classic example of this approach is the threat of shock task (Robinson et al., 2013), where participants are subjected to the possibility of experiencing a small (yet distressing) electric shock.

Another set of powerful stress manipulations uses social threats. The Trier Social Stress Test (TSST; Kirschbaum et al., 1993) is another approach used to elicit threat (Allen et al., 2016). The TSST has had various instantiations. However, the overarching idea is to ask participants to present something to observers (e.g., a short presentation or mental arithmetic) for which they are ill-prepared. The observers are often instructed to stare blankly at the participant and provide negative feedback. The threat here is a consequence of failing to meet the standards of some projected ideal of performance. Indeed, these manipulations demonstrate that social threats elicit powerful subjective and physiological responses (Allen et al., 2016; Park et al., 2023).

While these approaches to manipulating threat may succeed in placing participants in a threatened state, the threat is more of a distraction than an integral part of the task at hand. Further, exposure to a stressor is generally given before performance on the task is measured. In this sense, participants are not under threat, but rather recovering from it (Ting et al., 2022). In contrast, real-world situations require individuals to navigate scenarios where the source of the threat is dependent on task performance, e.g., integral (McCall & Laycock, 2024; Peterson et al., 2018). For example, a police officer attempting to disarm a knife-wielding adversary must balance goal-directed behaviour (e.g., making a successful arrest) with self-preservation in response to the threat of being harmed in the process. These factors are not independent. What is required by an individual to perform optimally is not to ignore stressors but to integrate this information into their decisions. Multiple competing goals are a fundamental feature of complexity (Funke, 2012), and so context and relevant environmental stressors should not be overlooked. This is particularly applicable to translational research (Finseth et al., 2022). To accurately model the effect of environmental threat on task performance, direct means are required that manipulate threat in ways that are theoretically or contextually associated with the behavioural task being performed. These direct methods arguably better reflect the way environmental threat can influence task performance as reported in many naturalistic settings (Asutay & Västfjäll, 2024).

Nevertheless, developing experimental paradigms that make threat an integral part of the behavioural task, while maintaining safety and experimental control, is challenging. Despite this difficulty, one promising approach is to use virtual reality (VR; Keren et al., 2013; Kothgassner & Felnhofer, 2020; McCall & Blascovich, 2009; Pan & Hamilton, 2018; Peterson et al., 2018). These technologies allow for the creation of digital virtual environments that, if done well, can promote realistic responses to synthetic situations (Brookes et al., 2024; Slater, 2009). Experimentally, virtual reality can be used to replicate the cognitive demands of complex (Dickerson & Kosko, 1994) and threatening environments (McCall et al., 2015; Pan & Hamilton, 2018). Virtual worlds can be used to model threat explicitly, for example, a fall from a simulated ice glacier (Baker et al., 2020) or ambiguously, for instance, by using subtle variations in light and sound (McCall et al., 2022). Critically, although these environments are fictitious simulations, participants exhibit behavioural, subjective, and physiological responses that map onto those of the real world (Slater, 2009). Further, researchers can embed threatening stimuli, such as the threat of shock, directly into a virtual world, allowing for the observation of a participant's behavioural response to aversive stimuli across contextual manipulations (Suarez-Jimenez et al., 2021). Thus, these methods allow for the development of experimental environments in which actions taken by participants can safely be associated with explicit threats directly (Peterson et al., 2018) while still retaining the ability to manipulate task-specific demands (Binsch et al., 2021).

Immersive virtual reality has also shown promise as a way of modelling dangerous environments that can aid both research and preparation of Individuals to help negotiate hazards such as a fear of heights (Krupić et al., 2021). This approach was recently employed to review the existing emergency spaceflight procedures within a threatening environment. Participants were asked to react to a series of unfolding scenarios that included extinguishing fires within a VR simulation based on the International Space Station (Finseth et al., 2022). Although full size mock-ups of the International Space Station are often used in developing task acquisition skills (Eichler et al., 2006), these simulations are restricted in terms of constructing threatening scenarios. This use of VR demonstrated the utility in recreating a safe yet operationally relevant environment that facilitates the experimentally controlled observations of human cognitive and behavioural task performance during hazardous situations in ways that incidental methods of manipulating threat could not.

Effects of threat on complex decision-making

Previous sections dealt with complex decision-making and environmental threat somewhat independently. Yet it's what happens when these two factors are intertwined experientially, which is the focus of this thesis. The remainder of this chapter will provide a selected narrative review of research that is theoretically relevant to tackling the question of how complex decision-making is affected by being in a threatened state. This diverse body of work includes research examining the effects of threat on complex decision-making itself, as well as threats' effects on the underlying psychological mechanisms that have already been discussed. This literature provides a platform for the development of a range of hypotheses regarding how threat can shape complex decision-making.

Threat is often argued to have a detrimental effect on higher-order cognitive abilities, e.g., (Arnsten, 2009), a perspective that seems to pervade the popular imagination (Arnsten et al., 2012). One core idea here is of threat rigidity, whereby being in a threatened state renders cognition less flexible (Giovanniello et al., 2023; Staw et al., 1981). Intuitive or automatic processes are employed to promote self-preservation at the cost of analytic reasoning (Yu, 2016). However, the threat rigidity hypothesis lacks contextual nuance, and evidence for it often lacks methodological rigour (Mazzei et al., 2025).

Indeed, research on the effect of threat on complex decision-making is mixed and often counterintuitive (Giovanniello et al., 2023; Morgado et al., 2015; Starcke & Brand, 2012; Yu, 2016). This point may be in part due to the way complex decision-making has been operationalised as well as variation across studies in the methods used to manipulate threat (Wake et al., 2020).

For example, Simonovic (2017) observed IGT performance of non-clinical participants across threat versus no-threat conditions. An anticipatory speech task (a modification of the TSST) was used as an incidental manipulation of threat. Threat here was associated with poorer

IGT performance over time, as indexed by a decrease in the selection of advantageous decks in contrast to matched controls. However, other research paradigms have resulted in a different pattern of results. Byrne et al., (2020) manipulated threat using a socially evaluated cold pressor task. Following this manipulation, participants took part in a decision-making under uncertainty task. Individuals in the stress condition demonstrated an improved ability to make choices that led to better outcomes over time, suggesting that acute stress can actually improve performance on complex decision-making tasks.

What explains the discrepancy between these two studies' conclusions? One possibility is the difference between tasks. The decision-making task in Byrne et al., (2020) was arguably less complex, given that reward was ultimately governed by a simple rule (Dörner & Funke, 2017). Unlike the IGT, where the payoff of a given choice is independent of its unpredictability, the more unpredictable choice in Byrne et al., (2020) was always more lucrative in the long run. This is distinct from the IGT, where a participant is required to adapt behaviour with experience to be successful, as initially good choices turn bad as trials progress. Thus, the processes, or cognitive mechanisms, required by the participants to minimise error during the task were different. Because the results in both these studies are limited to final performance scores, we do not know how the various mechanisms involved in wrestling with uncertainty were affected.

As mentioned above, a variety of cognitive mechanisms underlie complex decision-making. Threat has the capacity to affect any of these. For instance, research has long investigated the effect of threat on attention (Baddeley, 1972). Although these effects have been overwhelmingly studied in populations with some form of anxiety disorder (Bar-Haim et al., 2007; Clauss et al., 2022), disruption to attentional processing as a function of threat is likely universal (LeDoux, 1995). One fundamental way that threat may affect complex decision-making is via threat-related attentional bias. Threat-related attentional bias (Bar-Haim et al., 2007) is a term used to describe how distinct attentional systems can be biased towards threatening stimuli (Dennis et al., 2008). This rigid fixation towards the stressor may lead to restricted processing of goal-oriented, non-threatening visual information (Azriel & Bar-Haim, 2020).

Further, this attentional bias is seen across sensory modalities and has been shown to be influenced by prior exposure to threatening experiences. This point was supported by observing the magnitude of the bias using audible stimuli within a population that had previously been exposed to life-threatening events, in this case, psychologically healthy military veterans (Ranes et al., 2017). Here, background details about combat history were recorded before participants were exposed to a series of aversive and neutral sounds. Their attentiveness to this threat was monitored by asking participants to rate their expectancy for aversive stimuli (as a proxy for covert attention) continuously throughout the task. Results from this simple paradigm demonstrated that the intensity of an individual's experience in threatening environments predicted attention to threat. That is, more combat exposure was associated with a greater fixation for the aversive stimuli.

Threat might also affect decision-making via attentional narrowing (Derryberry & Tucker, 1994), driving individuals to only review a subset of available information when making decisions (Lewinsohn & Mano, 1993; Luce et al., 1997). One study demonstrating this effect monitored performance on decision-making tasks developed to assess the degree to which individuals accrue pre-decisional information (Wichary et al., 2016). The decision-making task asked participants to review four job candidates for a prospective role and make choices on who would be the most productive worker. Participants were free to sample the information using a computerised display. This allowed the researchers to code the way information was sampled (the amount of information and the pattern of time spent on different cues). The task was presented to participants following induced emotional stress through exposure to aversive pictures. Results showed that, in contrast to participants in a control condition, threatened individuals preferentially focused on a restricted amount of information (associated with task saliency) being presented. At first glance this may be interpreted as adaptive, yet ultimately this myopic behaviour led to a reduction in pre-decisional information being gathered and resulted in simpler decision strategies being employed (contingent on less evidence). However, this attentional narrowing perspective is contested, with alternative viewpoints (e.g., uncertainty-reduction) suggesting that being in a state of high arousal is likely to increase the drive to seek information to reduce uncertainty and risk (Anselme, 2010), thereby broadening attention (Han et al., 2007).

Like attention, memory has also been shown to be influenced by threat (Thomas & Wulff, 2024). Threat has been shown to disrupt functioning across many different facets of memory. This disruption has been suggested to result in a more vivid recall of threatening events and the false perception that those memories are more accurate (Talarico & Rubin, 2003). Additionally, memory for threatening events may exhibit an over-generalisation effect (new memories mistaken for old) across competing memories, leading to inaccurate categorical representations (Starita et al., 2019). These effects of threat on memory may affect complex decisions, as optimal performance on complex tasks requires a non-biased recollection of environmental dynamics, e.g., problem space representations (Wilson et al., 2014). In other words, given that complex decisions are based on information gathered from experience, the information accrued must be accurate to be useful, because in a complex environment, the exact same stimulus is rarely encountered over repeated experiences (Starita et al., 2019). Accurate memory representation may be critical to ensure an individual can learn from a version of experience that truly reflects the situation as it presents itself.

Threat can also disrupt working memory capacity (Angelidis et al., 2019; Vytal et al., 2012). This effect is demonstrated in work that shows that threat manipulated by the TSST, reduces activity in brain regions known to support working memory (e.g., bilateral dorsolateral prefrontal cortex, superior parietal cortices, and intraparietal sulcus) and hinders working memory performance (van Ast et al., 2016). Further, incentivised action (e.g., following reward) is suggested to improve working memory capacity (Savine et al., 2010). Yet, under threatening conditions, these advantages are lost (Choi et al., 2015). As mentioned previously, working memory appears to be important for complex decision-making and is associated with performance (Bagneux et al., 2013). Threat-induced restrictions on working memory capacity may interfere with one's ability to keep relevant information in mind and use that information when making decisions.

Along similar lines, a variety of evidence suggests that executive functioning more generally is affected by threat. Attentional control theory (ACT; Eysenck et al., 2007) argues that anxiety (trait and transient) affects attention control, i.e., the capacity to flexibly focus and shift attention according to explicit goals (Muris et al., 2008). ACT proposes that this effect influences three cognitive abilities: task inhibition, shifting, and updating. Impact here may disrupt the exploration-exploitation trade-off, as performance and brain activity of "expert" decision makers on a multi-armed bandit task demonstrate that exploration is associated with brain regions that underpin attentional control (Laureiro-Martínez et al., 2015). Further evidence for changes in the balance between exploration and exploitation is suggested by Simonovic et al., (2018), who used an anticipatory speech task (again, a modified version of the TSST) to manipulate threat before the IGT. Amongst other measures, the authors recorded the inspection time of contingencies (e.g., decks) via eye tracking. Results demonstrated that participants in the threat condition performed more rigidly, were slower to disengage from disadvantageous decks (that had previously been rewarded), and examined them for longer. Conversely, those in the nonthreat condition spent longer inspecting advantageous decks and ultimately performed better. However, time spent examining choices is perhaps a better proxy for effort (Bailey et al., 2016) than actual choice behaviour. Other measures of choice preservation, i.e., sticky selection strategies (Senftleben et al., 2021), may be more informative as an index of choice behaviour.

Regardless, threat may lead to a fixation on threatening stimuli at the cost of goal-directed exploration (Eysenck et al., 2007). However, some environmental stressors may be critical to a decision, and a threat-related focus on them might actually support decision-making (Matthews et al., 2000). Independent of the wider theoretical implications of a fixation on environmental stressors to real-world scenarios, disruption in the way individuals negotiate the exploration-exploitation trade-off (i.e., sampling behaviour) may be one mechanism that helps explain how performance in complex situations may be moderated as a function of threat. Especially given that threat impacts the degree to which information is used when negotiating unfavourable choices, one example being that weaker evidence is needed to reach conclusions following social threat manipulations (Globig et al., 2021). Put differently, threat may affect not only the amount of information accrued before being acted on, but also the degree to which it informs behaviour.

Previous work has also suggested ways in which a high-stress environment can influence reward processing following exposure to dangerous events (Nan et al., 2025), as well as manipulations of threat (Berghorst et al., 2013). Foundational to this body of work is the idea that threat impacts loss aversion (Ben Hassen et al., 2023). Loss aversion refers to the tendency to weigh losses more heavily than equivalent gains, leading to risk-averse behaviour (Kahneman, 1979). Conversely, an alternative idea is that under duress, behaviour becomes less (Berghorst et al., 2013) or more fixated on immediate rewards (Ben-Zur & Zeidner, 2009), impacting the accrual of the information required to deal with complexity. Such disruption may promote behaviours like not considering all available options (premature closure), non-systematic scanning of all available options, and not allocating an appropriate time (temporal narrowing) to make an informed judgment (Keinan, 1987). This point is also supported in work using the TSST as an incidental manipulation of threat, which suggests stressed individuals employ simple heuristics when learning that leads to an incomplete awareness of all response options available (Schwabe et al., 2007).

The ability to learn from experience may be particularly challenging in a threatened state. Evidence suggests that anxious individuals show deficits in learning associations between choice behaviour and feedback, which manifest as impairments in the ability to differentiate signal from noise (Huang et al., 2017). Along similar lines, threat may interfere with context-dependent learning. Evidence here can be found in novel work using populations with a clinical diagnosis of anxiety. Participants were asked to navigate a virtual world while brain activity was monitored using non-invasive techniques (Suarez-Jimenez et al., 2021). During the task, individuals were asked to pick virtual flowers. However, the environment was split into two sections (e.g., safe and unsafe). When in the unsafe section of the world, picking a flower was paired with a subtle (yet effective) electric shock. This paradigm aimed to observe the ability to distinguish between safe and dangerous contexts. To achieve this, the participants had to make use of contextual features (mountainous features on the horizon) and environmental predictive cues (shock-associated flowers). Results demonstrated that during crucial periods of the task under threat, anxious individuals displayed weaker engagement of areas implicated in memory processing and emotional regulation (anterior hippocampus and ventromedial prefrontal cortex). More research is required to understand if these changes in neural activity are generalisable to non-pathological individuals, where anxiety is a typical response to a dangerous situation. Yet, the findings do suggest a candidate mechanism accounting for the impact on learning as a function of being threatened state.

Evidence can also be found in non-clinical populations of the impact of threat on cognitive flexibility, as indexed by reversal learning ability (e.g., recognising prediction error and adapting behaviour accordingly). This process may be essential in cases where an individual must resolve conflict and remain goal-oriented following previous exposure to environmental threat (Bublatzky et al., 2017). To observe the effect of threat on reversal learning, researchers typically employ a reward reversal task to investigate how participants adapt to change whilst in a threatened state, as operationalised by a task using the threat of shock (Paret & Bublatzky, 2020). Here, findings suggest that threat promotes rigidity by disrupting reversal learning, leading to delayed switching following negative feedback. However, research in this area is limited and mixed, particularly regarding how the nature of the threat shapes performance, the constraints of the task, and the demands of the context (Giovanniello et al., 2023; Hurtubise & Howland, 2017).

Finally, threat might affect the motivational mechanisms underlying complex decision-making. Individual differences in motivation to engage with uncertainty are suggested to be fixed dispositional traits (Bavolar et al., 2021). However, this position is disputed, with some evidence that aspects of tolerance for uncertainty are malleable following therapy (Laposa et al., 2022; van der Heiden et al., 2012), and indeed changes in context (Djikic et al., 2013; Ladouceur et al., 2000). Research has consistently shown a correlation between anxiety and Intolerance of Uncertainty (McEvoy et al., 2019) and Need for Closure (White, 2022). Much of this research interprets these correlations as different aspects of the same phenomenon. But these correlations may also emerge because being in an anxious or threatened state actually reduces the motivation to engage with complexity (Birrell et al., 2011). Indeed, in many situations, it may be a functional response to a perceived threat.

If a stronger drive to embrace uncertainty boosts complex decision-making performance, it may be helpful to identify factors that promote or limit that motivation. One possibility is that social threat, such as evaluation (Park et al., 2023), might negatively affect the motivation to embrace uncertainty. This point was a central theme in interviews we conducted with military personnel (McCall & Laycock, 2024). Here, participants reported that individuals overly concerned with negative evaluation are less motivated to engage with complexity. If someone fears being judged or criticised for making mistakes, they might avoid taking risks or considering unconventional approaches, leading to restricted exploration and reduced creative problem-solving (Ben-Zur & Zeidner, 2009; Bonetto et al., 2021). In a sense, a fear of failure (Smith, 2020) might inhibit the motivation to engage with complexity (Finn, 2019). However, empirical work supporting a mechanistic account of how sensitivity to evaluation might impact complex decision-making performance is currently lacking.

As this broad overview of the literature suggests, threat likely disrupts complex decision-making via its underlying cognitive processes. Nevertheless, results are mixed, and questions remain regarding when and how the effects emerge across variations in research paradigms used to manipulate threat and index performance, as well as how dispositional individual differences may moderate the impact of threat on performance.

Conclusion

Research suggests that being in a threatened state may change and even hinder decision-making performance by affecting both the cognitive and the motivational factors that underlie complex decision-making. Nevertheless, the findings are mixed, perhaps because of the use of indirect, incidental methods of manipulating threat, coarse indexes of complex decision-making performance, and a lack of research on motivational influences.

The first two empirical chapters of this thesis address some of these gaps by using virtual reality and computational modelling to investigate how complex decision-making is impacted by being in a threatening environment. I developed and validated a virtual world that allowed me to manipulate threat in a way that was integral to the decision-making task. I also went beyond traditional methods of performance to study the effects of threat on cognitive mechanisms that underlie complex decision-making.

The last empirical chapter of this thesis examines the effects of social evaluative threat on motivational factors relevant to complex decision-making. In five studies, I examined the effects of real-world military training on the motivation to engage with complexity. I further investigated the moderating effects of different training strategies on these effects.

Chapter Two: The Effects of Threat on Complex Decision-Making: Evidence from a Virtual Environment

Author contributions

Conceptualisation: Aaron Laycock., Cade McCall. Methodology: Aaron Laycock., Cade McCall., Guy Schofield. Software development: Aaron Laycock., Cade McCall., Guy Schofield. Data acquisition: Aaron Laycock. Data curation: Aaron Laycock. Data analysis: Aaron Laycock., Cade McCall. Visualisation and figures: Aaron Laycock., Cade McCall. Writing, coordination/editing: Aaron Laycock., Cade McCall. Project administration: Aaron Laycock.

Code availability

Analysis code, preregistration, and supplementary materials are available on the OSF repository.

Follow: https://osf.io/jg2qv/?view_only=0d42f9fce5d0466685e205fde92354d2

Abstract

Individuals living and working in dangerous settings (e.g., first responders and military personnel) make complex decisions amidst serious threats. However, controlled studies on decision-making under threat are limited given obvious ethical concerns. Here, we embed a complex decision-making task within a threatening, immersive virtual environment. Based on the Iowa Gambling Task (IGT), a paradigm widely used to study complex decision-making, the task requires participants to make a series of choices to escape a collapsing building. In Study 1 we demonstrate that, as with the traditional IGT, participants learn to make advantageous decisions over time and that their behavioural data can be described by reinforcement-learning based computational models. In Study 2 we created threatening and neutral versions of the environment. In the threat condition, participants performed worse, taking longer to improve from baseline and scoring lower through the final trials. Computational modelling further revealed that participants in the threat condition were more responsive to short term rewards and less likely to perseverate on a given choice. These findings suggest that when threat is integral to decision-making, individuals make more erratic choices and focus on short term gains. They furthermore demonstrate the utility of virtual environments for making threat integral to cognitive tasks.

Introduction

Some of life's most difficult decisions are made within dangerous environments. Indeed, individuals such as first responders or military personnel make critical, complex decisions amidst threats to life and limb (Harman et al., 2019; Lieberman et al., 2005; Penney et al., 2022; Williams, 2010). Nevertheless, many questions remain regarding the effects of threat on complex decision-making (Starcke & Brand, 2016). Research in this area furthermore presents considerable methodological challenges in terms of both manipulating threat and measuring decision-making's underlying processes.

Complex decisions are characterized by uncertainty and the need to balance multiple competing goals (Funke, 2012). This combination presents a significant cognitive challenge that is likely amplified in the presence of threats. In the most basic sense, complex decisions require attention, which itself is shaped by threat in many ways (Azriel & Bar-Haim, 2020). Furthermore, optimal decision-making frequently requires an individual to apply their existing knowledge and to learn about the given situation. These processes are also likely influenced by threat, given its effects on memory (Vytal et al., 2012) and learning (Fanselow, 2018).

More specifically, complex decision-making often involves reinforcement learning; optimal choices depend upon accommodating for feedback from the environment (Yechiam, 2020) and from the outcomes from one's choices (Silver et al., 2021). Threat may disrupt this ability to flexibly respond to changes in rewards and losses over time. For example, research demonstrates that under threat of shock, participants are slower to switch away from disadvantageous choices following negative feedback (Paret & Bublatzky, 2020). Other work suggests that threatened individuals are less likely to explore their options when problem-solving (Keinan, 1987) and instead employ simple heuristics (Schwabe et al., 2007).

Despite these clues that threat negatively affects many of the cognitive processes underlying complex decision-making, research directly examining the effects of threat is limited and the findings are mixed (Starcke & Brand, 2012). This inconsistency may in part be due to fundamental differences in decision-making paradigms. Some work focuses on decisions from description, where participants respond to hypothetical questions (e.g., participants are asked, "If X happened, how would you respond?"), while other work focuses more directly on decisions from experience (e.g., participants are put in X situation and must make an actual decision). These different approaches sometimes yield different conclusions (Gonzalez, 2022). For example, research using decisions from description supports the idea that individuals overweigh low probability events when making risky decisions (Kahneman, 1979), while research using decisions from experience suggests that individuals actually underweigh them in those circumstances (Hertwig et al., 2004).

Research on decisions from experience has frequently used multi-armed bandit tasks (Buelow & Suhr, 2009), tasks in which participants' decisions incur costs and benefits. The Iowa Gambling task (IGT; Bechara et al., 1994), for example, requires individuals to select cards from four separate decks over a sequence of trials. Each selection leads to losses and gains, usually in financial terms. Initially, individuals have no information about the probability of payoff from the decks. Instead, they gain information from sampling the decks across trials. Maximising net reward depends on an individual's ability to learn which decks are most profitable over time and to adapt their choices accordingly. Performance on the this task is ostensibly representative of real-world decision-making and its underlying cognitive processes (Bechara et al., 1994; Brevers et al., 2013). When it comes to the effects of threat on multi-armed bandit tasks, findings are mixed. Some studies find that threat or acute stress decreases overall performance (Wemm & Wulfert, 2017), while others find it can improve aspects of decision-making (Byrne et al., 2019). Still others find no effect at all (Sokol-Hessner et al., 2016).

This diversity in results is perhaps not surprising given the variety of methods used across studies (Starcke & Brand, 2016). Critically, threat is manipulated in different ways, often via a task that is incidental to the decision-making task itself. For example, some studies have examined the effect of threat on IGT performance by manipulating anticipatory stress; participants complete the decision-making task while knowing that afterwards they will be asked to complete a public speaking task (Preston et al., 2007; Simonovic et al., 2017, 2018; Wemm & Wulfert, 2017). Other studies have participants complete a decision-making task after a stressful experience, such as a cold pressor task (Nowacki et al., 2019).

The manipulations in these examples are indeed effective, eliciting both subjective and physiological responses and oftentimes affecting decision-making performance (Preston et al., 2007; Simonovic et al., 2017, 2018; Wemm & Wulfert, 2017). Nevertheless, the relationship between the threat and the decision-making task in these paradigms is incidental; the outcome of decisions is not directly related to the outcome of the threat. While these incidental manipulations of threat can elicit a threat response, they may only tell us about situations where threat is a distractor (i.e., situations in which optimal performance might rely upon ignoring the threat). But they may not tell us about situations when threat is integral to the decision-making task (i.e., when optimal performance determines success in dealing with the threat). This distinction is likely critical in the "real world". For example. threat might act as a distractor for a military medic making decisions about how to deliver care to a patient in the midst of a hostile environment. On the other hand, threat is integral to decision-making when that medic is choosing the safest route out of hostile environment. In this sense, real world threats might play a very different role depending on whether they are incidental or integral to the decision at hand.

Virtual reality provides one means of creating paradigms where threat is integral to decision-making. Virtual environments can elicit subjective and physiological responses via simulations of threats such as being perched precariously on the edge of a cliff or being surrounded by dangerous animals (Baker et al., 2020; Kane et al., 2012; McCall et al., 2015). They can also immerse individuals in ambiguously threatening environments which create a prolonged experience of anxiety and unpredictability (McCall et al., 2022). As such, they potentially allow researchers to create decision-making tasks in which one's decisions are integrally connected to threats in the surrounding environment.

Another potential cause for the diversity of results in the threat and decision-making literature is the disparity in decision-making measures. Multi-armed bandits have been used to examine learning from feedback, loss aversion, risk taking, and other features of complex decision-making that are potentially affected by threat. Even within a given paradigm, approaches differ. With regards to the IGT, measures have evolved over time. Performance was initially quantified by simply calculating the ratio of advantageous over disadvantageous choices (Bechara et al., 1994; Franken & Muris, 2005). More recent research has shifted the focus to the cognitive processes underlying those choices via computational modelling of the behavioural data (Haines et al., 2018; Steingroever, Wetzels, & Wagenmakers, 2013; Sullivan-Toole et al., 2022). Depending on the nature of the model, these methods derive a range of parameters that reflect learning and decision-making mechanisms such as reward sensitivity, loss aversion, and choice perseveration.

Recent work along these lines suggests that threatened individuals may be less responsive to feedback from their choices in the IGT (Ben Hassen et al., 2023). Although the detrimental effect of threat on overall IGT score reported in previous studies (Simonovic et al., 2018; Wemm & Wulfert, 2017) was not replicated in Ben Hassen and colleagues' study (Ben Hassen et al., 2023), their computational modelling suggests that threatened participants were less sensitive to feedback. However, the particular computational model used in their study did not distinguish between loss aversion and sensitivity for reward, which might be critical in our understanding of the general effects of threat (Byrne et al., 2019; Molins, Ayuso, et al., 2021). The novel Outcome-Representation Learning model (ORL) (Haines et al., 2018) addresses this concern. Developed to reflect the cognitive strategies that underpin performance on the IGT, the ORL has separate parameters for loss sensitivity, reward sensitivity, and reward frequency sensitivity. Moreover, the ORL also has parameters for memory and choice perseveration, key features of complex decision-making that the abovementioned research suggests could be influenced by threat.

In the studies presented here, we sought to address open questions regarding the effects of threat on complex decision-making. Specifically, we sought to test if a threat that is integral to the decision-making process would lead to decrements in decision-making performance. We also sought to test if, as the literature suggests, threat would lead to a change in reward or loss sensitivity (Ben Hassen et al., 2023) as well as a reduced tendency to explore the range of options (Keinan, 1987; Paret & Bublatzky, 2020; Schwabe et al., 2007). To do so, we developed a VR-based complex decision-making task structured like the IGT, in which decisions were tied directly to threats in the environment. We furthermore used the ORL computational model to explore the effects of integral threat on cognitive underpinnings of complex decision-making.

Study 1

The aim of Study 1 was to pilot a virtual world for observing complex decision-making in a naturalistic and potentially threatening environment. We embedded a task based on the IGT within a virtual reality environment (the VRIGT). We then tested whether performance would improve over the course of trials, in line with the traditional IGT (Bechara et al., 1994). We also tested whether the ORL computational model used to fit IGT data would also work with VRIGT data.

Method

Participants

Seventy-one (71) participants were recruited to take part in Study 1. One participant was excluded due to medical concerns. Of the full sample (70), 58 participants (34 females, 24 males) completed all measures without technical difficulties. All participants were over 18 years old. Participants were given the opportunity to receive either course credit or a six-pound gift card as reimbursement for their time. Approval for the data collection was granted by the ethics committee of the Psychology Department at the University of York, and all experiments were performed in accordance with relevant guidelines and regulations.

VRIGT

We developed the VRIGT on the gaming development platform Unity (version 2020.3.15f2) using the SteamVR plugin (version 2.7.3). Some materials were imported from the Underwood Project (McCall et al., 2022). The VRIGT was constructed around three scenes: a pre-scene (for practice and instructions), a test-scene (task) and an end-scene (end task).

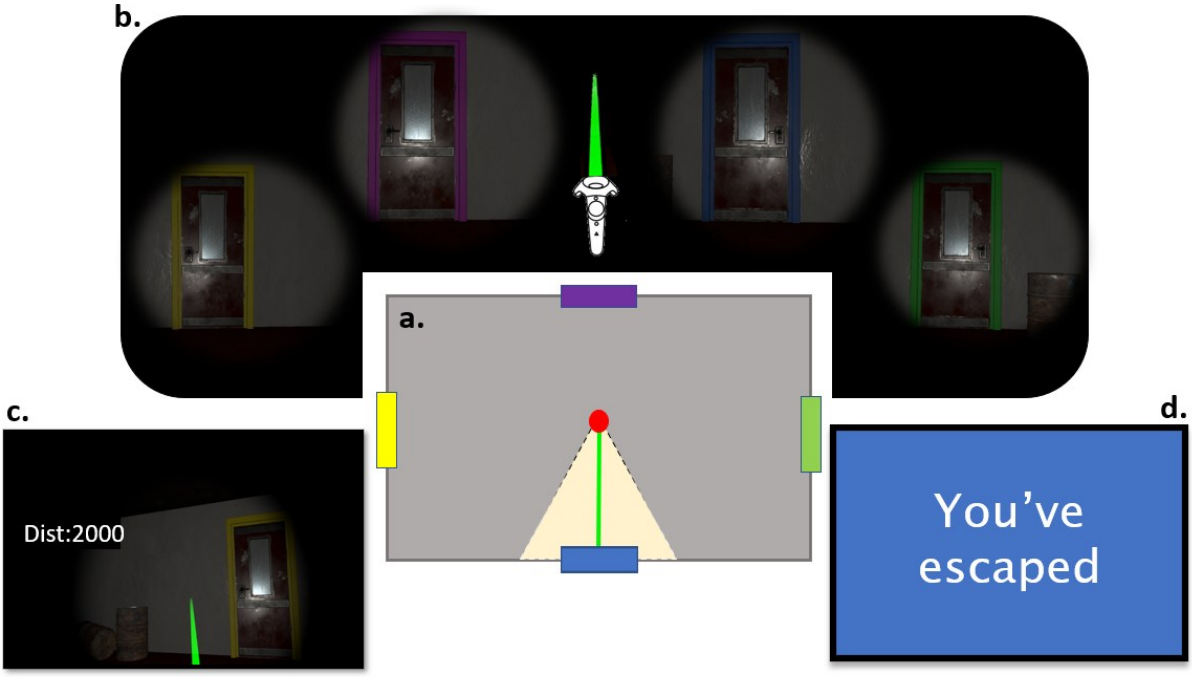
The VRIGT test-scene was based on the Iowa Gambling Task (Bechara et al., 1994). In each trial of the classic IGT, participants select a card from one of 4 decks (A, B, C or D). Each turn of a card results in losses and or rewards. While the amount returned from any given deck changes between trials, some decks offer greater rewards over time. Success on this task relies on the ability to identify the pattern of reward and adapt choice accordingly (Aram et al., 2019).

In Study 1’s version of the VRIGT, participants are told that they are in a building that is slowly collapsing, and their goal is to maximise distance from a danger zone. On each trial they enter a new room. Their task is to choose which door they will use to exit that room. These doors (4 doors = A, B, C and D) replace the decks of the IGT; the financial rewards of the IGT are replaced with meters from the danger zone. In other words, each door selection changes the participant's distance from the danger zone. Participants see their current distance via a display located within their view (see Figure 1c).

Scoring starts at a distance of 2000 meters from the danger zone (as opposed to $2000 facsimile US bills in the traditional IGT) (Bechara et al., 1994). The specifics of each selection, e.g., "You gained 100 meters from the danger zone, but you lost 250 meters", are also played via an audible recording. This recording is paired with a 4-second visual fade to darkness, which alerts the participants that a selection has been made. As with the classic IGT, the VRIGT measures the participant's ability to identify the pattern of rewards and losses (e.g., the door that statistically offers the largest distance increase from the danger zone) and to adapt choice accordingly. In study 1, the VRIGT test-scene consisted of 80 trials (as in Halfmann et al., 2014). We chose the 80 trial version of the IGT instead of the original 100 trial version to avoid any fatigue that might arise from wearing the HMD. These 80 trials were subdivided into four 20-trial blocks for analysis.

Figure 1

Illustrations of the VRIGT Test Scene



Note. a) During the VRIGT test-scene, participants are orientated in the centre of a single room and have 360° freedom to rotate. Colour-coded doors are presented at the centre of each of the surrounding walls. The red dot represents the players' location. b) Door choices are made by placing the raycaster (green laser graphic) over the appropriate door and selecting using the trigger button. c) At all stages during the test-scene, participants can view their current distance from the danger zone (index of reward) on a visible graphic. d) At the end of the test-scene (i.e., after 80 trials), participants are instructed that they have safely escaped the building (end-scene).

Doors in the VRIGT were colour-coded to ensure that participants recognized them from trial to trial (see Figure 1b). Two versions of the VRIGT were created to counterbalance door colour. As with IGT decks A and B, VRIGT doors A and B are low-scoring in the long run. Although selecting these doors offers the highest maximum per-trial reward (100 meters gained from the danger zone), choosing these doors leads to greater losses over time. Doors A and B are identical in sum of losses but differ in frequency of loss. Door A has a series of regular small losses, whereas Door B is associated with rare but large losses. As with decks C and D in the IGT, doors C and D return a low reward initially (50 metres gained), yet the loss over time is less. The frequency of loss received from selecting doors C and D are matched to doors A and B, respectively. The scoring matrix used in the VRIGT is based on the original IGT (as used in Oberdörfer et al., 2021).

Study 1's version of the VRIGT was made to be ambiguously threatening following the approach of McCall et al. (McCall et al., 2022). Following every door selection, the volume of a background audio clip (of a building demolition) increased by a small amount. Lighting was dim, the air appeared dusty (via a particle system), and participants used a torch to illuminate the room.

Hardware

Participants experienced the immersive virtual environment via a VIVE head-mounted display (HMD) with an integrated Dual AMOLED 3.6'' diagonal screen, a resolution of 1080 x 1200 pixels per eye (2160 x 1200 pixels combined), refresh rate of 90 Hz, and a 110 degrees field of view. Participants also used a wireless VIVE controller with dual-stage trigger and integrated HD haptic feedback. For audio, participants wore wireless headphones set to a maximum volume of 80%.

Questionnaires

In the pre-task questionnaire, participants were asked about their age, gender, and video game experience. For exploratory purposes, participants also completed individual difference scales related to intolerance of uncertainty; these are not reported here. In the post-task questionnaire, participants completed a series of questionnaires about their experience in the VRIGT. To rate subjective experience of affect, participants used a slider (ranging from "not at all = 0" to "a great deal = 100") to rate the extent to which the VRIGT was: "frightening", "creepy", "unpredictable", "amusing", "funny", "engaging", "confusing", "disgusting", "interesting", "surprising", "frustrating", "sad", "boring", and "enjoyable" (McCall et al., 2022). Participants also completed user engagement (O'Brien et al., 2018), and tension (Lehne & Koelsch, 2015) scales (see Supplementary Materials for details).

Procedure

After providing informed consent and completing the pre-task questionnaire. Participants were then told what to expect over the course of the task. This included information regarding the VRIGT layout and how to make selections with the controller. Participants were then helped into the HMD and were introduced to the virtual world. We counterbalanced the starting orientation between conditions across four possible orientations. This was done to control for any bias resulting from the spatial start location.

The initial VRIGT pre-scene provided participants with instructions and an opportunity to practice using the controller. Participants were instructed to identify all four doors in the visual display, confirm that the audio level was sufficient, and make a door selection using the appropriate trigger located on the controller. At the end of the pre-scene, participants listened to a set of instructions presented within the virtual world that provided the task narrative, aims, and rules (see Supplementary Materials). Participants then completed the task itself for 80 trials (door selections).

After the trials were completed the task ended and participants were told that they had successfully escaped the building. Participants then completed the post-task questionnaire. Participants were then debriefed and given information about payment or course credit.

Analysis

Performance during the VRIGT. As with the traditional IGT (Bechara et al., 1994), performance was calculated as a difference between advantageous and disadvantageous selections (C + D) – (A + B). Positive scores (> 0) demonstrate an overall trend of selecting doors that minimise net loss. We calculated an overall score (all 80 trials) as well as a score for each 20-trial block, as is the convention (Gansler et al., 2011). In the traditional IGT, participant performance improves over the blocks (Pasion et al., 2017) with improvement expected after about 40 trials in non-clinical populations (Bechara et al., 2005). This increase reflects participants learning to discriminate between advantageous and disadvantageous choices over consecutive trials (Bechara et al., 1994).

Computational modelling. We also tested whether the ORL computational model used to successfully model IGT data in prior research would also fit our VRIGT data. Prior research (Haines et al., 2018) tested the ORL's model performance (e.g., short- and long-term prediction accuracy and parameter recovery) using data from multiple IGT studies with diverse samples. The ORL showed comparable or better performance than other models used to analyse IGT data.

Moreover, the ORL provides parameters that are theoretically important for understanding the influence of threat on complex decision-making. 1) Reward sensitivity (Arew), where higher values represent a greater influence of reward on learning (from 0 to 1); 2) loss sensitivity (Apun), where higher values represent a greater influence of punishment on learning (from 0 to 1); forgetfulness (K), which represents how quickly decision makers forget their past choices (from 0 to 242), with greater values representing shorter retention; reward frequency sensitivity (betaF), in which selections are based on win frequency (from—∞ to + ∞) and positive values demonstrate a preference for options with a high win frequency; 5) choice perseveration (betaP), where higher values reflect a tendency to repeatedly select from the same option (from—∞ to + ∞).

To confirm that the ORL had at least equivalent fit for VRIGT data as other IGT-related computational models, we also tested the Prospect-Learning Valence Delta model (Ahn et al., 2008), the Prospect-Learning Valence Decay model (Ahn et al., 2014), the Value-Plus-Perseverance model (Worthy et al., 2013), and the Outcome Representation Learning model (Haines et al., 2018).

Modelling was run using the "hBayesDM" in R (Ahn et al., 2017), following Haines et al. (Haines et al., 2018). All models were sampled for 4,000 iterations, with the first 1,500 as warmup (i.e., burn-in) across four sampling chains (10,000 posterior samples for each parameter total). Model convergence was judged visually by inspection of trace-plots and assessment using the Gelman-Rubin test (Gelman & Rubin, 1992), $\hat{R}$ values <1.1 suggest adequate model convergence.

Software. All analyses were done in RStudio 4.2.1 [64] (R Core Team, 2022), using R basic or the "lme4" 1.1-26 package (Bates et al., 2015). For all LMMs, p values were calculated using the "lmerTest" package 3.1-3 (Kuznetsova et al., 2017) and the ANOVA function using Satterthwaite's method for F tests. Pairwise post hoc comparisons were calculated using the "emmeans" package 1.6.3 (Lenth, 2023). All pairwise post hoc comparisons used a Tukey correction for p-values. A MVT correction was applied when adjusting for multivariate comparisons. The "bayesplot" package was used visualise posterior predictive checks (Gabry et al., 2019). Finally, univariate outliers were identified using the median absolute deviation (MAD) using the "routliers" package (Delacre & Klein, 2019).

Results

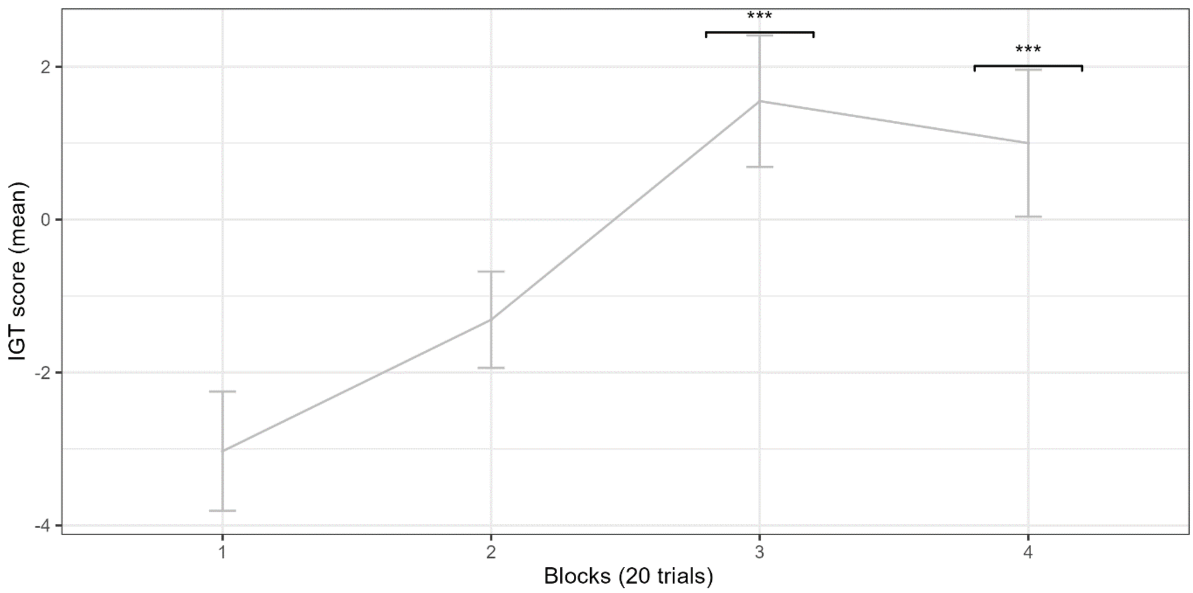
Regarding previous gaming experience, the sample was generally balanced: none = 19 (26.76%), some = 31 (45.07%), and lots = 20 (28.16%). See Supplementary Materials for details of reported user engagement, tension, and subject experience (see Supplementary Figure 3) during the VRIGT.

Performance during the VRIGT

Performance scores (M = -1.79, SD = 16.72, 95% CIs: -6.19, 2.60) were normally distributed. Using the threshold of 3* MAD (+/- median), no outliers were identified. To model the effect of performance over time, we used a linear mixed-effects model (LMM) using block as a fixed factor and performance score as the dependent variable. Block 1 was used as the reference level. Intercepts were allowed to vary as a random factor at the level of the individual. Results demonstrate (see Supplementary Table 2) that the effect of block on performance was significant ($F(3, 171) = 9.26$, $p < .001$). Performance improved as blocks progressed. Performance in blocks 3 (t.ratio (171) = 4.65, p<.001) and 4 (t.ratio (171) = 4.09, p<.001) was significantly better than block 1 (see Figure 2). Post hoc comparisons between sequential blocks show that the performance significantly improved between blocks 2 and 3 (t.ratio (171) = - 2.90, p =. 022, all other p's> .05). As with the traditional IGT (Bechara et al., 2005; Gansler et al., 2011), these results demonstrate that at group level, the ability to discriminate between advantageous and disadvantageous choices improved after approximately the 40^{th} trial.

Figure 2

Performance over time



Note. Error bars represent -/+ standard error. *Indicates differences from baseline with a significant p value *< .050, ** < .01, *** <.001.

Computational modelling

As with prior research on the IGT, each of the computational models we tested had adequate model convergence with all $\hat{R}$ values <1.1. The best model fit (see Supplementary Table 1), however, was with the Outcome Representation Learning model (Haines et al., 2018) (see Supplementary Figures 1 and 2 for posterior distributions of the hyper (group) parameters and posterior predictive checks).

Study 2

Building on the findings of Study 1, we developed threatening and nonthreatening versions of the VRIGT. This allowed us to test if integral threat affects learning and overall performance in a complex decision-making task. We also used the ORL computational model to test for any effects of threat on key parameters in the decision-making process (e.g., reward sensitivity, loss sensitivity, choice perseveration, and etc.). The research questions and analyses for this study were preregistered [As predicted: 130526].

Methods

Participants

One-hundred (100) participants (male = 46, female = 51, non-binary = 3) with an average age of 20 (M = 20.84, SD = 4.23) participated in Study 2. This sample size was chosen based on power analyses using the “smir” package in R (Green & MacLeod, 2016); see

Supplementary Materials for details. Participants reported a range of gaming experience, none = 13%, some = 53%, and a great deal = 34%. Participants were randomly allocated into either the threatening (n=50) or nonthreatening (n=50) experimental conditions. Approval for data collection was granted by the ethics committee of the Psychology Department at the University of York and all experiments were performed in accordance with relevant guidelines and regulations.

Materials

The materials used in Study 2 were similar to those of Study 1 with a few changes. Study 2 used two versions of the VRIGT, a threatening version and a nonthreatening version. In the threatening version, participants were told that their task was to escape a collapsing building. In the nonthreatening version, participants were simply told that their task was to exit an office building; there was no mention of any threats (transcripts available in the Supplementary Materials). We used these two versions in a between-subjects design which we chose to avoid practice effects from repeating the IGT (Ernst et al., 2003).

In terms of the virtual world's content, both VRIGT versions used in Study 2 included a prime-scene before the task (150s). This prime-scene differed between conditions (see Figure 3). In the threat condition, participants entered an elevator that gradually descended 8 floors. The elevator had transparent doors and on each floor the doors opened. Over the course of the descent, participants encountered threats that gradually increased in intensity. First, an audible warning instructed participants not to enter the lift (although they had no control in doing so). This was followed by an exploding light fitting, fire in the hallway opposite the elevator, cracking lift windows, further encroaching fire, and finally, smoke which filled the lift compartment (video in Supplementary Materials). The nonthreatening condition included a different prime-scene (also 150s); participants still rode the elevator for 8 floors, but there were no threatening stimuli, they simply travelled past floors of a mundane office building.

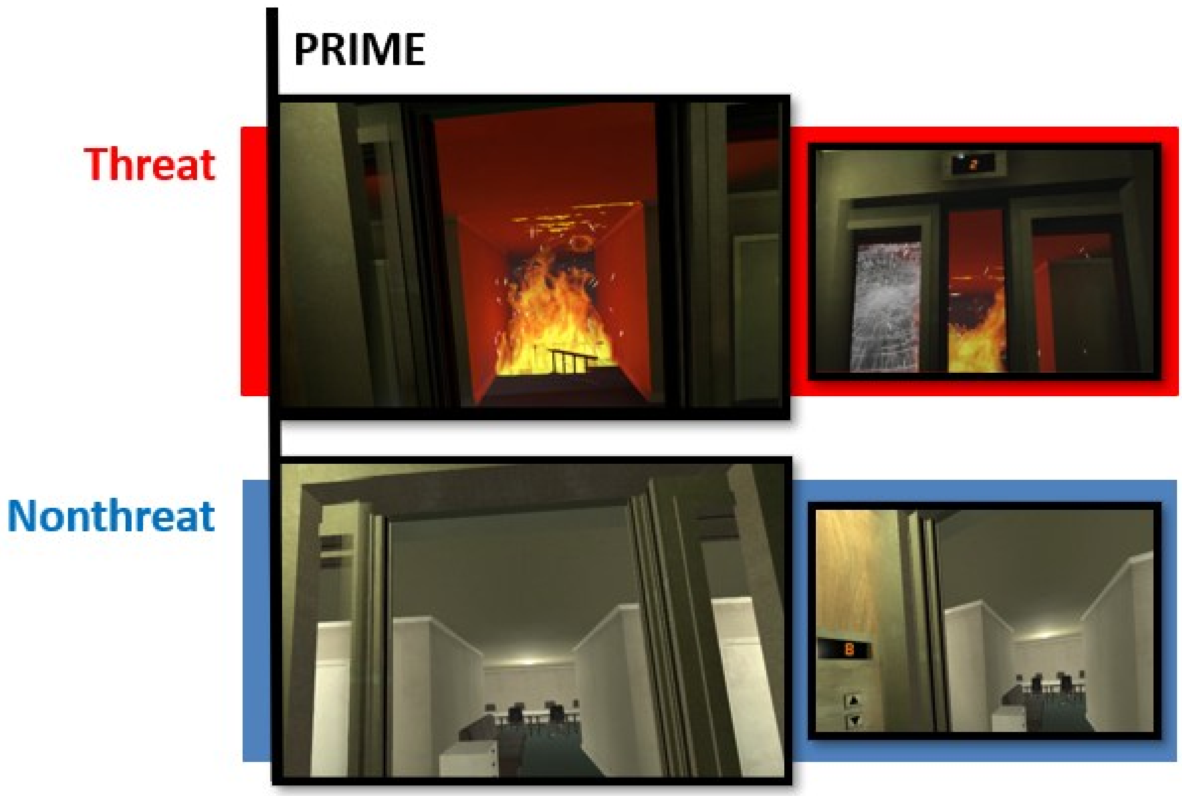
We also changed the content of the task rooms slightly in Study 2. To give a sense of moving through a series of different rooms (as stated in the narrative), the appearance of the task room changed between trials for both conditions. These changes included subtle adaptations to the walls, floor, and lighting. In between trials, the time of the fade was extended to 4.8 seconds. There were also minor differences between conditions in the task scenes. Capitalizing on the threatening effects of darkness (Grillon et al., 1997), lighting in the threat condition was the same as Study 1 (i.e., dim with a torch), while rooms were more brightly lit in the nonthreatening condition. The threat condition also included ambient red lighting to signal danger (Elliot & Maier, 2007) and some aversive sound effects. These effects were taken from sounds already experienced by the participants during the prime-scene but were not paired with any visual effects. Because participants did not report fatigue in the 80 trial version from Study 1, we increased the number of trials during the task to 100, meaning that the test-scene did not end until these trials had been completed.

To measure differences in physiological arousal between conditions, skin conductance level (SCL) and heart rate (via ECG) were recorded using AcqKnowledge 5.0 software (Biopac Systems Inc., Santa Barbara, CA) and the Biopac MP160 acquisition system. SCL was recorded using a wireless Biopac BioNomadix amplifier (BN-PPGED). The electrodes were attached to the middle phalanges of the left middle and index fingers. Heart rate was recorded using a wireless Biopac BioNomadix ECG (BN-RSPEC) amplifier with a three-lead set. Electrodes were placed on the sternal end of the right clavicle, left mid-clavicle, and lower left rib cage. These physiological data were collected during a 5-minute baseline before the task, during the prime scene, and during the task itself. Heart rate data were averaged over these blocks. For skin conductance data, we took the average skin conductance minus baseline over the given block. Due to a technical fault during recording, we excluded two participants' SCL data from this analysis.

To test for differences in timing between conditions, we also recorded trial time length.

Figure 3

Screenshots of Study 2 prime-scenes for the two conditions



Results

Manipulation checks

To test differences in subjective experience between the two conditions, we tested for group differences in the post-task questionnaire (see Supplementary Figures 4 and 5). For the affect variables, we used a LMM to predict rating, with fixed factors for affective category (e.g., "unpredictable"), condition, and their interaction. Intercepts were allowed to vary as a random factor at the level of the individual. A significant effect of condition ($F(1, 98) = 23.31$, $p <. 001$), and its interaction with affective category ($F(13, 1274) = 7.44$, $p <. 001$) was found on rating (see Supplementary Table 3). Post hoc comparisons of affective terms suggested that between conditions, only ratings that the VRIGT was "frightening" (t.ratio (1179) = – 7.46, $p <. 001$), "creepy" (t.ratio (1179) = – 6.51, $p <. 001$), and "surprising" (t.ratio (1179) = – 3.42, $p =. 009$, all other p's> .05) were significantly different, with those in the threat condition reporting higher ratings in all instances (see Supplementary Table 4). The MVT correction was applied to adjust for multivariate comparisons.

We compared mean heart rate between conditions for the baseline, prime-scene, and task-scene. At baseline there was no significant difference in heart rate between the threatening ($M= 89.77$, $SD = 15.62$) and nonthreatening ($M= 86.94$, $SD= 12.20$) conditions, $t (98) = -1.01$, $p =.314$. During the prime-scene, average heart rate was significantly higher in the threatening ($M= 98.85$, $SD = 17.08$) versus nonthreatening ($M= 92.76$, $SD= 13.28$) conditions, $t (98) = -1.99$, $p =.049$. During the task-scenes, heart rate was also higher in the threatening ($M= 101.29$, $SD = 17.04$) versus the nonthreatening ($M= 94.93$, $SD= 11.95$) conditions, $t (98) = -2.16$, $p =.033$.

We found no effect of condition on average SCL at baseline. There was also no difference between conditions in the baseline corrected averages for the prime or task scenes (all p's> .050).

The average trial time taken (in seconds) on each trial during the task-scene was not significantly different between the threatening ($M= 5.85$, $SD = 2.02$) and nonthreatening ($M= 5.47$, $SD= 1.33$) conditions, $t (98) = -1.10$, $p = .273$.

Performance

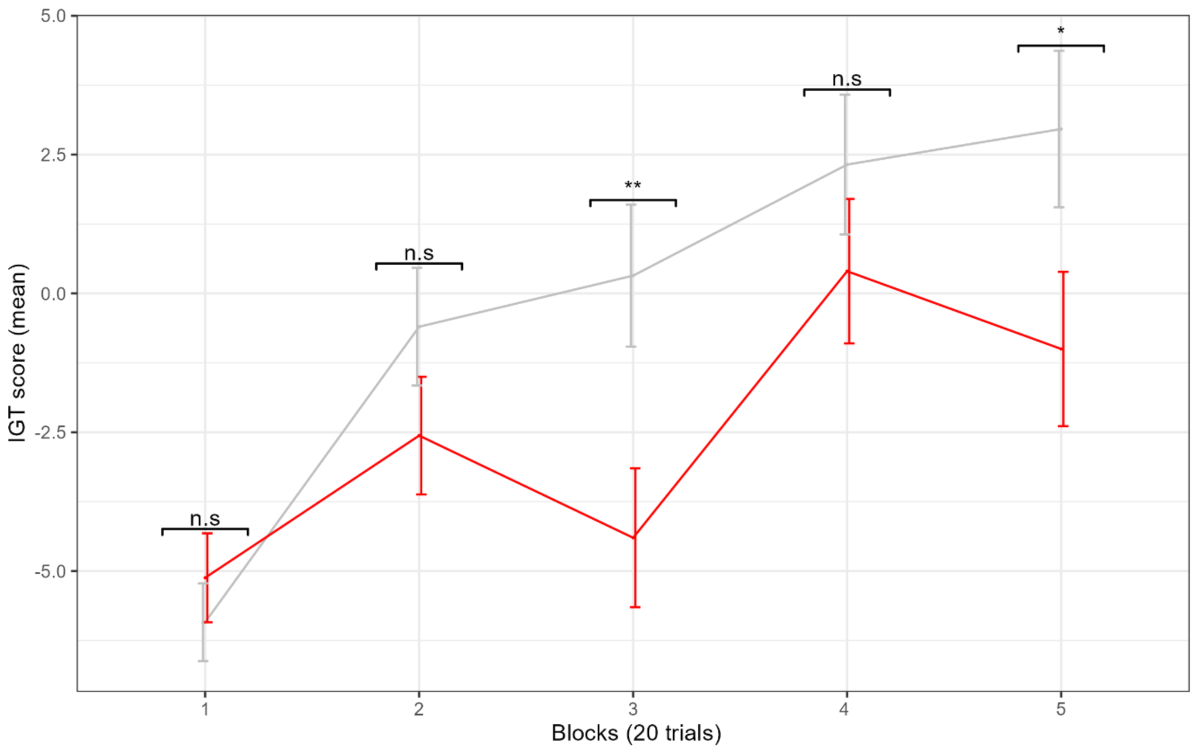
Performance scores were normally distributed, and no outliers were identified. A t test (two-tailed) comparing performance between threatening ($M= -13.04$, $SD = 29.13$) and nonthreatening ($M= -.92$, $SD= 22.22$) conditions showed a significant difference, $t (98) = -2.15$, $p =.034$. Participants performed worse in the threat condition.

To assess performance over time, we ran an LMM with block, condition, and their interaction as fixed factors. Block 1 was used as the reference level. Intercepts were allowed to vary as a random factor at the level of the individual. Results demonstrate (see Supplementary Table

5) that the effect of condition ($F(1, 98) = 4.62$, $p = .034$), and its interaction with block ($F(4, 392) = 2.47$, $p = .045$) on performance was significant. A significant block by condition interaction on performance (see Figure 4) was seen at block 3 (t.ratio (392) = − 2.87, p =. 004) and block 5 (t.ratio (392) = − 2.47, p =. 014, all other p's> .05). This suggests that the differences in performance between the threatening and nonthreatening conditions first emerged at the point when non-clinical individuals in traditional IGT studies begin to migrate toward the more advantageous decks, i.e., around the 50th trial (Bechara et al., 2005) and the point at which meaningful individual differences tend to emerge (Gansler et al., 2011). While participants in the threat condition seem to close the gap by block 4, their learning is apparently limited. As a consequence, participants in the nonthreatening condition have a higher score in the final block when the benefits of learning peak (Bechara et al., 2005).

Figure 4

Performance over time between conditions



Note. Conditions (red= threat, grey = nonthreat). Error bars represent −/+ standard error. *Indicates differences between conditions with a significant p value *< .050, ** < .01, *** <.001.

Breaking this down for participants in the nonthreatening condition only, we ran an LMM to assess the effect of block as a fixed factor on performance, with intercepts being allowed to vary as a random factor at the level of the individual. Results demonstrated a significant effect of block on performance ($F(4,196) = 12.87$, $p < .001$). Blocks two (t.ratio (196) =

3.84, p <. 001), three (t.ratio (196) = 4.50, p <. 001), four (t.ratio (196) = 5.94, p <. 001), and five (t.ratio (196) = 6.40, p <. 001) were significantly higher than block one (see Supplementary Figure 6 and Table 6). Post hoc comparisons between successive blocks confirmed that performance significantly improved between blocks 1 and 2 (t.ratio (196) = – 3.84, p =. 002, all other p's> .05).

A different pattern of results emerged in the threatening condition. Here again we used an LMM to assess the effect of block as a fixed factor on performance, with intercepts being allowed to vary as a random factor at the level of the individual. Results demonstrated a significant effect of block on performance (F(4, 196) = 5.37, p <. 001). Yet, only blocks four (t.ratio (196) = 3.87, p <. 001) and five (t.ratio (196) = 3.09, p =. 002) were significantly higher than block one (see Supplementary Figure 7 and Table 7). Post hoc comparisons between blocks confirmed that performance only significantly improved between blocks 3 and 4 (t.ratio (196) = – 3.33, p =. 009, all other p's> .05).

Computational model

We applied the ORL, to the data from Study 2 for a more fine-grained analysis of behaviour across conditions (as in (Haines et al., 2018). All $\hat{R}$ values <1.1 suggest adequate model convergence (see Supplementary Figure 9 and Figure 11 for posterior predictive checks). Here, the posterior distributions from components of the ORL model were compared between the threat and nonthreat conditions of the VRIGT. This approach allowed for comparison between groups on model components, given that comparisons that do not include 0 (no change) within the 95% highest density interval (HDI) can be considered strong evidence in support of a difference (Kruschke, 2018).

This approach revealed differences in reward sensitivity and choice perseveration between participants in the threat and nonthreat conditions of the VRIGT (see Table 1). Participants in the threat condition (on average) had a greater tendency to update expectations after experiencing reward. They also switched between options (doors) more frequently than participants in the nonthreat condition. We found no robust support for differences between conditions in terms of loss sensitivity, frequency sensitivity, or forgetfulness.

Table 1

ORL group level comparisons

	Threat	Nonthreat	95% HDI of comparison
Reward sensitivity	**.28 (.05)**	**.08 (.01)**	**.107, .299**
Loss sensitivity	.05 (.01)	.03 (.01)	-.018, .041
Forgetfulness	.87 (.34)	1.51 (.26)	-1.487, .175
Frequency sensitivity	1.24 (.21)	1.22 (.18)	-.540, .578
Choice perseveration	**.27 (.66)**	**1.97 (.57)**	**-3.481, -.046**

Note. Means and SDs (in brackets) of the ORL model components. 95% HDI based on the posterior distributions for the mean differences between groups. Bold indicates strong evidence of a difference between conditions.

Discussion

Here we used a virtual reality-based version of the Iowa Gambling Task (IGT) to test the effects of threat on complex decision-making. We made threat integral to decision outcomes and tested its effects using both traditional measures of performance and computational modelling. In doing so, we show that threat reduces decision-making performance, likely by increasing focus on short-term rewards and decreasing meaningful choice perseveration.

In Study 1, we piloted our virtual reality version of the IGT (VRIGT). As with the traditional IGT, performance improved over the course of the task as participants learned to make more optimal choices. We were furthermore able to fit the data using computational models that have been useful for modelling IGT data in prior research (Haines et al., 2018). Together, these findings are in line with previous work which shows that playing the IGT in a VR environment, when compared to a computerised desktop display, does not disrupt performance (Oberdörfer et al., 2021). Moreover, participants' subjective reports from Study 1 also suggest that the VRIGT provided a challenging and complex task.

In Study 2, we tested the effects of threat on complex decision-making performance and learning. We created two versions of the VRIGT, one threatening and one nonthreatening. The biggest differences between these conditions emerged before the task itself (i.e., in the prime-task scene), during which participants in the threat condition were immersed in a virtual building that was slowly collapsing. Participants in the nonthreatening condition were in a similar building but without the threatening details. This manipulation appears to have been effective, as participants rated the threatening environment as more frightening than the nonthreatening environment. Participants in the threatening condition also exhibited higher average heart rates throughout the experience (although we found no difference in skin conductance).

More importantly, the manipulation of threat negatively affected task performance overall and over time, with participants requiring more trials to significantly improve performance from baseline and scoring lower in the final block when compared with those in the nonthreatening condition. In general, these findings replicate previous work that demonstrated a negative impact of incidental threat on IGT performance and learning (Simonovic et al., 2017, 2018; Wemm & Wulfert, 2017). Our data add to this literature by showing these detrimental effects when threat is integral (and not incidental) to the decision-making task. Whereas prior work created situations where threat emerged as part of a separate task, divorced from the decision-making process itself, decisions in the VRIGT have direct relevance to threat imminence (i.e., on how far the participant is from danger).

To more closely investigate differences between conditions, we used the ORL computational model. These findings reveal that individuals in the threat condition were more driven by reward and displayed less choice perseveration (i.e., a greater tendency to switch between choices) over the course of the task. Indeed, these tendencies may be at the root of the poorer performance in the threat condition. Prior research has connected greater reward sensitivity with poor performance on the IGT (Bechara et al., 2000; Zeif & Yechiam, 2020). The aim during the IGT (as is often the case in real-world situations) is to maximise net reward over time, which sometimes requires individuals to sacrifice immediate gain (e.g., the magnitude of a single trial) in favour of longer-term goals. Moreover, these findings support suggestions by Wemm & Wulfert (2017) that acute stress enhances the salience of reward-associated behaviours. Here, a bias to only focus on the metaphorical “carrot” and neglect the “stick” leads to sub-optimal performance.

These data are also roughly in line with Ben Hassen et al. (Ben Hassen et al., 2023), who suggest that threat reduces loss aversion in the IGT, as evidenced by a lower index of the loss aversion parameter of the VPP computational model following threat manipulations. However, as noted by the authors, the VPP models loss aversion and sensitivity for reward as a single parameter (e.g., losses relative to gains). Therefore, the reported changes in this parameter in their study could be interpreted either as higher sensitivity to gains or as lower sensitivity to losses. We provide a degree of clarity here. We used the ORL model, which separates sensitivity for reward and loss (Haines et al., 2018), and found that threat did not disrupt loss sensitivity, but did affect reward sensitivity.

The way individuals sample information via choice switching also tells us something about how they deal with complex decisions (Yechiam, 2020). Some previous research suggests that when making decisions from experience, individuals in threat-related states sample more information before making a choice (Frey et al., 2014). Yet these findings are at clear odds with suggestions elsewhere in the literature that incidental threat promotes premature closure, whereby an individual perseverates on a given choice before sufficiently exploring their options (Dubois & Hauser, 2022; Keinan, 1987). Here, using a task in which threat is integral to decision-making, we find that threat increased individuals' tendency to switch between choices. That is, rather than early choice perseveration and limited exploration, threatened participants continued sampling from the different options. However, this greater exploration did not lead to improved performance. Instead, participants in the threatening condition performed worse than controls. With this in mind, the threatened group's switching between options may have been more impulsive than strategic (Ben Hassen et al., 2023). Future research could more directly test this claim with paradigms that evaluate impulsivity (e.g., Simon et al., 2021). Moreover, future research could also directly test the effects of integral versus incidental threat on complex decision-making. Factors such as choice perseveration may be different when threat is an incidental distractor versus when an individual's decisions have direct implications for the level of threat.

The findings presented here demonstrate the utility of combining existing decision-making paradigms with VR's ability to effectively make threat an integral part of a task. Here, we used the IGT as our starting point. Future work could take a similar approach with other multi-armed bandit paradigms. Further work could also look at variability in the effects of threat on performance based on individual differences. Indeed, decision-making is shaped by many factors including other forms of affect (Asutay & Västfjäll, 2022; Västfjäll et al., 2016), working memory (Bagneux et al., 2013), and intolerance for uncertainty (Luhmann et al., 2011).

Regardless, the current findings suggest that when threat is an integral part of complex decision-making, it can disrupt learning from feedback, focus attention on short term rewards, and reduce perseveration on adaptive choices. This pattern of effects may have real world implications for individuals living and working in threatening environments. Threat biases the decision-making process; knowing the nature of those biases might help people in dangerous settings keep themselves and others safe.

Chapter Three: Reversal Learning Under Threat

Author contributions

Conceptualisation: Aaron Laycock., Cade McCall., MaryAnn Noonan. Methodology: Aaron Laycock., Cade McCall., Guy Schofield., MaryAnn Noonan. Software development: Aaron Laycock., Cade McCall., Guy Schofield. Data acquisition: Aaron Laycock. Data curation: Aaron Laycock. Data analysis: Aaron Laycock., Cade McCall. Visualisation and figures: Aaron Laycock., Cade McCall. Writing, coordination/editing: Aaron Laycock., Cade McCall. Project administration: Aaron Laycock.

Code availability

Analysis code, preregistration, and supplementary materials are available on the OSF repository.

Follow: https://osf.io/a7gpj/overview?view_only=376bc53183eb4f798f6a3915f235df9d

Abstract

Complex decision-making is difficult, and this difficulty only intensifies when those decisions must be made under threat. The ability to adapt behaviour in response to feedback is an important process underlying complex decision-making performance. However, understanding how threat shapes flexibility and how it impacts performance, is extremely challenging given that paradigms often fail to distinguish between different types of threat, and indeed forms of uncertainty. We embedded a three-arm probabilistic reversal learning task into an immersive virtual environment to explore the effects of integral manipulations of threat on cognitive flexibility. This allowed us to manipulate threat in such a way that it was directly related to the task at hand, not incidental which is often the case experimentally. The probabilistic reversal learning task allowed us to isolate two specific types of uncertainty: volatility and stochastic variability. One hundred participants completed the task under both threatening and nonthreatening conditions. We also used a computational model to estimate learning rates and inverse temperature between conditions. While threat did not impair reversal learning performance, it was associated with higher learning rates and inverse temperature estimates, indicating increased responsiveness to feedback and a greater tendency to exploit current beliefs. This challenges the threat rigidity perspective, that proposes threat drives inflexibility and reduces responsiveness to feedback. This shift in process may support rapid action under volatile, well-defined constraints but may limit adaptive action when situations are complex.

Introduction

Optimal decision-making in dynamic environments requires cognitive flexibility. Imagine being trapped inside a building that has begun to collapse. Heading towards the closest exit may initially be a sensible choice. But the merit of that choice might change if one encounters an impassable stairway. In this situation, the individual must update expectations based on feedback from the environment. Moreover, they must balance the trade-off between exploiting previous choices and searching for novel solutions. In this sense, the processes of decision-making and learning are fundamentally intertwined (Levy & Schiller, 2021). Prior research suggests that threat disrupts both learning and decision-making in complex, volatile environments (Laycock et al., 2024). One way in which it might do so is by limiting cognitive flexibility, specifically the learning processes that allow an individual to flexibly respond to changes in the environment (Fellows & Farah, 2005; Krems, 1995).

Reversal learning

Reversal learning, the ability to adapt to changing associations between choice and reward, is one important form of cognitive flexibility. It is critical to success in dealing with an uncertain and dynamic natural world (Bland & Schaefer, 2012). Deficits in reversal learning are associated with maladaptive functioning, with impairment typical across a broad range of clinical conditions (Perandrés-Gómez et al., 2021; Uddin, 2021).

Lab-based studies using animals have a long history investigating reversal learning (Butter, 1969), and research using human subjects is increasingly common (Izquierdo et al., 2017). During a traditional probabilistic reversal learning task, subjects are asked to choose between two (Weiss et al., 2021) or more (Wittmann et al., 2023) stimuli, and the subject receives immediate feedback that the choice is either rewarded (or not). The stimuli are presented on a desktop visual display (Highgate & Schenk, 2021) and are often abstract such that they don't hold any pre-existing meaning or associations to participants (Weiss et al., 2021). A choice represents a single trial. In many human studies (Highgate & Schenk, 2021), reward is jittered probabilistically.

For example, in a two-option version of the task, the optimal option will be rewarded 70% of the time, while the suboptimal option will be rewarded 30% of the time. This is repeated over many trials. However, periodically, the associations between choices and rewards are reversed without participants being explicitly informed. Thus, to perform optimally, participants must negotiate stochastic variability (i.e., jitter), and volatility (i.e., reversals over time) during the task (Levy & Schiller, 2021; Simoens et al., 2024). The ability to identify this change and adapt choices to maximise reward is a feature of cognitive flexibility.

Historically, measures of reversal learning in these paradigms include the number of accurate responses (and mistakes) participants make after each reversal (Weiss et al., 2021), the number of errors made before reaching a learning threshold (Hampton et al., 2007; Murphy et al., 2003), the use of simple heuristics such as win-stay/lose-shift rates (den Ouden et al., 2013), and the differences in sampling behaviour such as perseveration across trials (Izquierdo et al., 2017). However, increasingly computational methods are being employed to index the cognitive processes that underlie decision-making. Computational modelling can estimate parameters such as learning rate and inverse temperature (Gläscher et al., 2009; Wittmann et al., 2023), and can therefore be used to identify the effects of experimental manipulations on these specific cognitive processes.

Learning rate and inverse temperature parameters (Rescorla, 1972; Siegel & Allan, 1996) describe how individuals update their expectations based on prediction error. Ideally, these parameters should adapt to enhance the ability to anticipate future events in a given environment, as this is the fundamental aim of the learning process (Behrens et al., 2007). The learning rate describes how much weight is given to new information, reflecting how quickly expectations adjust. Inverse temperature describes the consistency of decision-making. Higher values reflect choices more reliably following learned rewards, while lower values reflect greater variability and exploration. Together, these parameters isolate key aspects of the learning processes that underlie decision-making.

To illustrate, recall our scenario where someone is trapped inside a collapsing building. Learning rate indexes how quickly they incorporate new information, such as the appearance of hazards on a given route, into their decision-making process. Individuals with a high learning rate will be more influenced by the new information about hazards than those with a low learning rate. This difference reflects how their expectations for the different routes (i.e., the learned values for those different options) are updated with experience. Inverse temperature, on the other hand, describes the extent to which an individual's choices are based on those learned values. Individuals with a high inverse temperature stick more rigidly to the path that has been most rewarding in the past, while those with a lower inverse temperature are more likely to explore alternative paths.

Reversal learning under threat

Of course, being trapped inside a collapsing building is not only a computationally challenging situation to negotiate, but also profoundly threatening. Reports of individuals with lived experience dealing with dynamic choices under threat suggest that threat impacts decision-making performance (Harris et al., 2017; McCall & Laycock, 2024). This is perhaps not surprising, as threat impairs brain regions that support cognitive flexibility (Arnsten, 2009), and likely impact aspects of cognition that underly adaptive functioning, and leads to inflexibility (Giovanniello et al., 2023). Threat interference (i.e., fixation on aversive stimuli) can impair short-term retention of information (Angelidis et al., 2019), bias recall towards emotive stimuli, and narrow attentional scope (Starita et al., 2019; Thomas & Wulff, 2024). This excessive fixation on a stressor may lead to reduced informational search and restricted cognitive flexibility (Azriel & Bar-Haim, 2020; Wichary et al., 2016).

In general, these effects across a wide range of cognitive domains align with the threat-rigidity hypothesis, which posits that under threat, individuals constrict information processing and default to rigid behaviours that undermine adaptability (Kamphuis et al., 2011; Pahng & Kang, 2023; Staw et al., 1981). In essence, behaviour is less sensitive to feedback (Bogdan & Pizzagalli, 2006) and more dependent on intuition-based computation (Schwabe & Wolf, 2011; Yu, 2016). This would suggest that under threat we would expect worse performance on reversal learning tasks, given that success requires flexible updating of choice-reward associations across changing contingencies. Breaking this down, rigidity would result from individuals being slower to learn and less governed by feedback.

Yet, things aren't so simple to interpret through the lens of threat rigidity (Mazzei et al., 2025). The effects of threat on higher order cognition are often more nebulous. For example, using the Iowa Gambling Task to operationalise complex decision-making performance, we found that threat negatively impacted learning (Laycock et al., 2024). In this study, participants in the threat condition demonstrated increased sampling, despite worse performance when compared to those in a nonthreat condition. Given the worse performance, this was interpreted as erratic choice behaviour. Participants under threat appeared to be sampling more information, but not using that information to improve performance.

One explanation for these and other findings is that threat impairs reversal learning, which in turn impacts performance on complex tasks that require individuals to adapt choice behaviour in response to feedback. This would result in a disconnect between the amount of information acquired (i.e., increased sampling) and performance over time. Indeed, evidence suggests that an underlying cognitive flexibility deficit may play a significant role in impairing performance on complex decision-making tasks (Fellows & Farah, 2005; Krems, 1995). However, complex decision-making tasks (e.g., the Iowa Gambling Task) are not always suited to directly measure cognitive flexibility. Rather than presenting participants with discrete reversals that allow researchers to index adaptability to changes in feedback at specific points, they expose individuals to a nuanced, non-binary reward/loss schedule and focus on gradual behavioural shifts. This makes it challenging to attribute behaviour to specific events, since behaviour is shaped by many overlapping experiences.

Although research on how threat influences reversal learning remains limited and is primarily drawn from animal studies, findings suggest that threat can disrupt the ability to modulate behaviour as a function of feedback (Bogdan et al., 2011) and have a maladaptive effect on cognitive flexibility (Giovanniello et al., 2023). However, effects can vary depending on the type of stressor and task constraints (Giovanniello et al., 2023; Hurtubise & Howland, 2017). For example, in a study by Paret & Bublatzky (2020), participants performed a classic two-armed reversal learning task. Participants were instructed that they might receive (small) electric shocks when a specific background colour was present at various stages of the task. Thus, the task alternated between the contextual setting of threat and safety. Findings demonstrated that this manipulation of threat resulted in diminished reversal learning performance, as indexed by an increase in the number of non-rewarded stimuli chosen after reversal. In line with the threat-rigidity hypothesis, these findings were interpreted as enhanced habitual behaviour and failure to update behaviour based on feedback under threat.

The relationship between threat and decision-making

Research on the influence of threat on decision-making has taken a variety of different approaches. One key distinction between these approaches is the relationship between the threat and the decision-making task. In some studies, the threat manipulation occurs before the decision-making task (Wemm & Wulfert, 2017). In other studies, the threat is incidental to the decision-making task, occurring in parallel to it (Paret & Bublatzky, 2020; Simonovic et al., 2018). In a small set of studies, threat is integral to the decision-making process in the sense that the outcome of the decisions directly impacts the level of threat (Laycock et al., 2024). These different approaches may produce different results for good reason.

For example, when fleeing our metaphorical building, threat is integral to one’s decisions about the best escape route. Taking the wrong route directly impacts the level of threat. In this sense, threat is not merely a distraction from the task of escaping the building, it provides critical information. Success requires an individual to act on this information. Ignoring feedback from the environment (e.g., a loud bang or smoke bellowing from underneath a door) likely impacts performance. This is not the case with situations where threat is incidental to decision-making. It is analogous to trying to do a different task (e.g., attending to the needs of an injured person) in the midst of a collapsing building. Threat here is a distractor and only draws attention away from the task at hand.

With this in mind, the way threat is experimentally manipulated may be important for interpreting the results of a given study (Asutay & Västfjäll, 2024; Laycock et al., 2024; Västfjäll et al., 2016). As discussed, threat is often manipulated before the task. This time disparity makes it unclear if the effects of the threat manipulation are due to being in a threatened state or recovering from one (Ting et al., 2022). In paradigms where threat is incidental to the task (Paret & Bublatzky, 2020), the effects of manipulation on behaviour may tell us about the distracting nature of threat. But unless manipulation of threat is integrally tied to decision-making itself, then the task may fail to tell us how individuals make decisions directly related to threat.

The current study

In a preregistered study, we embedded a reversal learning task into an immersive virtual environment (the VRCF) to explore the effects of integral manipulations of threat on cognitive flexibility. The VRCF places participants in a virtual building. During the non-threating condition, the participant's task was to simply make their way outside of the building. During the threat condition, this was different; participants were informed that the building was slowly collapsing and that their task was to escape. Visual and audio effects were added to create the threatening experience. This world was based on a virtual world that was previously been shown to elicit feelings of creepiness and fear and increase heart rate, as compared to a nonthreatening condition (Laycock et al., 2024).

During the VRCF, participants completed a three-arm probabilistic reversal learning task. This was inspired by traditional reversal learning tasks, where the objective is to select the options associated with reward, but where the pairing of choice and reward changes over time (i.e., at reversals). The three-armed bandit task has previously been used in both animal (Noonan et al., 2010) and human populations (Wittmann et al., 2023). The three-armed version of the task limits the utility of simple strategies (Izquierdo et al., 2017), as at the point of reversal, participants cannot simply switch to the other choice, but must determine which of the other choices is optimal. This introduces a sense of ambiguity, and likely increasing cognitive demand (Daw et al., 2006).

In each trial of our version of the task, participants were placed in a room and asked to choose between doors on three different sides of the room. Each door represented travelling in a specific direction. After selecting a door, they learned if the door could be successfully opened or if it was locked. If the door opened, they could continue travelling in that direction. As such, doors that opened were optimal to achieving task success (i.e., exiting the building). Reward (and reversal of reward) were operationalised as the likelihood of doors in a given direction opening.

Using this experimental paradigm, we addressed a series of hypotheses, based on the theoretical claims of the threat-rigidity literature (Staw et al., 1981), previous research using an incidental manipulation of threat on reversal learning ability (Paret & Bublatzky, 2020), and prior work using a complex decision-making task (Laycock et al., 2024).

First, we predicted that participants in the threat condition would perform worse on the reversal task when compared to those in the nonthreat condition. We predicted participants in the threat condition would demonstrate significantly lower overall task performance when compared to those in the nonthreat condition.

Our second hypothesis was that participants in the threat condition would require more trial blocks to improve task score from baseline, when compared to those in the nonthreat condition, based on the expectation that participants in the threat condition would be slower to adapt behaviour to task demands.

We also used a computational model to estimate learning rates and inverse temperature across conditions. We predicted that learning rate and inverse temperature would be significantly lower in the threat condition, when compared to the nonthreat condition, based on the notion that participants in the threat condition would be less responsive to feedback (i.e., are less flexible), and that choice behaviour in the threat condition would be based to a lesser extent on feedback.

Methods

Participants

We recruited 100 participants (females = 72, males = 25, and non-binary = 3), all over the age of 16 years (mean = 21.81, SD = 6.87, median = 19, min = 18, max = 52). Every participant completed the measures without technical difficulties and received either course credit or an £8 gift card as compensation for their time.

The study received approval from the Ethics Committee of the Psychology Department at the University of York, and all procedures were conducted in accordance with relevant guidelines and regulations. The study was registered on OSF preceding data collection []. Finally, participants reported varying levels of gaming experience: 14% indicated no experience, 52% reported some experience, and 34% described having a great deal of experience.

An estimate of the sample size was determined by calculating the smallest effect size of interest, for a within-subjects analysis, using a dependent t-test at a reasonable Type II error rate (smallest d = .3, power = .80, α= .05). This power analysis suggested a sample size of 90.

VRCF

The VRCF was adapted from previous work which has demonstrated the effectiveness of manipulating environmental threat using VR (see for more detail: Laycock et al., 2024). The VRCF is comprised of three scenes: the prime-scene, test-scene, and end-scene.

During the prime scene, participants were placed in a virtual lift that travelled down eight floors over the course of 150 seconds. In the threat condition, participants experience a number of visually aversive events (e.g., windows breaking, fire spreading, and electrical failure). Audio was also used to instantiate the experience. These events were not present in the nonthreatening condition.

Next, participants enter the task scene. In the threat condition, participants were placed in a darkened room, although they had a torch, their field of view was restricted. Participants were exposed to aversive sounds taken from the prime scene, yet no visual events were used. The nonthreating condition included no emotive stimuli. During the task scene, the VRCF participants took part in a probabilistic reversal learning paradigm (Highgate & Schenk, 2021).

Here, participants were required to select “doors” from an array of three colour-coded options (A, B, and C). Following selection, doors either opened or were locked. At any point, one door was always optimal, that is, successfully opened on 70% of trials, and two where suboptimal (successfully opened on 30% of trials). After every 30 trials, this switched, and a new door became the optimal choice. During the task, optimal choices switched three times over the course of 120 trials (i.e., selections).

The central aim of the task was to open as many doors as possible. Therefore, improved performance on the task required participants to adapt their choice behaviour as a function of changes to the reward schedule. In both conditions, once all 120 trials had been completed, participants entered the end-scene and were instructed that they had escaped the building and the VRCF ended.

Hardware

Participants used a VIVE head-mounted display (HMD) with an integrated Dual AMOLED 3.6’’ diagonal screen, a resolution of 1080 x 1200 pixels per eye (2160 x 1200 pixels combined), refresh rate of 90 Hz, and a 110 degrees field of view to engage with the virtual task.

A wireless VIVE controller with dual-stage trigger and integrated HD haptic feedback was used to interact with task features during play. Participants wore wireless noise-cancelling headphones set to a maximum volume of 80% at all points during the task.

Participants wore a bHapticts (TactSuit X40) haptic suit in both conditions. The TactSuit X40 covers a subject’s torso. Using the Audio-to-Haptic feature (set to .05 feedback intensity), audio during the task was converted into haptic feedback via 40 synchronised vibro-tactile motors. This tactile stimulation was included to reinforce the feeling of immersion experienced during the VRCF. We made the addition in response to suggestions from participants following a pilot development phase.

Questionnaires

User Experience Scale (O'Brien et al., 2018).

The User Engagement Scale (UES) measures self-reported user engagement (O'Brien et al., 2018). As our intent was to use the UES as an index of real-world dissociation, we only used the three-item component of focused attention. A sample item from this scale was "I lost myself in this experience". Responses were recorded on a 5-point Likert scale (from 1 = not at all characteristic of my experience to 5 = entirely characteristic of my experience). Higher scores indicate greater focused attention during the VRIGT.

Affective response.

Affective experience was rated by participants following both the threatening and nonthreatening conditions. To do so, we asked participants to rate their overall affective response to the 14 categories of affect (e.g., frightening), using scales ranging from "not at all = 0" to "a great deal = 100" (as in: McCall et al., 2022).

Procedure

Participants first completed a composite survey (online: Qualtrics). The survey consisted of basic demographics questions (age, gender, gaming experience), as well as two individual difference questionnaires: the Intolerance of Uncertainty Scale (IoU; Carleton et al., 2007) and the Need for Closure Scale (NfC; Roets & Van Hiel, 2011). For analysis, see Supplementary Materials: Individual differences.

Next, participants had the opportunity to ask questions and test the equipment. This was achieved using a mock version of the VRCF. We ensured participants understood all functional aspects of the task.

All participants then completed both VRCF conditions of the study (threat/nonthreat). Before each condition of the VRCF, an audio narrative was played (see Supplementary Materials for transcripts). Participants then completed the VRCF. Each condition had two versions where the reward schedule differed. In version 1) doors B, A, C, A, were optimal with reversals every 30 trials. In version 2) the optimal door followed an alternative schedule, A, C, B, and C. Participants were randomly allocated into one of these versions for each condition. Further, the order of conditions (e.g., threat then nonthreat) was also randomised in a counterbalanced fashion.

After each VRCF condition, participants completed another survey online. This survey included questions about subjective experience regarding feelings of immersion and threat (e.g., "how threatening did you find the virtual world?"). Participants were also asked to complete a free-text section to provide additional details of their experience.

Analysis

We did all analyses in RStudio 4.2.1 [64] (R Core Team, 2022), using R basic, “lme4” 1.1 (Bates, 2006), and the “psych” package (Revelle, 2022). Here, p-values were calculated using the “lmerTest” package 3.1 (Kuznetsova et al., 2017), and the ANOVA function using Satterthwaite’s method for F tests. The “emmeans” package 1.6.3 (Lenth, 2023) was used for Pairwise post hoc comparisons. Post hoc comparisons used a Tukey correction for p-values. However, the MVT correction was applied when adjusting for multivariate comparisons (e.g., interactions). The median absolute deviation (MAD) was used to identify univariate outliers using the “routliers” package (Delacre & Klein, 2019). The “pwr” package was used for power analysis (Champely, 2020). Highest Density Intervals (HDI) of posterior draws were calculated at 89% (McElreath, 2016, 2018).

Computational model for learning rate and inverse temperature

We selected a simple Rescorla-Wagner model to robustly estimate parameters that best reflect fundamental associative learning processes via prediction error. While more sophisticated models may offer improved fits to behavioural variability (Browning et al., 2015; Huang et al., 2017; Weiss et al., 2021), our objective was not theoretical parsimony, but robust precision. The chosen model prioritised the interpretability and specificity of parameters indexing the core processes of interest, rather than generalised performance prediction.

We ran our model in STAN (Carpenter et al., 2017; Stan Development Team, 2023) using the R “cmdstanr” interface (Gabry et al., 2025), to estimate two parameters considered theoretically important for understanding the influence of threat learning (Supplementary materials). The model used Rescorla-Wagner delta rule to estimate, (1) Learning rate (alfa), where higher values represent a greater influence of reward on learning (from 0 to 1); (2), Inverse temperature (beta), where higher values indicate more deterministic decision-making, meaning choices are more strongly guided by learned values (Q-values) e.g., expectation formed over time based on past experiences (ranging from 0 to 5).

We used a Bayesian Hierarchical method to estimate parameters (Kruschke, 2014; Lee, 2011), given the demonstrated improved efficacy of MCMC and hierarchical process to search the parameter space, over regular optimisers (Ahn et al., 2011, 2017). This is especially useful when attempting to estimate multiple parameters and require group and individual-level parameter estimations. Parameter recovery across a range of simulated known parameter values was used to clarify model and method suitability (Wilson & Collins, 2019).

All models were sampled for 4000 iterations, with the first 1000 as warmup (i.e., burn-in) across four sampling chains. Model convergence was judged visually by inspection of trace-plots and assessment using the Gelman-Rubin test ($\hat{R}$), values < 1.1 suggest adequate model convergence (Gelman & Rubin, 1992).

Results

Subjective experience

Affect

To investigate differences in subjective experience between the threat and nonthreat conditions, we tested for differences in the post-task questionnaire across conditions. For the affect variables, we used an LMM to predict rating, with fixed factors for affective category (e.g., "unpredictable"), condition, and their interaction. Intercepts were allowed to vary as a random factor at the level of the individual. A significant effect of condition ($F(1, 2673) = 131.290$, $p <. 001$), and its interaction with affective category ($F(13, 2673) = 17.787$, $p <. 001$) was found on rating.

Next, we computed a series of post hoc comparisons of the effect of an affective term-condition interaction on ratings. To adjust for multivariate comparisons, we applied an MVT correction. Results suggested that between conditions, ratings that the VRCF was "frightening" (t. ratio (2673) = − 11.04, p <. 001), "creepy" (t. ratio (2673) = − 8.52, p <. 001), "surprising" (t.ratio (2673) = − 7.71, p <.001, "interesting" (t.ratio (2673) = − 4.31, p <.001, "engaging" (t.ratio (2673) = − 6.13, p <.001, and "unpredictable" (t.ratio (2673) = − 3.80, p =.002, where reported following the in the threat condition when compared to the nonthreat condition. Conversely, ratings that the VRCF was "boring" (t.ratio (2673) = 5.37, p <. 001) were higher following the nonthreat condition. All other were p's> .05 (Supplementary Materials: Linear Mixed Model Outputs).

Tension

In the threat condition, 10% of participants reported that their tension remained flat, showing no change over time. A similar proportion (13%) experienced increasing tension, while a larger portion (39%) felt their tension decrease as time passed, and 38% described tension as dynamic, fluctuating over time. In the nonthreat condition, 29% reported their tension as flat, while 20% experienced increasing tension. 31% of this noted a decrease in tension, and 20% felt that their tension fluctuated dynamically.

User experience

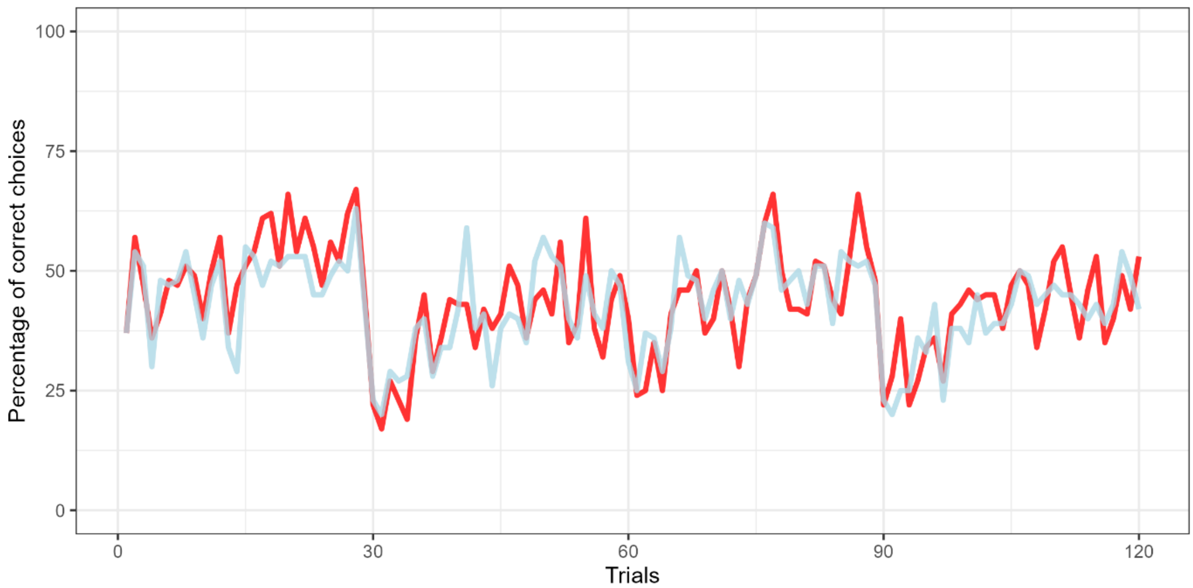
We also collected data on focused attention using the UES (O'Brien et al., 2018). The mean scores were M =3.59, SD = .82 in the threat condition and M =3.23, SD = .93 in the nonthreat condition. Both averages exceeded the threshold of 3 (the scaling midpoint, indicating "neither agree nor disagree"), suggesting participants experienced some degree of immersion (e.g., real-world dissociation).

Performance during the VRCF

Learning on the three-armed version of the reversal learning task is a considerable challenge. Despite this difficulty, at group level, participants did learn to identify the correct door in the final window of trials before a reversal, but it is apparent that as trials progressed, learning deprecated (Figures 1 and 2). This was demonstrated by estimating a critical value to determine a significant change. The critical value was calculated using a chi-square goodness-of-fit test to determine whether the observed proportion of correct responses significantly differed from the expected proportion under random chance (33.33% vs 66.66%). A significant deviation from expected values at a critical value of .40 was found: χ^2 (1) = 3.84, $p < 0.05$. This critical value represents the point at which the proportion of correct responses was significantly higher than expected under chance distribution. All points above this line may be considered above chance.

Figure 1

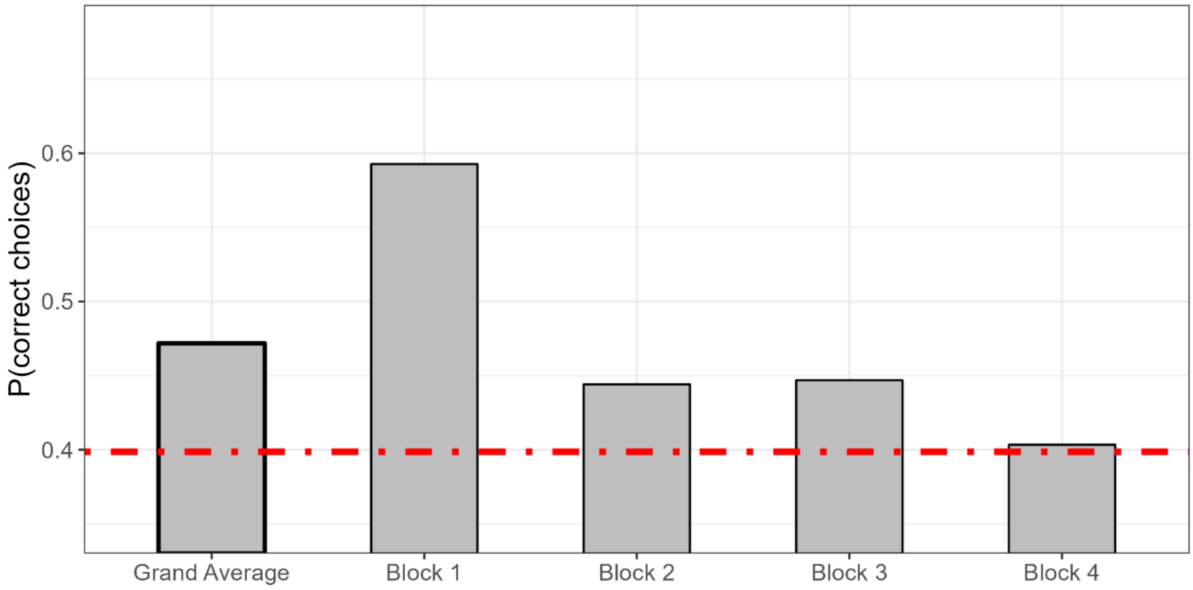
Percentage of correct choices by trial



Note. Illustrates the percentage of correct choices across trials. Coloured lines represent red (threat), blue (nonthreat), and black (average across conditions).

Figure 2

Learning: critical value



Note. The percentage of correct responses in the final five trials before a reversal was indexed, as a grand average (mean across all blocks) and individual blocks. These values are illustrated with reference to a critical value marked by the red vertical line.

Overall performance

To investigate our first hypothesis, that participants in the threat condition would have a significantly overall lower task score when compared to those in the nonthreat condition, we tested differences between the overall task score, i.e., the sum of doors opened (e.g., rewarded trials) during the task across all 120 trials. Results of a dependent t-test revealed no significant difference in overall score between the threatening (M = 53.67, SD = 8.36) and nonthreatening (M = 54.22, SD = 7.48) conditions, t (99) = .68, p = .498.

Further investigation was conducted using a Bayesian estimation of the posterior distribution of differences in means (threat – nonthreat) to test for the prospect of a null effect. No evidence was found (Supplementary Materials Figure 3). Indeed, this approach estimated an 81.8% probability that the true difference was below 0. $\hat{R}$ value <1.1 suggested adequate model convergence.

To assess the effect of order (e.g., which condition participants completed first) on overall performance, we ran an LMM with condition, order, and their interaction as fixed factors. The overall score was used as a dependent variable. Intercepts were allowed to vary as a random factor at the level of the individual. Results (Supplementary Table X) demonstrate that neither the effect of condition (F(1, 98) = .50, p = .479), nor order (F(1, 98) = 1.77, p = .186) score was significant. A significant condition-order interaction was found (F(1, 98) =

10.00, $p = .002$) on score. Further investigation using an MVT correction suggested that only when participants did the threat condition first (relative to second), did those in the threatening condition perform worse on the task (t. ratio (98) = 2.72, $p = .014$). However, this effect should be interpreted with caution, as we were not powered for a post hoc analysis and found a non-significant effect between the threat and nonthreat condition on overall performance when both were performed first (t. ratio (156) = 1.51, $p = .435$).

Performance over time

To analyse performance over time, we recalculated performance into four, thirty trial blocks. This allowed us to test our second hypothesis that participants in the threat condition would take more trial blocks to significantly improve their task score from baseline (block 1) compared to those in the nonthreat condition. To assess the effect of block on performance, we ran an LMM with block, condition, order (e.g., which condition participants completed first), and their interactions as fixed factors, and score as a dependent variable. Intercepts were allowed to vary as a random factor at the level of the individual. Results demonstrate that the effect of a block-condition interaction on score was not significant ($F(3, 686) = .611$, $p = .608$). The block-condition-order interaction was also not significant ($F(3, 686) = .74$, $p = .526$). To account for dependence between observations over time, an autocorrelation term was considered. The inclusion of an autocorrelation term did not significantly improve model fit ($p=.633$), AIC: autocorrelation term + (3957.37), autocorrelation term – (3955.60), indicating no meaningful improvement.

Computational modelling

To address our third and fourth hypotheses, that that group level learning rate and inverse temperature would be significantly lower in the threat condition when compared to the nonthreat condition, we used a computational model to estimate these parameter values between groups and at the level of the individual participant (see “Methods”). Supplementary Materials Figures 1 and 2 illustrate the posterior distribution of group-level parameter estimates.

Threat condition

In the threat condition group-level parameters were estimated as: learning rate (mu alpha) = .59 (SD= .038) and Inverse temperature (mu beta) = 2.48 (SD =.178). The 5th and 95th percentiles for learning rate were .52 and .65, respectively. The 5th and 95th percentiles for Inverse temperature were 2.19 and 2.77. At the level of the individual, variability across participants was also calculated: learning rate (sigma alpha) = .32 (SD =.026) and Inverse temperature (sigma beta) = 1.65 (SD =.125). Convergence diagnostics, as assessed by the Gelman-Rubin statistic ($\hat{R}$) were all < 1.01, indicating satisfactory convergence.

Nonthreat condition

In the nonthreat condition, the group-level parameters were estimated as: learning rate (mu alpha) = .507 (SD= .040) and Inverse temperature (mu beta) = 2.25 (SD =.178). The 5th and 95th percentiles for learning rate were .44 and .57. The 5th and 95th percentiles for Inverse temperature were 1.96 and 2.55. Again, individual variability across participants was also calculated: learning rate (sigma alpha) = .34 (SD =.027) and Inverse temperature (sigma beta) = 1.53 (SD =.120). $\hat{R} < 1.01$, indicated satisfactory convergence.

Comparison between conditions, learning rate

Learning rate was estimated as higher in the threat condition (Supplementary Materials Figure 4). To further explore differences in estimates between conditions, we first ran a dependent t-test on the estimates of individual learning rate and calculated bayes factor for this comparison, $t(99) = 3.47$, $p < .001$, $BF_{10} = 28.17$. This suggests that estimates of learning rate in the threat condition are significantly higher when compared to the non-threat condition. To provide a more comprehensive understanding of the uncertainty in this claim we calculated the HDI interval of difference between posterior draws (Supplementary Materials Figure 4). The probability of direction metric suggested a 92.45% likelihood that the true difference in learning rate between (Threat > Nonthreat) conditions was above 0.

Comparison between conditions, Inverse temperature

Inverse temperature was also estimated to be higher in the threat condition (Supplementary Materials Figure 5). As above, we ran a dependent t-test on the estimates of individual learning rate and calculated bayes factor for this comparison, $t(99) = 2.28$, $p = .024$, $BF_{10} = 1.32$. This suggests that estimates of Inverse temperature in the threat condition are significantly higher when compared to the non-threat condition. The HDI interval of difference between posterior draws was also calculated (Supplementary Materials Figure 5). The probability of direction metric suggests an 81.51% likelihood that the true difference in Inverse temperature between (Threat > Nonthreat) conditions was above 0.

Exploratory simulations

To understand how threat’s effects on learning rate and inverse temperature might influence performance in different environments, we next ran a series of simulations. Specifically, these simulations compared the effects of relatively higher versus lower learning rates and inverse temperatures in environments with different levels of volatility and/or stochastic variation. We implemented this series of reinforcement learning simulations (Wilson & Collins, 2019) using the Rescorla-Wagner model to analyse average reward under varying parameter sets (e.g., learning rate and inverse temperature), and task constraints (e.g., number of trials, volatility and stochastic variation) on a three-armed bandit task.

These model each agent selecting an arm using a softmax decision rule, where the probability of choosing an option is determined by its estimated value (Q-value) scaled by the inverse temperature (β) parameter. After selecting an arm, the agent receives a reward (1 or 0) based on the assigned probabilities and updates its Q-value using the learning rate (α) via a Rescorla-Wagner update rule (Supplementary Materials). All simulations were run for 10000 iterations (participants) per condition.

The simulations plotted in Figure 3 (a) represent performance (average reward) across 120 trials of a three-armed bandit task, where individuals select between three options based on reward probabilities (mirroring the task defined in "Methods") that change at predefined reversal points (30, 60, and 90 trials). This illustrates how parameter estimated from our experimental manipulations impact performance at scale.

Figure 3 (b, c, and d) simulated performance across 75 trials of a three-armed bandit task with two reversal points (at 30 and 60 trials) and modified levels of stochastic variation. This illustrates the accumulation of reward across various parameter sets.

Figure 3 (b) simulated no stochastic variation (i.e., noise). The reward probability across choice options was adapted (optimum = 1, suboptimum = 0 and 0). This demonstrates how, at both levels of inverse temperature, a high learning rate outperforms a low learning rate in an environment with no stochastic variation in choice reward.

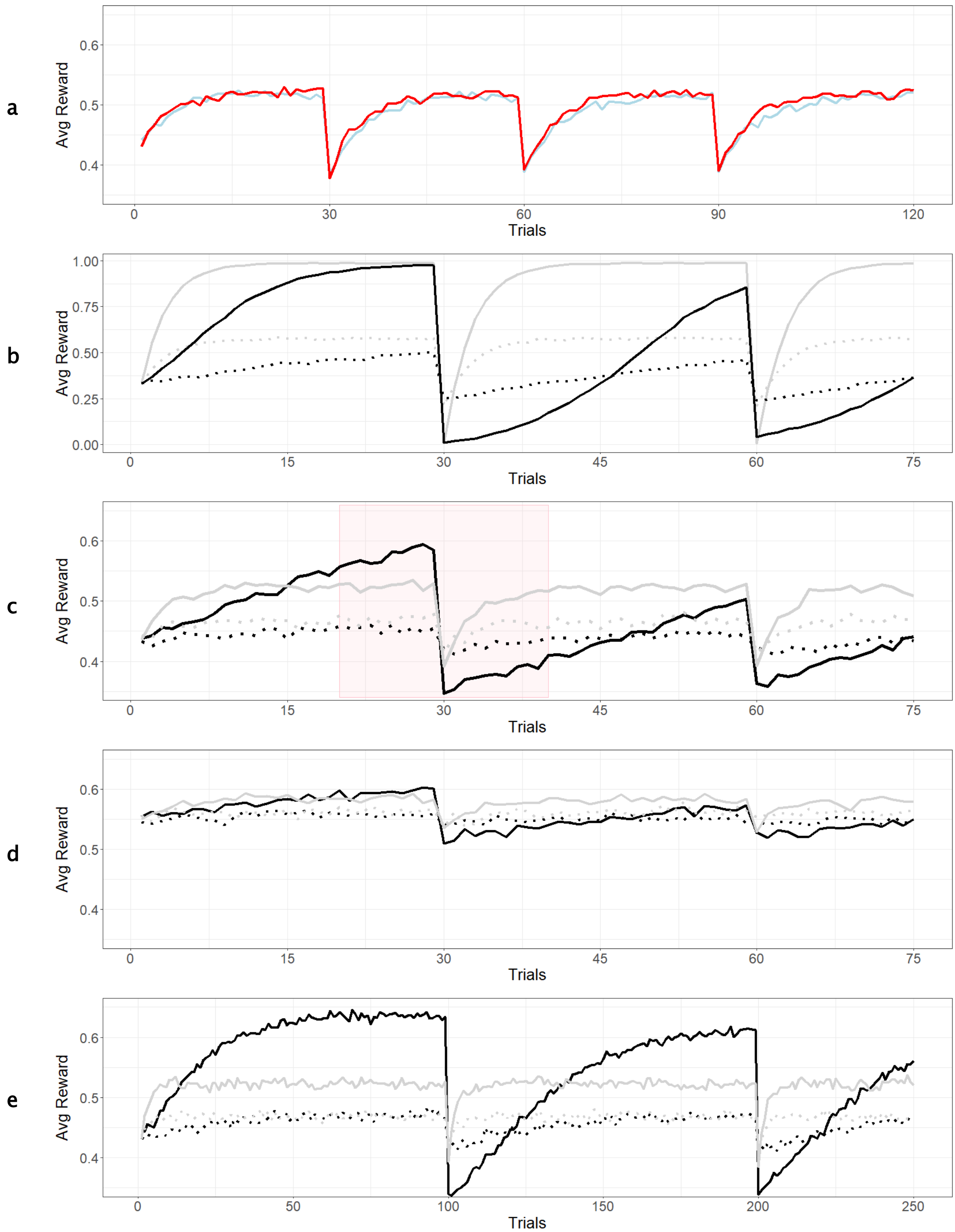
Figure 3 (c) simulated stochastic variation based on our task (optimum = .7, suboptimum = .3 and .3). This demonstrates how a low learning rate can outperform a high learning rate over time in a stable environment. However, following a reversal, the benefits of a high learning rate become apparent, and it is clearly more profitable in the short term (red-shaded region). We also see low inverse temperature impacts performance across both high and low learning rates in a similar fashion, blunting the effects of variations in learning rate on performance.

Figure 3 (d) simulated increased stochastic variation. Increased stochastic variation was introduced by adapting the reward probability across choice options, these changes represent a noisy environment (optimum = .7, suboptimum = .5 and .45). This makes identifying the optimum choice more difficult. When we add stochastic variation, the benefits of high rates diminish.

Figure 3 (e) simulated an extended time frame (i.e., more trials). We used the same constraints as Figure (c) but increased trials (250 trials, reversal points = 100, and 200). This highlights how a low learning rate can outperform high not only in stable environments, but also in noisy environments over an extended period. Focusing on the average reward, we see high learning rates level out after an initial benefit, with low rates doing a better job of identifying the most profitable option under noisy constraints.

Figure 3

Simulations



Note. Figure 3 (a). We used two sets of α and β values to observe how different learning rates and exploration tendencies affect total accumulated rewards. These sets represent parameters estimated from our experimental manipulations: Threat (red; $\alpha = .59$, $\beta = 2.47$), and Nonthreat (blue; $\alpha = .51$, $\beta = 2.25$). Figure 3 (b). No stochastic variation. We used four sets of α and β values to observe how parameter sets impact reward. However, sets are adapted to reflect a change in value and direction. Colours represent learning rate (black= low α, grey = high α). Variations in Inverse temperature are illustrated with dotted (low β) and solid lines (high β). Grey solid: $\alpha = .95$, $\beta = 5$; black solid: $\alpha = .1$, $\beta = 5$; grey dotted: $\alpha = .95$, $\beta = 1$; black dotted: $\alpha = .1$, $\beta = 1$. Figure 3 (c). Stochastic variation based on the task. The red-shaded region illustrates performance dynamics during the reversal phase. Figure 3 (d). Increased stochastic variation. Figure 3 (e). Increased trials and reversal windows.

Discussion

In the study presented here, we explored the effects of integral threat on cognitive flexibility by embedding a reversal-learning task into a virtual reality environment. This allowed us to observe the impact of threat on task performance and on specific associative processes that likely impact learning and decision-making. Although we successfully manipulated threat (Laycock et al., 2024), we found no significant effect of threat on performance, nor did we find solid evidence for the absence of an effect.

We also tested for the effects of threat on learning rate and inverse temperature. We predicted that group-level estimates for these parameters would be significantly lower in the threat condition compared to the nonthreat condition, as supported by the notion of threat-rigidity that claims threat promotes inflexibility (Giovanniello et al., 2023; Staw et al., 1981), reduces responsiveness to feedback (Berghorst et al., 2013; Bogdan & Pizzagalli, 2006), and moderates a shift from deliberation to intuition based computation (Schwabe & Wolf, 2011; Yu, 2016). Our findings were the opposite of our predictions. Learning rate and inverse temperature were both significantly higher under threat.

This increase in learning rate reflects faster adaptation to reward. Although it was not what we expected, this result echoes our previous findings that also reported increased sensitivity to reward following integral manipulation of threat (Laycock et al., 2024). Rather than being inflexible, individuals under threat were more responsive to feedback, albeit with a disproportional bias toward short team gains. This is in conflict with work that has demonstrated that threat reduces reward responsiveness (Berghorst et al., 2013; Bogdan & Pizzagalli, 2006). In those prior studies, following incidental manipulations of threat, participants were less sensitive to task-based reward (Bogdan & Pizzagalli, 2006), with some evidence that this effect was moderated by stress-induced anhedonia (i.e., a diminished motivation to initiate and sustain interest in rewarding tasks triggered by acute stress).

The disparity in findings regarding reward sensitivity may be a product of the method used to induce threat. In our study, reward takes the form of escaping from threat, making it integral to self-preservation and likely enhancing reward responsiveness. This is not the case with incidental threat paradigms, in which reward is not directly related to the threat and could even serve as an uninformative distractor from dealing with the threat. When reward is incidental to the threat, threat-induced anhedonia (Bogdan & Pizzagalli, 2006) may indeed be a common and adaptive response. This distinction highlights the value of thinking about the impact of threat on behaviour within context to better identify real-world implications, where function is likely influenced by the nature of the threat, task constraints, and concurrent demands.

We also found an increase in inverse temperature under conditions of integral threat. Changes in inverse temperature impact sampling behaviour. Our findings suggest that threatened individuals are more likely to exploit current understanding of choice values, and consequently, limit exploration. But why is threat increasing inverse temperature? One key aspect of integral manipulation of threat is that mistakes are consequential to self-preservation. In real-world contexts involving integral threat, the potential for absolute loss, such as a debilitating injury, outweighs the benefit of incremental gains (Fanselow, 2018). That is, surviving a dangerous situation is more important than incrementally optimising performance. Therefore, an increase in exploitative tendency may represent greater apprehension to extend choices beyond those recently experienced and rewarded. In a sense, this may reflect a desire to limit uncertainty and risk under threat (Wake et al., 2020; Walia & Gupta, 2025).

At face value, this explanation is aligned with prior work suggesting that threat promotes choice perseveration (i.e., repeated selection of a previously rewarded option despite changes in its value). Behaviourally, this can lead to premature closure, a term used to describe an individual's propensity to restrict exploration (Eva, 2009; Keinan, 1987; Pelaccia et al., 2014; Wemm & Wulfert, 2017). But, when learned values based on feedback do a poor job of predicting reward (e.g., in high noise environments), the opposite would be expected. High inverse temperature would result in increased sampling of different options as individuals follow the variability in rewarded options. This could explain why, using a complex task with dramatic shifts in reward values between options, we found that integral threat increased sampling of different options, seemingly promoting less perseveration (Laycock et al., 2024).

This is to say, the effect of an increase in exploitative tendency (increased inverse temperature) under threat may be task specific. In scenarios in which the relationship between choice and reward is clear and consistent, an increased inverse temperature would likely result in more choice perseveration (premature closure), given that learned values do a good job of predicting the environment, thus are stable. However, in complex situations, learned values do a worse job predicting the environment, and high inverse temperature would drive increased sampling as a product of unstable estimates of learned value. In other words, you should only follow a map closely if you trust it. In noisy environments, high inverse temperature may compound error, given that figuratively, the map is constantly changing. Again, this highlights that the effect of threat on behaviour is shaped by both the cognitive demands of the task and the underlying processing mechanisms it recruits (Kim et al., 2021; Mobbs et al., 2019).

These effects of threat on behaviour may have a broader impact on performance across different task constraints. Threat promotes a more responsive learning profile that leads individuals to prioritise short term rewards. This adjustment has clear benefits in well-structured (high signal-low noise) environments even when they're volatile, as it drives rapid adaption to change and uncompromising action. Our simulations illustrate this. Higher learning rates and inverse temperature are clearly optimal in high volatility environments with limited stochastic variability (Figure 3, b, and c), particularly after reversals in rewarded choices. This pattern is in line with the notion that a higher learning rate (Levy & Schiller, 2021) and inverse temperature are optimal (Tuzsus et al., 2024).

However, in a noisier environment, the benefits of a high learning rate for dealing with volatility diminish. Akin to the phenomenon of catastrophic interference (McCloskey & Cohen, 1989), error due to overwriting old knowledge with new feedback can lead to rapid loss of previously learned contingencies (especially in complex environments). Our simulations demonstrate that although high learning rates result in swift adjustment to changes in reward patterns (i.e., volatility) in the short term (Figure 3c and 3d), lower learning rates mitigate catastrophic interference and lead to greater performance in the long term (Figure 3e). These benefits arise because low learning rates are less sensitive to stochastic variation in feedback, enabling steadier integration of new evidence while preserving previously learned contingencies.

These findings have important implications, as they suggest that when threat is integral to the decision at hand, individuals are more responsive to reward and faster to adapt to change. Individuals are also more likely to use feedback to direct choice behaviour. Yet, in noisy environments, this shift in associative learning may render individuals at an increased risk of abandoning an action that has been consistently safe in favour of one that was merely recently rewarding. Learning in a complex environment is challenging, and in a noisy environment, good choices that give poor outcomes are increasingly common. Although threat might elicit a responsive and exploitative learning profile, one that is cautious yet curious may be more beneficial in the long term.

Strengths and limitations

A key strength of the present study is that the paradigm uses a decision-making task in which one's decisions are directly related to a threat. This factor is not trivial, as the nature of each stressor should be carefully considered when interpreting any influences on behaviour (Giovanniello et al., 2023). Integral manipulations of threat explicitly model scenarios where threat is not a distraction, but information. In this sense, the current research examines situations that require an individual not to ignore a stressor, but to use it to make an adaptive choice (Laycock et al., 2024; Ting et al., 2022). Using integral manipulations of threat is essential if we are to understand situations where individuals are making threat-critical decisions and don't have the luxury to neglect the threat. Many institutions look to the literature to develop training programs to prepare personnel to negotiate complex decisions within hazardous environmental constraints, e.g., military (Blacker et al., 2019; Delahaij et al., 2006; Männiste et al., 2019). Therefore, robust and contextually appropriate research is essential. To this end, future studies should extend our findings by testing participants across diverse simulated scenarios to rule out any scenario specific effects. Real-world data may further support our claims. For instance, video footage from actual events could be used to examine reward responsiveness and sampling behaviour in threatening and complex environments (Luxem et al., 2023; Philpot et al., 2025; Philpot & Levine, 2022).

Further, we employed a three-armed version of the classic reversal learning task. This was selected to best reflect the way information is used in support of decision-making (i.e., a choice between alternatives). The way a researcher decides to implement the functional features of the task selected to index performance is an important procedural choice. Features of the task influence performance. For example, the number of options an agent must choose from following reversal, the probability of reward across alternatives, and the stability of reward probabilities across trials (D'Cruz et al., 2011; Izquierdo et al., 2017; Walton et al., 2010), and the number of reversals experienced (Costa et al., 2015; Izquierdo & Jentsch, 2012). Future research should contrast performance across different versions of the reversal learning task to better understand how threat impacts associative learning performance as a function of changes in task demand, such as the number of choices presented to an individual, level of stochastic variability, and rates of volatility.

Conclusion

In this study, participants completed the same reversal learning task in both a neutral setting and a setting in which decisions were coupled with threats in the environment. We found that this integral manipulation of threat resulted in higher estimates of learning rate and inverse temperature, reflecting an increase in responsiveness to feedback and a greater tendency to exploit current understanding. This challenges the threat rigidity perspective, that proposes threat drives inflexibility and reduces responsiveness to feedback. This shift in process may support rapid action under volatile, well-defined constraints but may restrict adaptive action when situations are complex.

Chapter Four: Individuals Concerned with Negative Evaluation are Less Motivated to Engage with Complexity. Implications for Leadership Training.

Author contributions

Conceptualisation: Aaron Laycock., Cade McCall. Methodology: Aaron Laycock., Cade McCall. Data acquisition: Aaron Laycock. Data curation: Aaron Laycock. Data analysis: Aaron Laycock., Cade McCall. Visualisation and figures: Aaron Laycock., Cade McCall. Writing, coordination/editing: Aaron Laycock., Cade McCall. Project administration: Aaron Laycock.

Code availability

Analysis code, preregistration, and supplementary materials are available on the OSF repository.

Follow: https://osf.io/w2xdb/?view_only=835d55583b7c4a69b738af23b0578655

Abstract

Leadership training often involves evaluating trainees on their performance. However, excessive concern with being evaluated (i.e., concern with being right or wrong) might limit trainees' willingness to acknowledge and engage with real world uncertainties (i.e., the absence of clear right or wrong options). Across three correlational studies, we show that the higher individuals are in evaluation sensitivity (as measured by Fear of Negative Evaluation and Power Distance scales), the more intolerant they were of complexity (as measured by Intolerance of Uncertainty and Need for Closure scales) in both military and civilian samples: Study 1 (n = 211), Study 2 (n = 101), and Study 3 (n = 199). However, we also present two training-based studies, Study 4 (n = 87) and Study 5 (n = 76), to show that leadership training approaches focused on coaching and formative assessment can reduce complexity intolerance. Together, these data suggest that the evaluative nature of training programs might interfere with one's willingness to deal with real world complexity but that this risk can be mitigated by directly engaging with uncertainty during training.

Introduction

Individuals in critical roles frequently work in complex and unpredictable environments. Emergency healthcare workers treat patients with incomplete background information, air traffic controllers confront shifting schedules and weather vagaries, and military personnel face the unpredictability of hostile agents. To perform optimally, these individuals must identify key uncertainties, gather relevant information, modify plans or protocols, and take action, often in a limited amount of time (Comfort & Wukich, 2013; Lipshitz & Strauss, 1997). To state the obvious, these processes are cognitively demanding. And as research on complex decision-making demonstrates, performance relies not just on cognitive abilities and capacity, but on the motivation to grapple with complexity (Brun et al., 2023; Disatnik & Steinhart, 2015).

With all this in mind, preparing individuals to perform optimally in complex, high-stress environments presents a considerable challenge (Comfort & Wukich, 2013; Delahaij et al., 2006). Even if training successfully imparts the skills and knowledge necessary for a given role, trainees may lack the motivation necessary to confront complexity when it arises. That motivational hurdle may be even greater in training environments where performance is evaluated, and risk-taking involves material or reputational costs. In the studies presented here, we explored the possibility that the evaluative nature of training itself may work against one's motivation to deal with complex situations. We furthermore explored the potential for novel training approaches to address this risk.

Complexity intolerance

A variety of evidence demonstrates that people vary in their motivation to deal with complexity, i.e., their complexity intolerance and this variance has important implications for cognition and behaviour. The drive to find quick, concrete solutions associated with these individual differences can be maladaptive in complex situations when exploring multiple options, seeking out and integrating new information, and experimenting with possible solutions is optimal. While finding any answer may help reduce uncertainty in the short term, it might hinder the development of a better understanding of the best answer over time. Indeed, complexity intolerance is associated with deficits in information processing (Bartoszek et al., 2022), decision-making (Carleton et al., 2016; Kornilov et al., 2015), and real-world behaviour under uncertainty (Bavolar et al., 2021; Brun et al., 2023).

Both clinical and social psychological research has explored complexity intolerance through the constructs of intolerance of uncertainty and the need for closure. Intolerance of uncertainty is a trait defined by negative beliefs about uncertainty and its ramifications (Buhr & Dugas, 2002). From its theoretical conception, intolerance of uncertainty (Freeston et al., 1994) has been described as multidimensional, consisting of beliefs (e.g., uncertainty should be avoided), affective states (e.g., frustration and stress caused by uncertainty), and behaviours (e.g., uncertainty preventing action). Further work on the Intolerance of Uncertainty Scale (IoU; Carleton et al., 2007) suggests that it represents both a future-oriented, negative response to uncertainty (prospective anxiety), and its consequences to behaviour (inhibitory anxiety). Overall, IoU describes a tendency to find uncertainty aversive and disruptive.

While intolerance of uncertainty has often been studied as a feature of clinical phenomena (Freeston et al., 1994; Grupe & Nitschke, 2013; Hong & Cheung, 2015), recent research has explored how more general individual differences in IoU can impact cognition and behaviour. For example, IoU has been shown to predict risk-averse behaviour and performance deficits on complex decision-making tasks (Carleton et al., 2016), reduced confidence in decisions and behavioural inhibition in unpredictable situations (Jensen et al., 2014), more frequent selection of an immediate reward in decision-making tasks (Luhmann et al., 2011), and disruption in information-seeking behaviour (Bartoszek et al., 2022). Together, research suggests that IoU reflects both distress with uncertainty and the belief that everything must be planned (Bottesi et al., 2020), giving rise to behaviours that reduce, avoid, or remove uncertainty and its associated distress from a given situation, even when those behaviours are counterproductive.

Need for closure (Kruglanski, 1990) is a related trait-like construct focused more squarely on motivation (Berenbaum et al., 2008; Kruglanski & Webster, 1996; Roets et al., 2015). Individuals high in the Need for Closure Scale (NfC) are ostensibly driven to find "an answer on a given topic, any answer ... compared to confusion and ambiguity" (Kruglanski, 1990). Failing to do so can even result in physiological changes (increased systolic blood pressure and elevated heart rate) and subjective reports associated with distress (Roets & Hiel, 2008). In cognitive terms, when a tasks complexity exceeds an individual's capacity to meet the computational demands required to perform successfully, individuals high in NfC are inclined to strive for simplification and predictability (Kossowska, 2007). As a consequence, individuals high in NfC tend to use heuristics (Roets et al., 2015) and to show decreased effort when searching for information during decision-making tasks, resulting in restricted hypothesis generation and rapid decision-making (Kruglanski & Webster, 1996). Further, NfC has been associated with an unyielding adherence to pre-existing knowledge structures and a resistance to incorporating new information into their decision-making (Disatnik & Steinhart, 2015).

Both IoU and NfC provide insight into an individual's complexity intolerance. That is, they represent one's ability and willingness to grapple with uncertainty and to engage in decision-making and problem-solving when the best choice of action is not clear.

Evaluation sensitivity

Given the potential importance of complexity intolerance to working in critical environments, it is important to identify factors that amplify or exacerbate it. One potential factor is evaluation sensitivity, one's sensitivity to being negatively evaluated by others (e.g., team leaders, colleagues, etc.). Effective decision-making in complex situations often requires creativity, exploration, and a trial and error. But if an individual is afraid of the social consequences of making errors, they may not actively explore the full range of options (Ben-Zur & Zeidner, 2009) and creative problem-solving may be limited (Bonetto et al., 2021). This may be especially true in contexts in which a negative evaluation comes with a material cost, such as when one is being assessed during an important training exercise where poor performance may negatively impact future prospects. Two separate constructs are likely to influence evaluation sensitivity, particularly in hierarchical environments, fear of negative evaluation and power distance.

Fear of negative evaluation refers to an individual's concern with being judged disparagingly or hostilely by others (Watson & Friend, 1969). The Fear of Negative Evaluation Scale (FnE) not only predicts a stronger impact of negative judgement of interpersonal feedback on the individual (Smith & Sarason, 1975; Zhang et al., 2023) but also risk-avoidant decision-making behaviour (Maner et al., 2007). It is also negatively correlated with both social risk-taking and the belief in one's creative capacity (Bonetto et al., 2020). All of these data suggest that the fear of negative evaluation might impact one's ability and motivation to cope with handling complex situations.

The second potentially important construct here is power distance, the tendency to believe hierarchical divisions are both valid and required to maintain standards (Yoo et al., 2011). Hierarchical power structures are likely to encourage evaluation sensitivity, given that individuals working in those power structures are dependent on superiors' evaluations of their behaviour. In the military, for example, both direct and indirect cohesive power act as a method of maintaining cultural norms and institutional standards (Soeters, 2018). In this sense, military institutions reduce uncertainty in operational environments by utilising hierarchical power structures. As such, rank provides a chain of command for fast and decisive action. In this sense, military structures traditionally encourage some degree of power distance.

With all of this in mind, we hypothesize that the excessive concern with power differentials can amplify one's fear of negative evaluation. This potent combination, in turn, can increase complexity intolerance as one avoids the costs of failing via creative problem-solving, risk-taking, and trial-and-error learning (Ben-Zur & Zeidner, 2009; Bonetto et al., 2020).

Complexity intolerance and evaluation sensitivity in training contexts

This possible relationship between evaluative threat and complexity tolerance raises serious concerns for training individuals to cope with complex situations. The inherently evaluative nature of training may inadvertently undermine the acquisition of skills in dealing with uncertainty. Again, the military provides a key example here. Traditional military training doctrines generally aim to reduce uncertainty, not necessarily to better prepare individuals to negotiate it (Delahaij et al., 2006). This is partly achieved by training approaches that aim to develop a habituated response in individuals (as part of a team) to pre-scripted mock exercises via a process of structured and summative scenarios that mirror the operational environment. While this focus on habituation may be necessary for training and testing basic skillsets, habituated responses cannot prepare individuals to flexibly face dynamic and unpredictable operations (Männiste et al., 2019; Van Den Heuvel et al., 2012; Williams, 2010). As such, trainees are likely to benefit from cultivating their ability and willingness to negotiate complexity.

However, the complex nature of operational environments can require observation, cognitive engagement, and input from all levels of the hierarchy. With this in mind, excessive focus on evaluation and deference to one's superior (i.e., evaluation sensitivity) might reduce one's willingness to engage with complex problem-solving when it could be perceived as going against the leadership's plans (Pattni et al., 2019).

Conversely, if there is a relationship between evaluation sensitivity and complexity intolerance, then training approaches that minimize evaluation sensitivity might better prepare trainees for working in complex environments. While these traits are often assumed to be relatively fixed (Bavolar et al., 2021), prior research certainly suggests that both longer term therapeutic approaches (Laposa et al., 2022; van der Heiden et al., 2012) and short term manipulations of context (Djikic et al., 2013; Ladouceur et al., 2000) can affect complexity tolerance.

Evaluation sensitivity is similarly malleable (Busch et al., 2023; Dogaheh et al., 2011). As such, the prospect of a training environment positively affecting it is reasonable. Along these lines, recent developments in military training address these concerns by creating a training environment that promotes experimentation via a greater focus on formative assessment of decision-making processes and mentorship or coaching through complex scenarios (Finn, 2019). While some degree of summative assessment is unavoidable in training (especially safety-orientated training with a clear right/wrong outcome, e.g., medical), these new training approaches may reduce complexity intolerance and buffer the effects of evaluative threat.

The current studies

The first aim of the current studies was to test the hypothesis that evaluation sensitivity is positively associated with complexity intolerance. Given the relevance of this question to military training, we tested for this relationship in both early career and experienced military personnel as well as within a cross section of the general population. The second aim of the current studies was to test the possibility that different leadership training approaches might affect complexity tolerance. Here, we tested if military training interventions geared towards coaching and formative assessment might reduce complexity intolerance by encouraging individuals to engage with uncertainty.

Study 1, Experienced Military Personnel

Study 1 examined the relationships between measures of evaluation sensitivity and complexity intolerance within an experienced military sample. First, we hypothesised that within-construct measures of evaluation sensitivity (Fear of Negative Evaluation and Power Distance scales; FnE and PD) and complexity intolerance (Intolerance of Uncertainty and Need for Closure scales; IoU and NfC) would be correlated. Further, we hypothesised a between-construct correlations as well. That is, individual differences in both FnE and PD (evaluation sensitivity) would predict ratings of IoU and NfC (complexity tolerance).

Methods

Participants

Two hundred eleven (*n* =211) participants were recruited by internal staff from the ranks of the United Kingdom's military. All participants were male, and ages ranged from 18 to 40 years. All participants had completed at least eight months of pass/fail specialist training program conducted at the Commando Training Centre Royal Marines (CTCRM). As such, participants were experienced personnel. The data collection took place between 2020 and 2021.

The minimum sample size for this study was determined by the results of a power analysis to estimate the sample size necessary to detect the smallest effect size of interest for our correlational analyses at a reasonable Type II error rate (smallest r = .20, power = .80, α= .05), which suggested a sample size of 193. In the absence of an empirically grounded estimate of effect size, we chose the smallest effect size of interest (r = .20) following Funder & Ozer's (2019) guidelines for an effect that is of some explanatory and practical use. Because of the nature of recruitment (i.e., sharing a link to the questionnaire with the entire course cohorts), we did not use this power analysis as a cut-off for gathering data. Instead, we analysed all data we received once the link was distributed.

Approval for the data collection was granted by the ethics committee of the Psychology Department at the University of York and the CTCRM research steering group.

Materials

Evaluation sensitivity

Fear of Negative Evaluation (Leary, 1983)***.*** The original Fear of Negative Evaluation scale (Watson & Friend, 1969) was developed to measure the degree to which individuals feel apprehension at the prospect of receiving negative judgment. We used a brief version of the original scale, the Brief Fear of Negative Evaluation Scale (FnE; Leary, 1983), which consists of 12 items that require a response on a 5-point Likert scale from, 1 =" not at all characteristic of me", to 5 =" entirely characteristic of me". Higher scores indicate higher apprehension of judgment. A sample item is "I am afraid people will find fault with me".

Power Distance (Yoo et al., 2011)***.*** Individuals higher in power distance believe that authority figures should be respected and shown deference (Yang, Mossholder, & Peng, 2007). We used an index of power distance (PD) taken from the CVSCALE. The CVSCALE was developed to measure Hofstede's dimensions of culture (Hofstede, 1984) at the individual level (Yoo et al., 2011). It includes five items that required a response on a 5-point Likert scale, from 1 =" not at all characteristic of me" to 5 =" entirely characteristic of me". A sample item is "people in lower positions should not disagree with decisions". Higher ratings of power distance indicate a belief in the need for authority in the workplace.

Complexity intolerance

Intolerance of Uncertainty (Carleton et al., 2007)***.*** The Intolerance of Uncertainty short form (IoU; Carleton et al., 2007) assesses an individual's valanced associations to uncertainty and ambiguous situations. The IoU is a 12-item condensed version of the original 27- item Intolerance of Uncertainty Scale (Freeston et al., 1994). A sample item is "I can't stand being taken by surprise". Responses are rated on a 5-point Likert scale, from 1 =" not at all characteristic of me", to 5 =" entirely characteristic of me"). Higher scores indicate greater Intolerance of Uncertainty.

Need for Closure (Roets & Van Hiel, 2011)***.*** The Need for Closure Scale was developed by Webster & Kruglanski (1994), and revised by Roets & Van Hiel (2007) to index one's motivational drive to seek and maintain certainty when dealing with ambiguous informational constraints. Here we used a brief version (NfC; Roets & Van Hiel, 2011). This scale contains 15 items that are rated on a 5-point Likert scale (from 1 =" not at all characteristic of me", to 5 =" entirely characteristic of me"). A sample item is "I would quickly become impatient and irritated if I would not find a solution to a problem immediately". Higher scores indicate greater need for closure.

Procedure

To distribute the questionnaires online, a composite survey was created in Qualtrics (Qualtrics, Provo, UT). Participants were instructed to complete the questionnaire in a quiet environment and answer all the questions accurately and honestly, using their phone or computer. The online link and QR code used to access the survey was internally distributed by military staff.

Once directed to the questionnaire, participants received an online initial brief, an agreement of consent and personal data was requested. Following this stage scales were administered to participants (FnE, PD, IoU, and NfC) in a randomly selected order. The final section of the questionnaire issued a full debrief and contact details for further correspondence.

Analyses

A Cronbach's alpha statistic was calculated to provide a measure of internal consistency for each scale. In Study 1 and throughout the paper, we report any scales with $\alpha < .70$ (Nunnally & Bernstein, 1994). To identify any univariate outliers, we calculated the median absolute deviation (MAD) across subscales (Leys et al., 2013), using a threshold of 3* MAD (+/- median).

To evaluate the relationship between variables we used a series of Pearson's correlation coefficients. Effect sizes were interpreted as $r = .10$ (small), .20 (medium), >.30 (large), following guidelines suggested by Funder & Ozer (2019). We also used a Holm correction to adjust for familywise error (Holm, 1979). To further support results, bias corrected and accelerated bootstrapped 95% CIs are also reported. All bootstrap results are based on 9999 bootstrapped samples. Although some of the scales showed non-normal distributions, here we report parametric tests given their robustness against violations of normality and relatively large sample ($> n = 30$) sizes (Field, 2000; Runyon et al., 2000). Moreover, we also calculated Spearman's coefficients to confirm the pattern of results with non-parametric tests where appropriate (Rovetta, 2020; see Supplementary Table 1 for those analyses). Thus, parametric, non-parametric tests, and bootstrapped confidence intervals were calculated for all correlational analyses.

Transparency and openness

We describe our sampling plan, all data exclusions, all manipulations, and all measures used in the study. All analyses were performed in R (R Core Team, 2022), version 4.2.2. Packages used in analysis include: psych (correlations and Cronbach's alpha statistic: Revelle, 2022), confintr (bootstrapped CI's: Mayer, 2023), pwr (power analysis: Champely, 2020), and BEST (Bayesian estimation of the posterior distributions: Meredith & Kruschke, 2021). Finally, outliers were identified using routliers package (Delacre & Klein, 2019). This study's design and its analysis were not preregistered.

Results

All Cronbach's alphas in Study 1 were above the .70 threshold, and no univariate outliers were detected.

As expected, our two measures of evaluation sensitivity, FnE and PD, were positively correlated with a medium effect size (see Table 1). Our two measures of complexity intolerance, IoU and NfC, were also positively correlated but with a large effect size (see Table 1). Additionally, we found positive correlations between our measures of evaluation sensitivity and complexity intolerance (see Table 1). These findings all support our hypothesis that greater evaluation sensitivity predicts higher levels of complexity intolerance in an experienced military population.

Interim Discussion

Study 1 aimed to test the relationship between evaluation sensitivity and complexity intolerance. Our measures of evaluation sensitivity (FnE and PD) correlated with the measures of complexity intolerance (NfC and IoU), revealing a pattern whereby the more an individual is concerned about being evaluated negatively, the less motivated they are to engage with uncertainty.

Study 2, Early Career Military Personnel

Study 2 collected data from early career military personnel. Through this preregistered study we sought to replicate the relationship between evaluation sensitivity and complexity intolerance demonstrated in Study 1.

Methods

Approval for the data collection was granted by the ethics committee of the Psychology Department at the University of York, and CTCRM research steering group.

Participants

All participants (n =101) were recruited by internal staff from two cohorts of early career military personnel enrolled in initial officer training at CTCRM. All participants were male. Sample size was determined by opportunity; we had access to two training cohorts and attempted to recruit all individuals in these cohorts. Ages ranged from 18 to 33 years. The data collection took place between 2021 and 2023.

Materials

The questionnaires in Study 2 were the same as those in Study 1.

Procedure

Data were collected at CTCRM, across two cohorts of officer recruits. The QR code used to access the survey was distributed at CTCRM by research staff. All participants were told that questionnaires could be completed on a phone or any computer.

Transparency and openness

The approach to transparency and openness is the same as Study 1. In addition, this study was preregistered [81939].

Results

All analyses were performed using the statistical software packages and in accordance with the details specified in Study 1. Again, parametric, non-parametric tests, and bootstrapped confidence intervals were calculated for all correlational analyses. Regarding internal reliability, the scale of PD raised some concern with a Cronbach's α=.58. As α is sensitive to the number of items included within a scale and homogeneity of total scores across the sample (Streiner, 2003; Taber, 2017), and as PD only consists of five items, this is perhaps not surprising. As such, PD was not excluded from analyses but should be interpreted with caution. Regarding outliers, one potential univariate outlier was identified on the IoU scale, with a high score (relative to sample average) of 3.9. Although we included this data point in the analyses reported below, we also ran each relevant analysis without it to confirm that the pattern and significance of our reported results holds.

In terms of the main analyses, we first assessed the direction and significance of correlations within and across the theorised constructs of evaluation sensitivity and complexity intolerance of all participants at T0. The results here replicate Study 1, with regards to FnE, IoU and NfC (see Table 1). However, in this population of early career military personnel, the correlation between PD and FnE and between PD and IoU were no longer statistically significant.

Interim Discussion

In Study 2, we replicated most of Study 1's findings within a sample of early career military personnel. As with Study 1, Fear of Negative Evaluation was positively correlated with both Intolerance of Uncertainty and Need for Closure. In this sample, however, Power Distance was only significantly associated with Need for Closure, although that may be due to the relatively low internal consistency of the scale in this sample.

Study 3, Civilian sample

Studies 1 and 2 demonstrated a relationship between evaluation sensitivity and complexity intolerance in military samples. It is not clear, however, if this relationship generalizes to the wider public. Indeed, militaries have hierarchical structures, and promotion within those structures partly depends on performance in training settings. In principle, this fact could amplify the influence of evaluation sensitivity on other cognitive processes and motivations, including complexity intolerance. Moreover, military contexts are at least stereotypically associated with more authoritarian leadership (e.g., the notion that strict obedience to direct orders is required). While actual approaches to leadership within the military are far more diverse (Fosse et al., 2019; Hannah & Sowde, 2012; Kark et al., 2016), the fact remains that well-rehearsed drills, norms, rules, and plans are important to operational settings. In this type of setting, evaluation sensitivity might be uniquely coupled with complexity intolerance (i.e., sticking to known responses). On the other hand, the relationship between evaluation sensitivity and complexity intolerance may be a more general phenomenon.

To examine this possibility, Study 3 tested the degree to which the relationships that we observed between evaluation sensitivity and complexity intolerance in Studies 1 and 2 would replicate in a civilian sample. More specifically, we also wanted to explore whether the authoritarian nature of an individual's working environment influences complexity intolerance or contributes to the relationship between evaluation sensitivity and complexity intolerance. Put simply, could the previously reported findings be attributed to the fact that the participants were working in military settings, which are potentially more authoritarian compared to other work environments?

Methods

Approval for the data collection was granted by the University of York ethics committee.

Participants

For Study 3 (n = 199), non-military participants were recruited online via Prolific (Prolific, 2014), Male (n = 99), Female (n = 99), and Non-binary (n = 1). The sample size was determined by the results of a power analysis to estimate the sample size necessary to detect the smallest effect size of interest for our correlational analyses at a reasonable Type II error rate (smallest r = .20, power = .80, α= .05), which suggested a sample size of 193. As with Study 1, we chose the smallest effect size of interest (r = .20) following Funder & Ozer's (2019) suggestions for a reasonable estimate of a useful effect. Given the recruitment approach (i.e., via Prolific), we were able to cut off enrolment once we had reached that sample size. All participants were in full time employment (as per restriction), and UK residents. Ages ranged from 18 to 47 years (M = 36.7, median = 35, SD = 11.5). Ethnic identification included White (81%), Asian (7%), Mixed ethnicity (4%), Black (3%), and Other (2%). The data collection took place in 2022.

Materials

Study 3 used the same set of questionnaires as the previous studies. In addition, we included the addition of the Authoritarian Leadership scale.

Authoritarian Leadership (Cheng et al., 2004)*.* The authoritarian leadership (AL-9) subscale was taken from the paternalistic leadership questionnaire (Cheng et al., 2004). The scale consisted of 9 items that were rated on a 5-point Likert scale, from 1 =" not at all characteristic of my experience", to 5 =" entirely characteristic of my experience". A sample item is "My supervisor exercises strict discipline over subordinates." Higher AL scores indicate that the individual has experienced authoritarian leadership in the workplace. We included this variable to address the possibility that the effects observed in Studies 1 and 2 were a product of individuals working in a hierarchical organization (i.e., to test if AL moderated the previously observed relationships between our measures of evaluation sensitivity and complexity intolerance).

Procedure

Data were collected using the online platform Prolific. The platform directed participants to the composite Qualtrics questionnaire. Data were collected at one time point.

Transparency and openness

The approach to transparency and openness is the same as in Study 1. In addition, this study was pre-registered [103805].

Results

Again, all analyses were performed using statistical software packages, and in accordance with details specified in Study 1. This also included checks of internal reliability, and univariate outliers across all scales. No issues with internal reliability were identified across scales (all $\alpha > .70$). However, a number of univariate outliers were highlighted amongst the ratings of PD and AL-9. As with study 2, these data points are included in all reported analyses, although all analyses were re-run while excluding outliers to confirm that all effects and their significance levels held. We found no significant differences in ratings across recorded measures between genders (all p's > .05; see Supplementary Materials).

The results of the correlational analysis replicate Study 1 (see Table 1). Regarding evaluation sensitivity, a significant and positive relationship was found between FnE and PD (see Table 1). Regarding complexity intolerance, a significant and positive relationship was found between IoU and NfC (see Table 1). Finally, robust associations were observed between theorised constructs, with strong effect sizes of positive relationships between the component scales of evaluation sensitivity and complexity intolerance (see Table 1). These findings all support our hypothesis, that higher levels of evaluation sensitivity are associated with greater complexity intolerance in a non-military population.

In addition, we were interested in investigating if the authoritarian nature of an individual's working environment might affect complexity intolerance or might drive or amplify the relationship between evaluation sensitivity and complexity intolerance. In other words, can the reported effects in Study 1 and Study 2 be partially explained by the fact that individuals in those samples were working in military environments which may be more authoritarian than other work environments. To test this possibility in the civilian sample, we ran a series of multiple linear regressions. Each regression predicted one measure of complexity intolerance (IoU or NfC) from one measure of evaluation sensitivity (FnE or PD), AL, and their interaction. If working in an authoritarian environment moderates the relationship between evaluation sensitivity and complexity intolerance, then the interaction term in these models should be significant. However, none of the interaction terms in these models was significant (all p's > .05, see Supplementary Materials for analyses). Furthermore, there was no significant effect of AL in any of these models (all p's > .05). As such, these analyses provide no evidence for the presence or absence of an influence of authoritarian working environment on complexity intolerance or its relationship to evaluation sensitivity.

Interim Discussion

Study 3 tested the relationship between evaluation sensitivity and complexity intolerance in a civilian sample. As with the military samples in Studies 1 and 2, evaluation sensitivity was associated with complexity intolerance. Furthermore, we found no statistically significant effect to support the idea that evaluation sensitivity, complexity intolerance, or the relationship between them, is dependent on working in an authoritarian environment. But regardless of whether or not working an authoritarian environment affects these phenomena, it appears that both military personnel and civilians tend to show reduced motivation to deal with uncertainty when they are high in evaluation sensitivity.

Study 4

The relationship between evaluation sensitivity and complexity intolerance shown in Studies 1 through 3 raises a serious concern regarding training individuals to work in critical environments. On the one hand, training for critical roles necessarily involves some element of evaluation. On the other hand, it is possible that the evaluation process itself might reduce individuals' willingness to engage with the inherent complexity of those environments. Our final two studies explored the possibility that training approaches might reduce this risk by openly engaging with uncertainty and using formative assessment as a coaching tool. Study 4 was an exploratory study using a subset of the participants from Study 1 to test if such a training program could reduce evaluative threat and complexity intolerance in a sample of experienced military personnel.

Methods

Participants

Data collected from experienced military personnel in Study 1 was used. Here, data were gathered at two time points (T0 and T1). A subset of participants (*n* =149) were enrolled on the operational command course (see below: "Training"). Their T0 data were gathered at the beginning of the six-week training. A further subset of those participants (n = 87) completed the training and, at that point, completed a second questionnaire (T1). This process was followed on three iterations of the operational command course.

Training

Operational command course

The operational command course is designed specifically to equip trained military personnel with the skills needed by tactical commanders in operational combative operations. The course is a six-week program delivered from multiple locations in the United Kingdom. During this intense course, trainees conduct daily simulated operational combat scenarios in a leadership capacity. These scenarios are designed to closely recreate the cognitive demands of leadership on dynamic task-based military operations. The operational command course is particularly focused on developing the process of adaptive decision-making.

Formative assessment focused training

This training intervention aimed to better prepare personnel for the uncertain and complex nature of operational exposure by creating a training environment that promotes experimentation by trainees (Finn, 2019). This program has been developed around changes specifically designed to decrease evaluation sensitivity within trainees and create a supportive learning environment. This involved a marked commitment to formative assessment. Here, training exercises were developed to better reflect the ambiguity inherent in operational environments. In some cases, summative evaluation is unavoidable (especially safety-orientated training with a clear right/wrong outcome, e.g., medical). Nevertheless, this was kept to a minimum and taken away specifically where it was deemed as being unsuitable to operational realities (i.e., dealing with probabilistic outcomes). Finally, the training staff received specialist training in mentorship. Historically, military training staff have held an authoritarian role amongst trainees (Finn, 2019; Hall, 2012; Soeters, 2018). A move towards a model of mentorship represents a paradigm shift in the deliverance of military command training and aims to create a supportive environment in which trainees feel empowered to learn as a function of trial and error, and reflection. The increased emphasis on these factors differentiates this program from traditional training methodologies.

Results

To test for effects of training, we calculated difference scores for each measure (T1 - T0). We then conducted a series of one sample t-tests (one-tailed) to examine if the training intervention resulted in a significant reduction (<0) in any of the scales. When we found significant effects with these t-tests, we explored them further using a Bayesian framework to model parameter values (Kruschke, 2018). These models were run with parallel processing using the default BEST package minimally informative default priors and model specifications (Meredith & Kruschke, 2021). This allowed us to comment on the presence of a null effect, report findings in terms of probability, and provide a summary of uncertainty around parameter estimates not constrained by the traditional Frequentist approach (e.g., power).

To explore changes in evaluation sensitivity and complexity intolerance following the operational command course, we used the T0 and T1 data collected from participants who had completed the training. Results showed a non-significant reduction in both measures of evaluation sensitivity, FnE: M= −.13, SD =.75, $t(86) = -1.63$, $p = .053$, and PD: M= −.02, SD = .71, $t(86) = -.24$, $p = .405$. However, we found a significant reduction across measures of complexity intolerance following the training intervention, NfC: M= −.15, SD = .63, $t(86) = -2.27$, $p = .013$, and IoU: M= −.11, SD = .57, $t(86) = -1.85$, $p = .034$. Bayesian estimations of the posterior distributions also were also calculated. All $\hat{R}$ values were <1.1, suggesting adequate model convergence. This allowed us to comment on the presence of a null effect of training. Results here were in support of previous findings. These results estimate an 86.9% probability that the true reduction in IoU was below 0 and a 93.2% probability that the true reduction in NfC was below 0 (see Supplementary Figure 1).

Interim Discussion

In Study 4, we assessed the effects of a new training designed to facilitate complex problem solving via more formative (and not strictly summative) assessment. That group showed an increase in complexity intolerance after training. On one hand, these results may have emerged because the novel training approach successfully increased trainees' ability to grapple with uncertainty. On the other hand, Study 4 did not have a control group for comparison, so we cannot make this causal attribution. Study 5 addressed this shortcoming with a suitable control group.

Study 5

In Study 5, we again sought to examine the effects of a training intervention on evaluative threat and complexity intolerance. We followed two training cohorts of officer initial training (A and B). The length, curriculum, and pass criterion remained the same across both cohorts. However, cohort B received training adapted in line with the formative assessment focused training (see Study 4) with the aim of reducing evaluation sensitivity. Based on the findings of Study 4, we expected the novel training to significantly reduce complexity intolerance.

Methods

Participants

Data collected from early career military personnel in Study 2 was used. Here, data were collected at two time points, T0 (start of training), and T1 (end of training), and across two cohorts (A and B). Cohort A (n =45) received traditional training; 39 of those completed the training and completed the post-training survey. Cohort B (n = 56) completed training designed to reduce evaluation sensitivity (see below); 37 of those completed the training and post-training survey. Critically, the course length, curriculum, and pass criterion remained the same across both cohorts of training.

The minimum sample size for this study was determined by the results of a power analysis to estimate the sample size necessary to detect the smallest effect size of interest in training-related change in the questionnaire variables at a reasonable Type II error rate (smallest d = .50, power = .80, one-tailed α= .05), which suggested a sample size of 26. Here, the smallest effect size of interest (d = .50) is typical of training-related research (Kraft, 2020). As with Study 1, our recruitment consisted of passing out a link to the questionnaire to each training cohort. As such, we used the sample size estimate as a minimum and analysed all data we received once we had distributed the link.

Training

Officer initial training

Officer basic training takes early career personnel (civilians) and aims to equip them with all the skills and knowledge required to deploy as a military commander operationally. The course lasts a minimum of 12 months and covers all aspects of tactical leadership. Like the operational command course, trainees conduct regular simulated operational combat scenarios in a leadership capacity. Generally, the development of adaptive decision-making skills is fundamental to officer initial training, however, unlike the operational command course, a wide array of capabilities are also addressed (e.g., fitness, weapons proficiency, administrative competence). All participants were enrolled in initial officer training. Cohort B received training with a greater focus on formative assessment (see Study 4: "Training").

Results

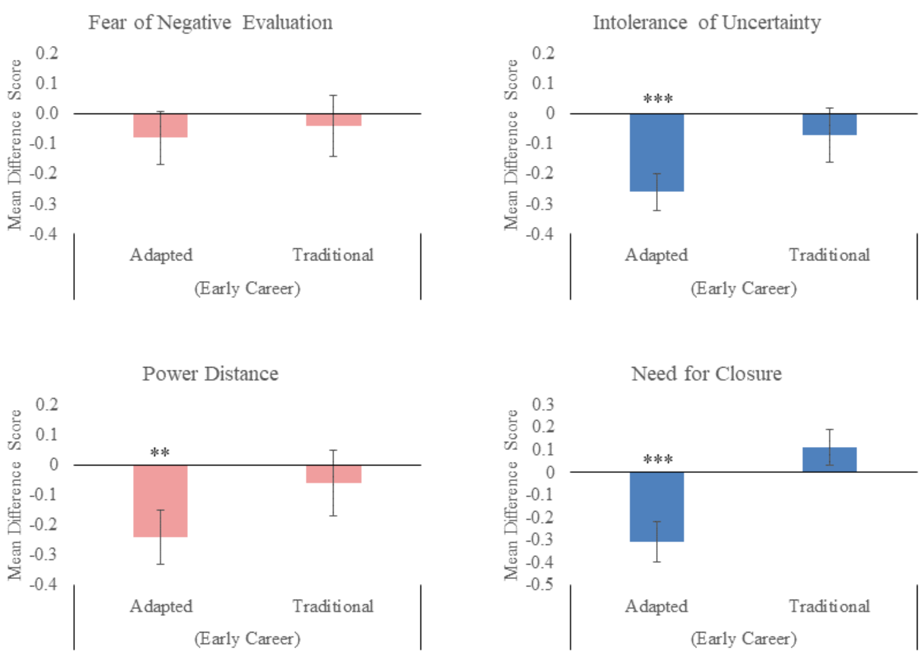
Next, we tested changes in evaluation sensitivity and complexity intolerance following training (see Figure 1). We found no change in Cohort A. As with Study 4, a difference score was calculated (T1 - T0) for all scales, FnE (M= -.04, SD = .64), PD (M= -.06, SD = .71), IoU (M= -.07, SD = .54), NfC (M= .11, SD = .53). Neither the measures of evaluation sensitivity, FnE: t(38) = -.42, p = .339 or PD: t(38) = .54, p = .705), nor complexity intolerance, NfC: t(38) = 1.29, p = .898; IU : t(38) = .86, p = .803) showed significant change following the traditional training intervention.

However, changes in several measures emerged in Cohort B; FnE (M= -.08, SD = .54), PD (M= -.24, SD = .57), IoU (M= -.26, SD = .29), and NfC (M= -.31, SD = .57). Here we found a significant reduction in PD, $t(36) = -2.61$, $p = .006$, and both measures of complexity intolerance, NfC: $t(36) = -3.28$, $p = .001$ and IoU: $t(36) = 4.09$, $p < .001$. Results showed a non-significant reduction in, FnE: $t(36) = -.86$, $p = .198$. The Bayesian analyses further support these results (see Supplementary Figure 2), estimating a 99.5% probability that the true reduction in IoU was below 0, a 99.7% probability that the true reduction in NfC was below 0, and a 98.8% that the true reduction in PD was below 0. All parameter estimates did not include 0 (no change) within the 95% HDI and can be considered strong evidence in support of a difference (Kruschke, 2018). Together, these data suggest that early career military personnel who underwent the novel training experienced a decrease in complexity intolerance.

Finally, we compared changes in evaluation sensitivity and complexity intolerance between cohorts. In accordance with our pre-registered analysis plan, difference scores were entered into a series of Welch's t-tests (one tailed). Results demonstrated that participants in Cohort B (when contrasted with Cohort A) had a greater reduction in PD: $t(71.96) = -2.08$, $p = .021$, and both measures of complexity intolerance, NfC: $t(72.90) = -3.29$, $p <.001$; IoU: $t(69.32) = -3.13$, $p =.001$. We found no significant difference between Cohorts A and B in FnE: $t(73.14) = -.25$, $p = .402$. Thus, findings suggest participants who underwent the formative assessment-focused training showed a reduction in PD, IoU and the NfC compared to those trained with traditional methods.

Figure 1

Effects of adapted and traditional training



Note. Mean difference of post-training from pre-training scores for early career personnel. Error bars represent standard error. *Indicates differences with a significant p value: *< .050, ** < .01, *** <.001.

Interim Discussion

Individuals who completed that specific training showed reduced complexity intolerance in both measures. Moreover, these changes were significantly greater for individuals who underwent traditional training. These findings support the idea that novel training approaches may increase one's ability to grapple with uncertainty.

Pooled Analyses

In the final set of analyses, we pooled the data from studies 1, 2, and 3 to address three points. First, we wanted to report relationships within and between measures of evaluation sensitivity and complexity intolerance with a larger and more diverse sample than any of the individual studies provided. Second, we wanted to use the greater statistical power of the larger sample to test the degree to which change in evaluative threat predicted change in complexity intolerance in the training samples. If the two constructs are causally related, then these changes should correlate. Third, we sought to test any differences between military and civilian samples in our key measures. Given stereotypes of military personnel as discipline-focused, rigid and lacking creativity (Harrell & Berglass, 2012), one might expect that military samples would be high in evaluative threat and complexity intolerance. We sought to call these stereotypes into question by comparing our military and civilian samples.

Method

The pooled correlation analyses (*n* = 511) for the relationships between evaluative threat and complexity intolerance, and the comparisons between samples on these measures, used data from Study 1 (*n* =211), Study 2 (*n* =101) and all participants from Study 3 (*n* = 199). For the training-related analyses, we used all participants who had completed the formative-assessment oriented training (*n* = 124).

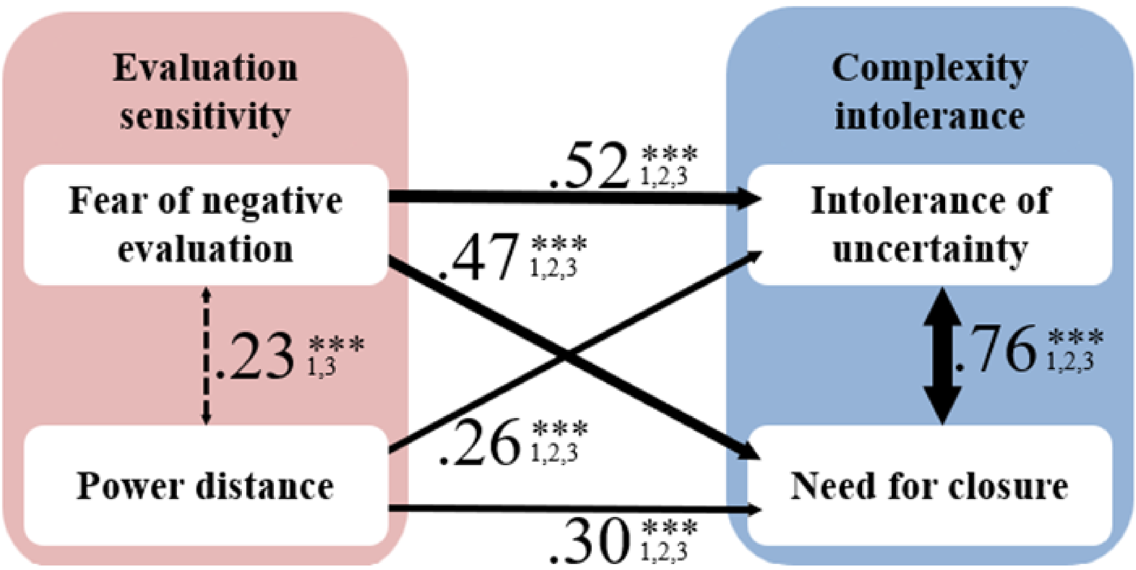
All correlational analyses were conducted as described in Study 1. In addition, we used a Welch's ANOVA (with Bonferroni correction for post hoc pairwise t tests) to compare the values on each variable between military personnel and civilians.

Results

Table 1 and Figure 2 illustrate the pooled correlations.

Figure 2

Correlations from pooled data



Note. Arrows represent weighted associations within and between latent constructs. Dashed arrows highlight pattens of associations that did not replicate across all studies (indexed numbers indicate which study showed a significant relationship). *Indicates correlations with a significant p value < .050, ** < .01, *** <.001.

Comparisons between samples revealed some differences between experienced, early career military and civilian populations (see Table 1). There was a significant difference between groups in measures of FnE, ($F(2, 282.58) = 6.45$, $p < .001$). Post hoc analysis revealed a significant difference between experienced military and civilian samples only ($p <.001$, all other comparisons, p's > .05), with experienced military participants reporting lower FnE than civilians. No significant difference between groups was found for measures of PD, ($F(2, 296.33) = .72$, $p = .488$). Regarding indexes of complexity intolerance, we found significant differences between groups for NfC, ($F(2, 298.31) = 45.42$, $p < .001$). Post hoc analysis revealed that experienced ($p < .001$), and early career military personnel ($p < .001$) reported significantly lower ratings of NfC than civilians, but we found no significant difference between military groups ($p = .50$). Finally, we found significant differences between groups for IoU ($F(2, 299.4) = 49.99$, $p < .001$), with significant difference between both experienced ($p < .001$), early career military ($p < .001$), and civilian samples. Here, military personnel reported lower IoU than civilians. Further, a significant difference between experienced and early career military groupings also emerged ($p = .009$), with experienced military participants reporting higher ratings of NfC than early career personnel.

Table 1

Means, standard deviations, and correlations with bootstrapped confidence intervals

	Study	Mean (SD)	FnE	PD	IoU
1.FnE	1	2.45 (.62)			
	2	2.52 (.52)			
	3	2.66 (.59)			
	P	2.55 (.59)			
2.PD	1	1.62 (.65)	.22***[.07, .35]		
	2	1.55 (.49)	.10 [-.09, .27]		
	3	1.62 (.67)	.30***[.18, .41]		
	P	1.61 (.63)	.23***[.15, .31]		
3.IoU	1	2.26 (.62)	.50***[.39, .60]	.33***[.20, .44]	
	2	2.03 (.48)	.52***[.36, .63]	.05 [-.12, .23]	
	3	2.74 (.76)	.52***[.41, .62]	.29***[.16, .41]	
	P	2.40 (.72)	.52***[.45, .58]	.26***[.17, .35]	
4.NfC	1	2.49 (.64)	.43***[.31, .54]	.33***[.21, .44]	.75***[.68, .80]
	2	2.38 (.47)	.40***[.22,.54]	.23* [.04, .38]	.57***[.43, .67]
	3	2.97 (.67)	.49***[.37, .59]	.31***[.19, .43]	.73***[.65, .78]
	P	2.66 (.67)	.47***[.39, .53]	.30***[.21, .37]	.76***[.72, .79]

Note. *Indicates correlations with a significant p value < .050, ** < .01, *** <.001. BCa bootstrap 95% CIs reported in brackets. Degrees of freedom (n): Study 1 T0= 209 (211), Study 2 T0= 99 (101), Study 3= 197 (199), Pooled (P) = 509 (511). Borders highlight associations within (evaluation sensitivity = solid, complexity intolerance = dashed), and between (dotted) theorised latent constructs.

To test the possibility that a change in evaluation sensitivity leads to reductions in complexity intolerance, we used the pooled difference scores (T1-T0) for all participants who underwent the formative assessment-focused training intervention across Studies 4, and 5 (n = 124). Changes in the FnE scale were significantly positively correlated with changes in both IoU, r (122) = .40, p<. 001, [.17, .60] and NfC, r (122) = .43, p<.001, [.22, .63]. The same pattern emerged with PD where changes in PD predicted changes in IoU, r (122) = .39, p<. 001, [.17, .61] and NfC, r (122) = .44, p<.001, [.23, .63].

Discussion

In the first three studies presented here, we showed that the more an individual is concerned about being evaluated negatively, the greater their complexity intolerance (i.e., the less motivated they are to engage with uncertainty). The last two studies suggest that when training programmes focus on coaching, formative assessment, trial-and-error learning, and problem-solving, they can reduce trainees' complexity intolerance.

We measured evaluation sensitivity using the Fear of Negative Evaluation (i.e., an individual's tolerance for the possibility they may be judged disparagingly by others) and Power Distance (i.e., one's tendency to endorse the legitimacy of hierarchy in maintaining order) scales (Leary, 1983; Yoo et al., 2011). Given the moderate strength of the associations between these scales and the fact that they correlated in two out of the three samples, they appear to be related, yet dissociable, psychological constructs. In general, the more an individual endorses the legitimacy of authoritarian modes of leadership, the greater the fear of interpersonal evaluation.

Regarding complexity intolerance, we found robust and strong correlations across studies between Intolerance of Uncertainty (i.e., negative beliefs about uncertainty and its implications) and Need for Closure (i.e., one's drive to find a concrete answer in ambiguous situations) scales (Carleton et al., 2007; Roets & Van Hiel, 2011). As such, there is little doubt that these constructs are closely related, although one is more focused on valenced responses to uncertainty (e.g., I don't like uncertainty) and the other more on behavioural intentions (e.g., uncertainty impacts my behaviour).

More critically, we found a consistent pattern of associations between evaluation sensitivity and complexity intolerance across all three studies. Pooled analyses across the training-based studies further suggest that reductions in evaluation sensitivity predict reductions in complexity intolerance. Taken together, these findings are consistent with the notion that concern about being evaluated negatively can reduce one's motivation to engage with uncertainty. This observation points to a serious risk in training individuals for critical roles in complex environments. If the stresses of training reduce an individual's willingness to cope with the inherent uncertainties of their work environment, training might undermine the very skillset it is attempting to build.

In all but one of the studies reported here focused on military samples. We did so for a couple of reasons. For one, we felt that the present work is particularly relevant to military populations. Both researchers and individuals within the military have highlighted the considerable challenge and need to train personnel on how to cope with complex environments (Delahaij et al., 2006; Dörner & Funke, 2017; Finn, 2019; Williams, 2010). While operational environments have probably always been complex, it has likely grown in complexity in recent years with the increasing presence of autonomous systems in battlespace and the increasing incidence of "nontraditional warfare" (Delahaij et al., 2006; Simonetti et al., 2020). It is critical that all levels of the military hierarchy have the ability and motivation to deal with complexity under extremely stressful situations so that they can make decisions and take actions that are the best (most reasonable and ethical) in the given circumstances. With all this in mind, effective training for complexity is essential.

The second reason why we focused on military samples was because it seemed likely, a priori, that the evaluation sensitivity would be particularly strong in military training environments, where performance evaluation has serious implications for one's reputation and career. But the fact that these effects emerged in both military and civilian samples suggests that the link between evaluation sensitivity and complexity intolerance is not dependent on the hierarchical nature of the military. Indeed, we found no significant relationship within the civilian sample between working in a hierarchical organization and complexity tolerance. Furthermore, comparisons between samples revealed that the military samples were lower in the Fear of Negative Evaluation, Need for Closure, and Intolerance of Uncertainty. These data clearly contradict stereotypes of military personnel as rigid, discipline-focused, and low in creative thinking that sometimes haunt service people and veterans (Harrell & Berglass, 2012). More generally, the consistency across samples suggests that evaluation elements of training programmes, no matter what the area, might run the risk of reducing one's ability to deal with real world complexity.

Our results suggest that complexity intolerance is not a rigid trait, but can be adapted through training. Much debate can be found in the literature on the feasibility of manipulating dispositional constructs like intolerance of uncertainty (Bavolar et al., 2021; Laposa et al., 2022). Studies 4 and 5 suggest that leadership training approaches can influence complexity tolerance when they involve coaching individuals to cope with complexity and a focus on formative (instead of merely summative) assessment. In Study 4, Individuals became more tolerant of complexity following such a training intervention. Study 5 replicated this effect and demonstrated that while the adapted training approach reduced complexity intolerance (both IoU and NfC), the traditional method of training, built around an authoritarian model of instruction and summative assessment, had no effect. Further, pooled analyses showed that changes in evaluation sensitivity strongly predicted those in complexity intolerance following novel training. These data are consistent with a causal relationship whereby reductions in evaluation sensitivity promote tolerance for complexity.

Together, these data have implications for leadership training. Preparing individuals to deal with dangerous and unpredictable environments is challenging. Traditional approaches, particularly in the military, often aim to expose individuals to repeated summative evaluations that require trainees to closely adhere to decision-making frameworks (Shortland et al., 2018). But these types of approaches may fail to translate from scripted rehearsal within a highly controlled training environment to real world situations that involve more degrees of freedom and real consequences of action (Delahaij et al., 2006; McCall & Laycock, 2024; Van Den Heuvel et al., 2012). To meet this challenge, training approaches might better reflect the dynamic and uncertain nature of complex environments.

The data presented here suggest that training can encourage trainees to embrace uncertainty via mentorship, open discussions about problem solving, and formative assessment that allows for more trial-and-error than summative assessment. Although this research focused on members of the military, the observations likely generalize to other domains. Conversely, other domains provide examples of training approaches that encourage complexity tolerance (Cannon-Bowers & Salas, 1998). For example, medical educators have long struggled with teaching practitioners to cope with making treatment decisions in spite of the inherent uncertainty of disease diagnosis and prognosis. But educators can encourage trainees to deal explicitly with this uncertainty in a given situation and can evaluate them not on correctness of a decision per se, but on their reasoning in coming to that decision (Simpkin & Schwartzstein, 2016). Along these lines, actively setting up diagnostic exercises where errors are likely to occur allows learners to engage in trial-and-error learning and reflect on their decision processes (Eva, 2009). The domain of emergency management provides further insight that may be particularly useful in the military domain. For example, Comfort and colleagues suggest that training can build up from basic skill acquisition, to thinking and discussing the application of those skills in complex scenarios, to simulation-based learning in which trainees face complex simulations in which they must choose when and how to deploy their skillset under stress and uncertainty (Comfort & Wukich, 2013).

Regardless, a fixation on avoiding the negative evaluations of others is not conducive to the types of trial-and-error learning (Eva, 2009) and exploration (Cohen et al., 2007; Mehlhorn et al., 2015) that complex roles require. When leaders fail to grapple with complexity, they run the risk of reaching decisions prematurely, before other alternatives have been considered (Keinan, 1987; Pelaccia et al., 2014). Our data suggest that training interventions that provide formative assessment and mentorship can help limit constraining effects of evaluation concerns so that individuals are more willing (and perhaps able) to deal with complexity.

Chapter Five: General Discussion

Across disciplines, the literature reviewed in Chapter One highlights many psychological mechanisms underlying complex decision-making and how threat may disrupt these processes. This literature underscores the importance of accounting for the nature of threat when interpreting research findings, identifying processes that may underlie decision-making ability, as well as identifying potentially malleable individual differences that could improve performance through targeted training.

Building on these insights, I presented three empirical chapters that examine how integral and evaluative threats influence key behavioural and motivational processes that shape complex decision-making. In this final chapter, I summarise the novel insights into how individuals make complex decisions under threat, as explored throughout this thesis, highlight opportunities to improve real-world application, and outline key directions for future research.

Summary of chapters

The empirical chapters present evidence that being in a threatened state results in a "*profile of effects*" (Broadbent, 1984) that impacts behaviour and motivation when making complex decisions. Together, in eight studies, this body of work has two key objectives: 1) to test the effects of environmental threat on the complex decision-making process and 2) to test the effects of evaluative threat on the motivation to engage in complex decision-making.

I developed a virtual experimental paradigm in Chapter Two where threat was integral to a decision-making task (VRIGT). I first validated the VRIGT, a complex decision-making task embedded in virtual reality. Next, I used the VRIGT to run a series of studies that probed the effect of our threat manipulation on performance. Results demonstrated that threat disrupted performance, learning, reward processing, and the way individuals sampled information from the environment during the task (described in greater detail below).

In Chapter Three, I further probed the effect of learning under threat. However, we used an alternative task that indexed reversal learning ability (VRCF) and adopted a within-subjects research design. This change enabled more precise modelling of distinct dimensions of complexity, such as when sound choices lead to poor outcomes due to stochastic variability, and when sound choices rapidly turn bad, under conditions of volatility. This study used the same integral manipulation of threat as in Chapter Two's studies but integrated haptic feedback to enhance the experience of immersion. I found that individuals adopted a distinct responsive and exploitative learning profile under threat. Using simulations to support, we showed that the impact of these changes on performance varies depending on the level of stochastic variability and volatility in the environment.

In Chapter Four, I investigated the idea that individuals concerned with negative evaluation are less motivated to engage with complexity. Building on interviews we conducted in earlier work with experienced elite military personnel (McCall & Laycock, 2024), I explored the possibility that social evaluation might affect individuals' motivation to engage with complexity. I used established questionnaires to index evaluation sensitivity (i.e., concerned with negative evaluation) and complexity intolerance (i.e., motivational drive to engage with complexity). We collected data from both military and civilian samples. This opportunity to gather data from a military population was particularly fortunate given the unique environment (traditionally hierarchical) in which they work and train. We found that individuals sensitive to negative evaluation were less motivated to engage with uncertainty, but that targeted training can increase this motivation.

Theoretical and methodological contributions

Manipulating integral threat experimentally

The review in Chapter One demonstrated the need for integral experimental manipulations of threat. Experimental control remains, as ever, important in empirical research (Sidman, 1960). But laboratory-based manipulations of threats often differ from the naturalistic contexts in which they naturally arise (Mobbs et al., 2019). The way a stressor is presented to an individual experimentally can affect results (Giovanniello et al., 2023). Integral manipulations of threat, where one's decisions directly affect the level of threat, are required to understand how stressors impact performance in situations where the threat is not just a distraction but goal-relevant information.

In general, integral manipulations are more appropriate to study how a threatened state impacts cognition in critical settings (i.e., for military personnel or first responders), where treating threat as incidental would misrepresent its functional role in the decision-making process. This is because threat is presented as task-relevant information in integral manipulations of threat. This is not the case with incidental manipulations, as in the moment, threat is simply not task-relevant. This changes task dynamics and impacts performance. More specifically, integral manipulations present the threat as a source of information. Indeed, every smash, crash, bang, and creak are all data points that could feasibly help (or hinder) performance. For example, think about the scenario of walking across a frozen lake, where the sound of a crack in the ice necessarily informs choices and impacts behaviour.

In this sense, integral manipulations of threat better replicate the dynamics of a complex environment by maintaining the demand of negotiating multiple competing goals. Demand to negotiate multiple competing goals is a defining characteristic of complexity (Funke, 2012). For example, under integral threat, at any point in time, an individual may be required to move toward a goal that is in conflict (or not) with self-preservation. This requires an individual to strategically allocate resources, dynamically interrogate feedback, and adapt behaviour.

In our first three experiments, we operationalised integral threat by placing individuals in a simulated building that was dangerously collapsing. In these experiments, the goal to "escape the building" was always aligned with the threat. This is in contrast to incidental manipulations of threat, where a threat is operationalised as a distractor and is not conceptually related to the task being performed in any way. For instance, one popular way to study threat and complex decision-making is to get participants to concurrently prepare a presentation (a stressor) while taking part in a complex decision-making task (Simonovic et al., 2018). However, doing poorly on the presentation has no bearing on the complex decision-making performance (and vice versa). Distinguishing between these paradigms helps explain inconsistencies in the literature.

A second novel aspect of the integral threat manipulation we developed is that it allowed temporal alignment between the stressor and task. Traditionally, incidental manipulations of threat principally present participants with a stressor before the task is performed (Wemm & Wulfert, 2017). This is problematic, given that when the task is being performed, participants are not actually under threat. The limitations with this approach are well documented (Ting et al., 2022). Real-world manipulations of threat (e.g., parachute jumping) can address this issue, however, with the cost of giving up experimental control (Idzikowski & Baddeley, 1987).

I demonstrated in Chapter Two that aversive subjective experience, arousal, and real-world dissociation (O'Brien et al., 2018) could be successfully manipulated experimentally using the VRIGT (replicated in Chapter Three with the VRCF). This gave me the opportunity to manipulate threat at the same time as the task, and in such a way that it was conceptually related to the task being performed. By developing and validating an experimental paradigm using virtual reality to manipulate integral threat, I offer a novel experimental paradigm that helps overcome significant challenges that have restricted prior research. That is, it allowed us to observe the effect of threat on complex decision-making performance in ways that have been neglected, such as in situations where threat is present in the moment and is directly related to task performance.

The following sections demonstrate how our research using integral manipulations of threat helps clear up disagreements in the literature and provides empirical grounding for claims about the functional role of threat in shaping complex decision-making performance.

The effect of integral manipulations of threat on complex decision-making performance

In Chapter Two, I demonstrated that integral manipulations of threat negatively impacted complex decision-making performance. Participants under threat performed worse on a complex decision-making task. This was evident in disrupted learning across trials, with individuals under threat taking longer to significantly improve performance from baseline, and scoring worse in the final block of trials, when compared to participants in a nonthreat condition. This adds to the current literature using only incidental methods of manipulating threat when investigating complex decision-making performance (Ben Hassen et al., 2023; Simonovic et al., 2018; Wemm & Wulfert, 2017).

One interpretation of the performance-related findings from Chapter Two is that when individuals are under threat, they are too rigid to respond to changes in the rewards of specific choices. In the context of the IGT, they do not adjust to the new information that the seemingly profitable decks, in fact, come with large losses. Indeed, the ability to respond effectively to change in the environment (i.e., cognitive flexibility) is a fundamental capability underlying complex decision-making performance (Fellows & Farah, 2005; Krems, 1995). To explore this possibility in Chapter Three, I investigated how integral threat impacts cognitive flexibility performance using a reversal learning task. Results suggested no clear effect of threat on cognitive flexibility performance. On the surface, this finding fails to build upon, or even contradict, the findings of Chapter Two.

However, the seeming contradiction only remains when we limit analyses to gross measures of performance. A key challenge when addressing a broad category, such as complex decision-making, is that no single task can practically measure all aspects of behaviour. For example, even the concept of uncertainty can be broken down into many subtypes (Levy & Schiller, 2021, 2023; Simoens et al., 2024; Soltani & Izquierdo, 2019), all of which are likely to impact negotiating complexity in unique ways. Complex decision-making tasks (e.g., the Iowa Gambling Task used in Chapter Two) are developed to be holistic representations of behaviour. Specifically, they use non-binary reward/loss schedules and focus on indexing gradual shifts in behaviour. This makes isolating a discrete response to a specific type of uncertainty difficult.

I argue that in lieu of a grand theory of the impact of threat on cognition, empirical work should aim to reveal the subtle ways that being in a threatened state impacts (or doesn't) the many cognitive processes that underlie decision-making (Giovanniello et al., 2023). This call to systematically study effects across a range of cognitive domains and task contexts is not a new idea; however, it's an important one. Broadbent made a similar comment when discussing drug efficacy back in 1984, "one cannot predict effects … from a single performance test. One must either simulate the practical situation exactly or use a battery of analytical tests to form a profile of effects…" (Broadbent, 1984). His point highlights the need to consider context and use multiple methods to capture the full scope of cognitive impact, especially when effects may vary across tasks.

At its core, decision-making is of course a process of choosing from an array of alternatives. Yet, given the ambiguous, dynamic, and sometimes emotionally challenging realities of "real world" choice behaviour, it's rarely ever just about an isolated selection. These complex decisions require an individual to construct and prune a body of knowledge on which to base action (Behrens et al., 2007) under uncertainty.

With this in mind, in both Chapter Two and Three, I estimated aspects of associative learning using a computational approach. In Chapter Three, I further demonstrated, using parameter estimates and simulations, that threat may have different effects on complex decision-making performance based on the nature of the task's constraints, such as the number of choices presented to an individual, the level of stochastic variability, and rates of volatility. Critically, this helps to explain contradictions in the literature, specifically regarding reward processing and sampling behaviour. Following this theme, Chapters Two and Three present novel theoretical contributions that add to current debates about how threat impacts the processes underlying complex decision-making performance.

Reward processing

Threat has been suggested to impact responsivity to rewards. A popular idea in the decision-making literature is loss aversion under threat. That is, at a basic level, participants are more sensitive to punishment than reward (Kahneman, 1979). This effect has been shown mainly in research examining decisions from description, more than actual experience (Hertwig et al., 2004). However, the idea is still influential, perhaps because it seems logical that a focus on self-preservation could plausibly promote loss aversion.

Recently, work has used computational modelling to suggest threat drives loss aversion in complex decision-making tasks (Ben Hassen et al., 2023; Yu, 2016). Yet, critically, distinguishing between loss aversion and sensitivity to reward is challenging. Is it that participants don't like loss, or do they really like reward? To add clarity here, in Chapter Two (Study Two), I used a computational model with parameters for punishment and reward (Haines et al., 2018). Threat led to an increase in reward sensitivity, not punishment sensitivity. In other words, participants under threat focused on (short-term) rewards.

This finding suggests reward sensitivity, not loss aversion, is responsible for threat-related differences in reward responsivity in complex tasks. These findings contradict some of the other findings in the literature (Berghorst et al., 2013; Bogdan & Pizzagalli, 2006; Morgado et al., 2015). This disparity may be method-induced. Experimental paradigms employing incidental threat manipulations have demonstrated reduced reward sensitivity (Bogdan & Pizzagalli, 2006), a phenomenon ostensibly driven by stress-induced anhedonia, namely, a diminished drive to initiate and maintain engagement with rewarding activities under acute stress. Under incidental conditions, threat is a non-informative distractor, lacking relevance to the task at hand. In such contexts, stress-induced anhedonia may reflect an adaptive response to the experimental framing of threat. In contrast, integral threat manipulations might produce different results because they link reward directly to threat avoidance, making it instrumental to self-preservation. This distinction underscores the importance of considering threat's behavioural impact through the lens of situational dynamics, where the nature of the threat, task constraints, and contextual demands shape behavioural outcomes.

These findings challenge the theoretical position of threat-rigidity, that argues threat drives behavioural inflexibility (Giovanniello et al., 2023; Kamphuis et al., 2008; Staw et al., 1981). Our findings in both Chapters Two and Three suggest that under integral threat, individuals are particularly responsive to recent feedback. This shift to a more responsive learning profile may be advantageous in well-defined environments, allowing individuals to rapidly maximise reward. Yet, when associations between stimulus and reward in the environment are unstable, performance is disrupted. This detriment in performance might look inflexible in tasks such as the IGT where individuals must track long-term gains to perform well. This is not inflexibility per se (at the level of behaviour), but a product of a threat driven change in learning strategy that's not suited to the task's constraints.

Sampling behaviour

Individuals sample information over time when making a series of decisions. Sampling behaviour is a core feature of many mechanisms that underlie complex decision making (e.g., reinforcement learning, exploration and exploitation, context-dependent learning, cognitive flexibility, and motivation). Previous literature has suggested that threat is associated with premature closure (Eva, 2009; Keinan, 1987). That is, under threat, individuals are driven to pick a choice and stick with it, independent of feedback. The idea here is that in threatening situations, errors carry survival costs. In contexts like escaping a collapsing building, minimising uncertainty becomes paramount, such that individuals favour familiar options to reduce risk, prioritising self-preservation over incremental gains (Fanselow, 2018). This exploitative tendency would represent greater apprehension to extend choices in order to minimise uncertainty under threat (Wake et al., 2020). However, evidence of this effect of threat on sampling behaviour is mixed (Frey et al., 2014).

Following an integral manipulation of threat on a complex decision-making task (Chapter Two, Study Two), we found no evidence of premature closure. Rather, under threat individuals were less likely to perseverate on a given choice, resulting in more sampling of information, although this increased sampling did not improve performance. Conversely, in Chapter Three, we found a shift to higher estimates of inverse temperature under threat. This represents an increase in the extent to which threatened individuals exploited their learned values.

To explain this contradiction between my findings, I argue that the impact on sampling behaviour of higher inverse temperature reported depends on the nature of the task (Chou et al., 2024). In well-defined situations, where current understanding accurately reflects the environment, increasing exploitation of learning rates will lead to an increase in repeated choice (premature closure). In contrast, in tasks such as the VRIGT, where the reward associated with different choices changes frequently and dramatically, being more exploitative of recent rewards results in more variable choices over time due to unstable value estimates (as seen in Chapter Two, Study Two). Critically, this accounts for what appears to be increased explorative behaviour without a performance improvement in Chapter Two.

Therefore, I present evidence that the effect of threat on sampling behaviour is adaptive in well-defined environments, given that it helps maximise reward. Yet, in some environments, where choices with short-term gains might result in bad long-term outcomes, a more curious or exploratory approach may be more appropriate. This reduces the risk of overfitting to noise and guards against the impact of updating value estimations too quickly with new feedback (i.e., catastrophic interference).

The effects of evaluation on the motivation to engage with complexity

The motivation to engage in complex decision-making is distinct from the behavioural impacts of threat covered in Chapters Two and Three. Motivation is an antecedent to behaviour. These individual difference measures of motivation, such as intolerance of uncertainty and the need for closure, may help explain variations in overall complex decision-making performance (Kornilov et al., 2015), as well as aspects of underlying process, such as the ability to respond flexibly to changes in the environment (Huang et al., 2017; Williams et al., 2024).

Over three studies reported in Chapter Four, we showed a relationship between evaluation sensitivity and complexity intolerance. Here, in military personnel and civilians, evaluation sensitivity predicted complexity intolerance. This provided empirical evidence that people concerned with negative evaluation were less motivated to engage with complexity.

Although evaluative threat is an unavoidable reality of many situations in which one must make choices with real consequences, the way that threat is internalised may directly impact our behaviour. This is particularly relevant to situations where good decisions necessarily involve a degree of failure. An overemphasis on being judged and getting things "*right* "can discourage people from confronting the ambiguities of real-world situations, where clear-cut answers often don't exist. I suggest that when individuals are overly concerned with being judged negatively, they are likely to resist confronting complex situations.

Critically, by demonstrating that evaluation sensitivity predicts complexity intolerance, we present a mechanistic account of how evaluative threat impacts one's motivational propensity to engage with uncertainty.

Training can increase motivation to deal with complexity

Building on the idea, in Chapter Four, we also demonstrated that the degree to which an individual is motivated to engage with uncertainty is malleable following training. Training is used to prepare individuals to make complex decisions in threatening environments. This training often relies on repeated summative evaluations that enforce strict adherence to predefined decision-making frameworks. This is especially true in military contexts (Delahaij et al., 2006; Männiste et al., 2019; Williams, 2010). Yet such scripted rehearsal within controlled settings may not equip individuals to navigate the dynamic situations they face outside of training. It may lead trainees to be overly concerned with negative evaluation and diminish their complexity tolerance.

While the literature continues to debate the malleability of constructs such as intolerance of uncertainty (Bavolar et al., 2021; Laposa et al., 2022), I present evidence that leadership training can enhance complexity tolerance, particularly when it emphasises coaching for adaptive coping and prioritises formative assessment over rigid summative evaluation.

The extent to which individual traits like complexity tolerance can be “trained” is challenged by some individuals with lived experience. Here, such attributes are not seen as learned but innate characteristics that should be selected in, not cultivated (McCall & Laycock, 2024). Our findings challenge the notion, suggesting complexity intolerance can be shaped through targeted training.

Implications

This thesis offers insights relevant to ongoing academic inquiry and real-world practice. Throughout, I have attempted to coordinate my research with practitioners in mind. Individuals are often required to make decisions under threat. Many organisations look to research to build training programs that aim to help people make tough decisions in dangerous environments, such as in the military (Blacker et al., 2019; Delahaij et al., 2006; Männiste et al., 2019). To ensure the best possible preparation and support for those negotiating hazardous environments, the impact of threat on cognition must be better understood.

Following my interviews with experienced military personnel (McCall & Laycock, 2024), it was apparent that to have a real-world impact, it was not enough to just identify the influence of threat on complex decision-making performance. It was also essential to acknowledge where function is likely influenced by the nature of the threat and task-contextual constraints and focus on factors that are plausibly malleable following education and/or training. Below, I highlight specific examples of how this aim might be achieved.

Moving beyond incidental exposure to threat

As informative as incidental manipulation of threat has been in modelling the relationship between threat and decision-making, they are limited in terms of inferential scope. This restricts their ability to inform researchers and practitioners of the effects of threat on behaviour in some applied contexts. The development of a method of manipulating integral threat gives us the opportunity to move closer to addressing this issue.

The studies presented in this thesis suggest that disparities in prior findings may be method induced. As highlighted in Chapter One, this may (in part) be explained by the nature of the method used to manipulate threat and the task used to index complex decision-making. This is critical because it suggests that widely held assumptions about threat and decision-making may rest on methodological artefacts rather than genuine psychological mechanisms. To offer clarity, research must confront this possibility directly, ensuring that experimental design does not obscure or distort the phenomena under investigation.

But it's not just about clarifying how manipulating threat integrally impacts the results of previously studied paradigms. By manipulating threat integrally, more ecologically valid models of decision-making under threat can be operationalised, allowing researchers to capture the interplay between evolving threat cues and shifting reward contingencies in dynamic contexts. Of course, multi-arm bandit tasks like the IGT (Bechara et al., 1994) capture aspects of this conflict. Notably, the IGT presents participants with options that are highly rewarded in terms of magnitude but risky in terms of punishment (e.g., deck B). Yet, critically, reward (e.g., moving closer to a predefined goal) is presented to an individual as a fixed value statement, like a price tag in a food market. In contrast, in real-world situations, individuals need to work harder to establish a subjective value based on estimates that are not easily quantifiable (Xu et al., 2024).

Using integral manipulations of threat, it's feasible to develop paradigms where different aspects of threatening event (e.g., smoke billowing from behind a door) are paired with specific rewards and/or punishments in the environment, as a task unfolds. For example, options that are both rewarding and threatening could be presented to individuals. Further, these associations could change randomly over time or based on the frequency of selection by the individual. This is distinct from the approach taken in our paradigms, where threat was ever present, and doing well on the task was always associated with a decrease in threat (e.g., escaping the building).

Psychoeducation for relevant practitioners

Psychological research has a great deal of promise in helping address real-world problems faced by society, but to do so, it must look to develop closer ties with relevant practitioners (Prather et al., 2022). One specific application with reference to this thesis is to prepare individuals better to make complex decisions under threat. Promoting the view that the nature of a threat, behavioural response, and context interact to produce variable outcomes. It's convenient (and valuable) to think about threat simply in terms of escaping predators (Mobbs et al., 2009). Perhaps, in the case of incidental threats, in some situations it is also important to ignore them, but what about when escape is not an option? Many practitioners, such as first responders and military personnel, must run into hazard, not away.

The work presented helps bridge the gap between theory and application. We have built on lived experience (McCall & Laycock, 2024) to develop experimental paradigms that reflect exposure to complexity under threat. We have focused on highlighting the impact of threat on processes that underlie complex decision-making performance, not necessarily the outcome. Critically, we highlight how the effect of threat may lead to maladaptive and/or adaptive performance depending on environmental constraints. This results in a body of work that has application in applied settings, specifically, in the form of psychoeducation delivered to individuals before being asked to perform in dangerous and uncertain situations.

What might this look like in practice? A starting point is to build around a key argument of this thesis: that threat should not always be dismissed as a mere distraction, but instead recognised as a source of information that can scaffold performance. Introducing the mechanisms of behavioural sampling (e.g., the distinction between exploitation and exploration of information) and reward responsiveness gives a nuanced understanding of how threat might influence complex decision-making. It's not necessarily about learning new skills, but how old skills developed in training might need refining when actioned in real-world situations under the influence of threat. In other words, delivering a comprehensive psychoeducational package of instruction aimed at translating psychological insights into practical understanding that's directed at the individuals who need it most.

Evidencing training approaches

In Chapter Four, we had the opportunity to sample from a population of military personnel. This allowed us to observe the impact of a novel training program developed to help practitioners make challenging decisions in hazardous situations. Although the primary objective was to test the effects of evaluative threat on the motivation to engage in complex decision-making, this work also has broader implications. The research project was developed with application in mind. That is, it was created in part to inform training staff and practitioners about psychological constructs that underlie the skills they are looking to promote in training and provide an evidence base on which to critique training methods employed and aid in developing new ones.

We made some important accommodations intended to boost impact. We also worked hard to replicate findings over three independent populations, two of which were longitudinal, requiring several years of data collection. Further, methods of analysis throughout the project were selected to give stakeholders an interpretable statistic (training-related change), as opposed to one that required some level of abstraction (e.g., mixed effect modelling). In this sense, we selected methods that were intuitive to interpret and in line with best practice for working with stakeholders in participatory research (Duea et al., 2022).

Training and evaluation approaches

Chapter Four also provides evidence in support of a shift from summative assessment to formative assessment in leadership training. Specifically, we support the claim that evaluation elements of training programmes might run the risk of reducing one's ability to deal with real-world complexity. It also offers suggestions of alternatives to traditional modes of training that promote experiential diversity, e.g., exposing trainees to novel situations (Harris et al., 2017; McCall & Laycock, 2024) that elicit error to encourage trial-and-error based learning (Eva, 2009). In principle, this work has the potential to support trainers and practitioners in grasping aspects of psychology behind the skills they're trying to teach. It could also stimulate discussion and reflection about the difficulties of negotiating complex and dangerous situations.

Limitations

In general, this thesis has several limitations. One is the capacity of virtual reality to simulate real-world threatening situations. Indeed, individuals must always, on some level, know the experience is not real (Pan & Hamilton, 2018).

Nevertheless, virtual environments can promote physiological, psychological, and behavioural responses that are similar to responses in actual threatening situations (Baker et al., 2020; Finseth et al., 2022; McCall et al., 2015, 2022). Along these lines, we demonstrated an increase in subjective aversion and arousal in the threatening conditions in Chapters Two and Three. We also demonstrated an increase in physiological arousal in the threatening condition in Chapter Two.

Furthermore, threatening experiences can even feel "unreal" in the real world. The reports from military personnel dealing with threat in the real world often describe the experience as surreal or dreamlike, with the gravity of the event only becoming apparent much later (McCall & Laycock, 2024). Regardless, while virtual paradigms offer controlled manipulations of threat, individuals remain removed from real danger.

In Chapter Four, I investigated the relationship between individual differences in evaluation sensitivity and complexity intolerance. I suggest that individual differences in the motivation to engage with uncertainty may help explain variation in complex decision-making performance. However, I do not provide empirical evidence that these individual differences predict performance.

We used the existing literature to support the claim that individual differences in motivation to engage with uncertainty may help affect complex decision-making performance. Critically, both measures of complexity intolerance used in Chapter Four account for individual variability in processes that underlie complex decision-making performance. For example, individual differences in motivation predict reward sensitivity (Williams et al., 2024), being risk-averse (Carleton et al., 2016), repetitive thinking (McEvoy & Erceg-Hurn, 2016), reduced confidence in decisions, being behaviourally inhibited in unpredictable situations (Jensen et al., 2014), decisiveness (Kruglanski, 2013), and resistance to change (Roets et al., 2015).

However, these studies provide little direct evidence that, under threat, complexity intolerance affects complex decision-making, given that in this work, threat is either not manipulated or is incidental (Kornilov et al., 2015). Although we did exploratory analyses to address this issue and investigate the relationship between complexity intolerance and complex decision-making performance in secondary exploratory analyses (see Supplementary Materials: Chapter Four), we were likely underpowered.

Moving forward, the effect of individual differences in motivation to engage with uncertainty on performance warrants further investigation with an appropriately powered sample. This is especially true given Chapter Four's finding that complexity intolerance is malleable following extensive formative training. In other words, it would help broaden the claim that individuals concerned with negative evaluation are less motivated to engage with complexity; it would suggest they are also better able to deal with it.

Future research

The following subsections detail areas for future research. These suggestions are aimed at broadening the methodological scope and interpretive reach of our reported findings by addressing key aspects of complex decision-making not covered in this thesis.

Is a (complex) problem shared, a problem halved? Group-based decision making under threat.

This thesis primarily examines complex decision-making at the individual level. While Chapter Four considers socially evaluative contexts, findings are interpreted through the lens of individual experience and response. Yet, decisions are rarely made in true social isolation. A critical next step is to explore how threat modulates decision-making differently when operating alone versus within a group context. By dissecting the mechanisms of individual decision-making under threat, we gain the necessary tools to interpret broader collective behaviour with greater precision. Group dynamics may amplify or constrain the processes that underlie complex decision making, and perhaps even replace some.

Despite the inherent resource-related challenges with researching group performance (Cannon-Bowers & Salas, 1998), work in this area is extensive. Factors such as intelligence (Jones, 2008) and cooperation (Okada, 2020) have long been used to explain aspects of group-based decision making, yet increasingly, task complexity is being acknowledged (Almaatouq et al., 2021; Reverberi et al., 2022). Thinking about complexity at the group level raises a host of interesting questions, given that uncertain tasks and informational constraints likely impact social dynamics on many levels. However, literature concerning the effects of threat on group-based decision-making is harder to find. Yet, the work that has been done tends to argue that threat impacts how information is shared (Driskell & Salas, 1991) and acted on (Kamphuis et al., 2011) during team-based interaction. However, work investigating the effect of threat on prosocial behaviour highlights that behaviour is sensitive to the needs of a specific social situation (Faber & Häusser, 2022).

This last point is key, as it reaffirms the common theme in this thesis, that contextual demands moderate the impact of threat on complex decision-making. Moving forward, methods developed and used in Chapters Two and Three could be adapted to focus on group-based decision-making. This could be achieved by leveraging multi-user networked virtual environments (Moser et al., 2020), which allow threat to be manipulated during collaborative tasks. This change will permit researchers to index the processes that are driving performance at both the individual and collective levels. This scalable framework will allow for data to be gathered on team communication, a vital aspect of coordination at the core of group-level problem-solving (Cannon-Bowers & Salas, 1998). This will allow a deeper understanding of how people make decisions, not just on their own, but together. By studying how integral threat affects individuals, we build the groundwork to explore how group behaviour shifts, adapts, or even breaks down when danger hits.

Decision-making beyond the data: the impact on internally generated processes under threat.

Despite integrating a broad range of behavioural and motivational psychological mechanisms when evaluating complex decision-making performance, many important processes remain absent from this thesis. I focused on psychological mechanisms that were computationally tractable and experimentally convenient. But other psychological dynamics are also involved in complex decision-making. For example, consider how we anticipate emotions. This process, known as affective forecasting (Pilin, 2021), often involves misjudgements about future emotional states. These misjudgements can distort long-term planning, valuation, and goal prioritisation. Another important factor is how we evaluate our own certainty. This relates to metacognitive abilities (Dunlosky & Metcalfe, 2008), the capacity to monitor and adjust one's own confidence. Such abilities shape learning, adaptability, and error correction. These mechanisms are central to how decisions unfold in real-world contexts, especially under uncertainty, over time, and in socially embedded environments.

The impact of threat on prospection is particularly relevant. Prospection is the ability to imagine future scenarios to guide decisions (Gilbert & Wilson, 2007). It draws on memory and executive systems to support planning and evaluate outcomes beyond experience. By generating novel possibilities, prospection enables flexible, goal-directed behaviour under uncertainty. It's not just imagination, it's a tool for intentional action. Police and military personnel report mentally simulating plausible, context-specific future states, selecting adaptive responses, and considering alternative courses of action when dealing with challenging situations (Harris et al., 2017; McCall & Laycock, 2024). Critically, they are going beyond the data using hypothetical mental simulations to help scaffold decision-making under ill-defined constraints.

Research investigating the impact of threat on prospection is limited. However, similar mechanisms are suggested to be negatively impacted by threat. For example, incidental manipulation of threat stress reduces cognitive reflection (Simonovic et al., 2017). This effect was proposed as the mechanism driving impaired complex decision-making performance over time. Such a mechanism may feasibly impact prospection and inhibit performance. Future research should incorporate data on internally generated processes such as prospection, affective forecasting, and metacognition, allowing for more comprehensive accounts of psychological function that extend beyond the scope of this thesis in understanding how threat influences complex decision-making.

Conclusion

This thesis explored the effect of threat on complex decision-making. Dealing with complexity under threat is difficult. Understanding how threatened states impact complex decision-making performance necessarily requires meaningful operationalisations of threat and task constraints that accurately model the conditions under which real-world decisions unfold.

The work in this thesis models the nuance of human cognition under duress, where motivation, performance, self-preservation, and learning converge. Empirically, I focus on manipulations of integral, environmental threat and socially evaluative threat. These two specific types of threat have been neglected in the existing complex decision-making literature. More importantly, lived experience suggests these threats lie at the heart of how people navigate uncertainty in hazardous environments (McCall & Laycock, 2024).

I had two central objectives: to test the effects of environmental threat on the complex decision-making process and the effects of evaluative threat on the motivation to engage in complex decision-making. I presented three empirical chapters, including eight studies, that meet these aims. The work views threatening emotional experience and evaluative pressure as structurally embedded in the decision-making process, not incidental distractions but functional constraints.

Being under threat impacts reward processing and sampling behaviour in ways that, on complex tasks, vary as a function of context. Learning is altered, as individuals become more responsive and exploitive. I demonstrate that the way these effects impact performance can differ depending on task constraints. Evaluative threat can also affect motivational processes that underlie complex decision-making. However, this may be mitigated by formative training, which may promote motivation to engage with uncertainty.

Identifying the psychological mechanisms disrupted by threat enables more accurate predictions of how decision-makers navigate complexity and danger in real-world situations. These findings, furthermore, help explain current contradictions in the literature and could offer a foundation for future work developing training and educational programs that better equip individuals who must make difficult decisions within hazardous environments.

Appendices

Supplementary Materials: Chapter Two

Study 1

Computational modelling

Supplementary Table 1

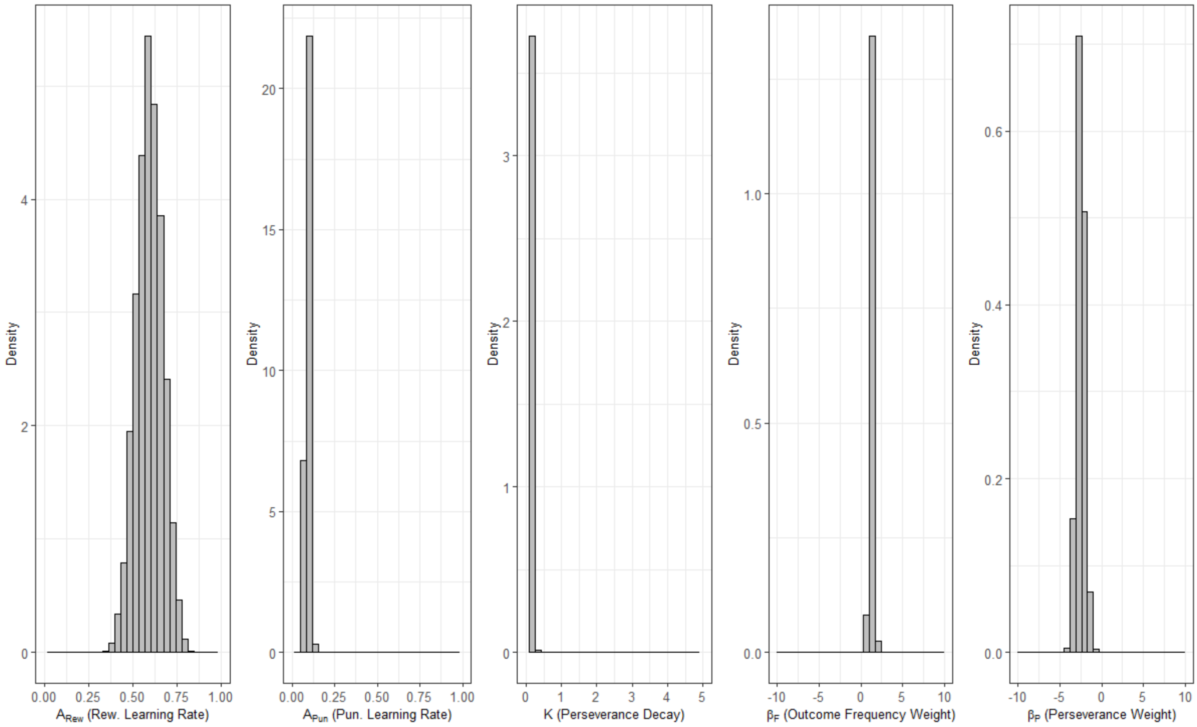
Model fits

Model	LOOIC
Prospect-Learning Valence Delta	12230.95
Prospect-Learning Valence Decay	11808.59
Value-Plus-Perseverance	11214.69
Outcome-representation learning	11173.73

Note. LOOIC, Leave-one-out information criterion. ORL was the best fitting model in Study 1 using the LOOIC.

Supplementary Figure 1

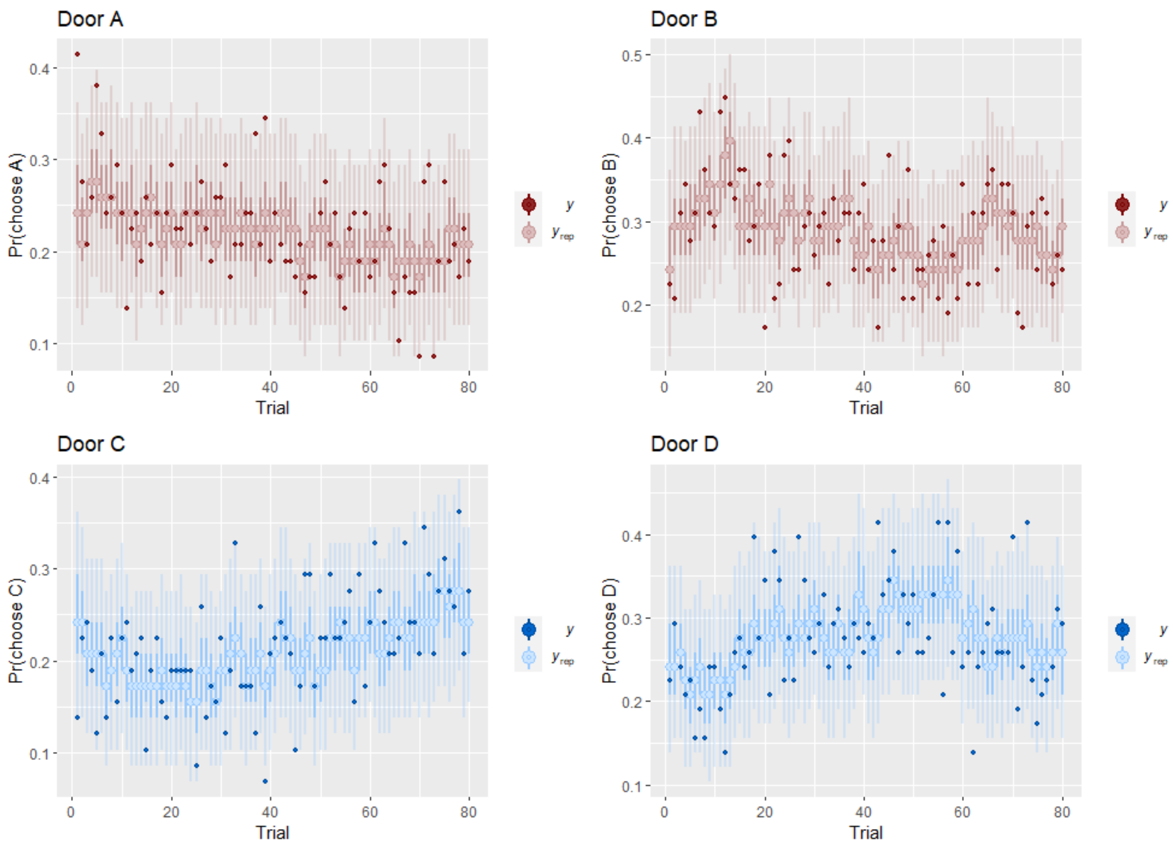
Posterior distributions of the hyper (group) parameters



Note. Arew (Reward sensitivity), Apun (Loss sensitivity), K (Forgetfulness), betaF (Frequency sensitivity), betaP (Choice perseveration).

Supplementary Figure 2

Posterior predictive checks



Note. Observed versus the posterior predicted probability of choosing each deck over trials. To illustrate performance trends “disadvantageous” doors are presented as red and “advantageous” doors in blue. The x-axis represents trial number, and y-axis the proportion of participants who selected the deck on each trial (Y), along with predictions (Yrep) computed from draws from the posterior predictive distribution of the same values. 50% (dark) and 95% (light) prediction intervals illustrate uncertainty.

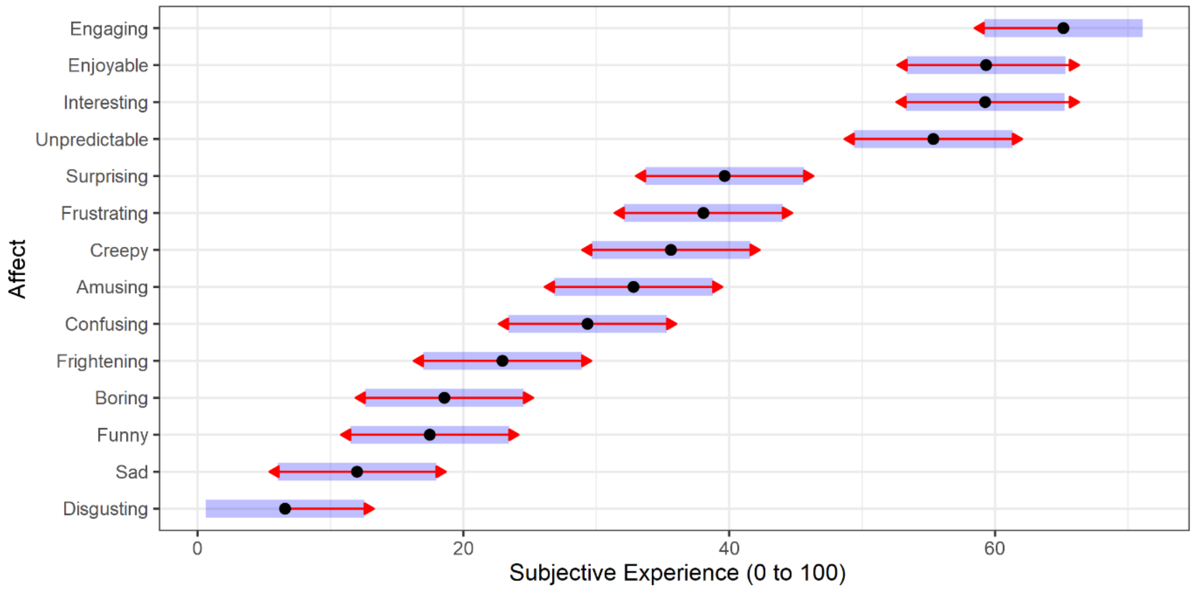
Subjective Experience

Affect

We used a LMM to analyse the relationship between the affect terms participants used to report their overall feeling during the VRIGT. The model predicted rating with a fixed effect of affect category (e.g., "frightening"). Intercepts were allowed to vary as a random factor at the level of the individual. The affect category with the lowest average rating score ("disgusting") was used as a reference level. A significant effect of affect category was found ($F(13, 897) = 45.21$, $p < .001$). Pairwise comparisons are reported using the Tukey correction for p-values highlighted significant contrasts between affect categories. Results displayed in Figure 1.

Supplementary Figure 3

Ratings of affect



Note. Measures of affect (14 categories, scaled from 0 to 100) was used from all 70 participants that completed the VRIGT. Means (black dots), CIs (purple line) for ratings of the overall experience of the VRIGT. The red arrows highlight the comparisons between means. Overlapping arrows between two means indicate a difference is not significant based on the adjusted p-values.

User Experience Scale

The User Engagement Scale (UES) was developed to measure self-reported user engagement (O'Brien et al., 2018). A sample item from this scale was "I lost myself in this experience". Responses were recorded on a 5-point Likert scale (from 1 = not at all characteristic of my experience to 5 = entirely characteristic of my experience). Higher scores indicate greater focused attention during the VRIGT.

The User Engagement Scale (UES) was broken down into four subsections to isolate the components of: focused attention (FA), perceived usability (PU), aesthetic appeal (AA), and reward factor (RF). Study 1: UES (M =3.53, SD = .51), FA (M =3.43, SD = .80), PU (M =3.77, SD = .91), AA (M =3.10, SD = .95), and RF (M =3.78, SD = .68).

Each of these averaged reported scores was above the threshold of 3 (scaling midpoint," neither agree or disagree"), suggesting the VRIGT was generally considered engaging. Regarding focused attention specifically, this suggests some degree of immersion was experienced by participants.

Tension

We also asked participants to select a graphic that best illustrated the degree and pattern of tension (a qualitative category selection) experienced during the task. Four options where visually presented to participants, depicting tension experienced during the VRIGT.

Here, results demonstrate, n= 4 [5.71%] participants reported the tension as flat (with no change over time), n= 8 [11.42%] reported tension increasing over time, n= 6 [8.57%] decreased over time, and n= 52 [74.28%] that tension dynamically changed over time. A chi-square goodness of fit statistical test demonstrated this final category significantly accounted for most individuals' reported account of tension $\chi 2$ (3) = 91.14, p <.001, w = 1.14.

Performance

Supplementary Table 2

LMM Results: Performance over time

Predictors	Estimates	CI	p
(Intercept)	−3.03	−4.64 – −1.43	<.001
Block 2	1.72	−0.22 – 3.67	.082
Block 3	4.59	2.64 – 6.53	<.001
Block 4	4.03	2.09 – 5.98	<.001
Random Effects			
σ^2	28.25		
τ_{00} ppt	10.42		
ICC	0.27		
N ppt	58		
Observations	232		
Marginal R^2 / Conditional R^2	.081 / .328		

Study 2

Videos

Demonstration of the VRIGT used in Study 2 available on OSF.

Power analysis

The data we collected in Study 1 was used to estimate the sample size required via simulations (Kumle et al., 2021), using the SIMR package in R (Green & MacLeod, 2016) . The results of 1000 simulated samples suggested that using a linear mixed effect model with VRIGT score being predicted by blocks, with a random factor on participant (intercepts only) a n = 50 was required to achieve a power of .80. As this study requires the inclusion of second condition, in line with the suggestion of (Brysbaert & Stevens, 2018) this number was doubled.

Pre-registration

Link: https://aspredicted.org/PT3_L9V

Pre-scene instructional scripts

Threat script

"You are in a building that is slowly collapsing.

Your goal is to get as far away from the danger zone as possible.

In each room, there are four doors to choose from. Selecting a door might lead you to a room that is further from the danger zone and help you escape. However, some doors will lead you to rooms where the floor is weaker, leading the danger zone to spread in your direction and cancelling any gains in distance made.

You are being tracked so you can always view your distance from the danger zone. This means you can attempt to choose the door that best maximises your distance from the danger zone.

I won't tell you how long the task will take. You must keep choosing doors until the computer stops. The computer does not make you lose distance at random. All I can say is that you may find yourself losing distance on all of the doors, but some doors will make you lose more distance than others. You can escape if you stay away from the worst doors.

Now get ready to start your escape. Good luck. "

Nonthreat script

“You are in a building attempting to find the exit so you can meet friends in the outside car park.

Your goal is to make your way as far away from your current position as possible and find the exit.

In each room, there are four doors to choose from. Selecting a door might lead you to a room that is further from your starting position and closer to the exit. However, some doors will lead you away from your starting point, but also further from the exit, cancelling any gains in distance made.

You are being tracked so you can always view your distance from your starting point. This means you can attempt to choose the door that best maximises your distance from your start location.

I won't tell you how long the task will take, but they is no need to rush. You must keep choosing doors until the computer stops. The computer does not make you lose distance at random. All I can say is that you may find yourself losing distance on all of the doors, but some doors will make you lose more distance than others. You will find the exit if you stay away from the worst doors.

Now get ready to find the exit. Good luck. “

Subjective Experience

Supplementary Table 3

LMM Results: Affect between condition

Predictors	*Estimates*	*CI*	*p*
(Intercept)	8.64	1.88 – 15.40	**.012**
affect [Creepy]	8.46	−0.54 – 17.46	.065
affect [Frustrating]	33.22	24.22 – 42.22	**<.001**
affect [Unpredictable]	42.26	33.26 – 51.26	**<.001**
affect [Engaging]	41.72	32.72 – 50.72	**<.001**
affect [Confusing]	29.46	20.46 – 38.46	**<.001**
affect [Enjoyable]	44.52	35.52 – 53.52	**<.001**
affect [Boring]	21.16	12.16 – 30.16	**<.001**
affect [Funny]	9.24	0.24 – 18.24	**.044**
affect [Amusing]	21.24	12.24 – 30.24	**<.001**
affect [Sad]	−5.60	−14.60 – 3.40	.223
affect [Disgusting]	−4.38	−13.38 – 4.62	.340
affect [Interesting]	40.50	31.50 – 49.50	**<.001**
affect [Surprising]	18.70	9.70 – 27.70	**<.001**
Condition	36.34	26.78 – 45.90	**<.001**
affect [Creepy] * Condition	−4.64	−17.37 – 8.09	.475
affect [Frustrating] * Condition	−28.34	−41.07 – −15.61	**<.001**
affect [Unpredictable] * Condition	−23.62	−36.35 – −10.89	**<.001**
affect [Engaging] * Condition	−22.92	−35.65 – −10.19	**<.001**
affect [Confusing] * Condition	−30.72	−43.45 – −17.99	**<.001**
affect [Enjoyable] * Condition	−33.78	−46.51 – −21.05	**<.001**
affect [Boring] * Condition	−45.36	−58.09 – −32.63	**<.001**

affect [Funny] * Condition	−42.26	−54.99 – −29.53	**<.001**
affect [Amusing] * Condition	−34.02	−46.75 – −21.29	**<.001**
affect [Sad] * Condition	−26.32	−39.05 – −13.59	**<.001**
affect [Disgusting] * Condition	−33.86	−46.59 – −21.13	**<.001**
affect [Interesting] * Condition	−25.22	−37.95 – −12.49	**<.001**
affect [Surprising] * Condition	−19.68	−32.41 – −6.95	**.002**
Random Effects			
σ^2	526.59		
$\tau_{00\ PPT}$	66.64		
ICC	0.11		
N PPT	100		
Observations	1400		
Marginal R^2 / Conditional R^2	.373 / .443		

Supplementary Table 4

Pairwise comparisons: Affect between conditions

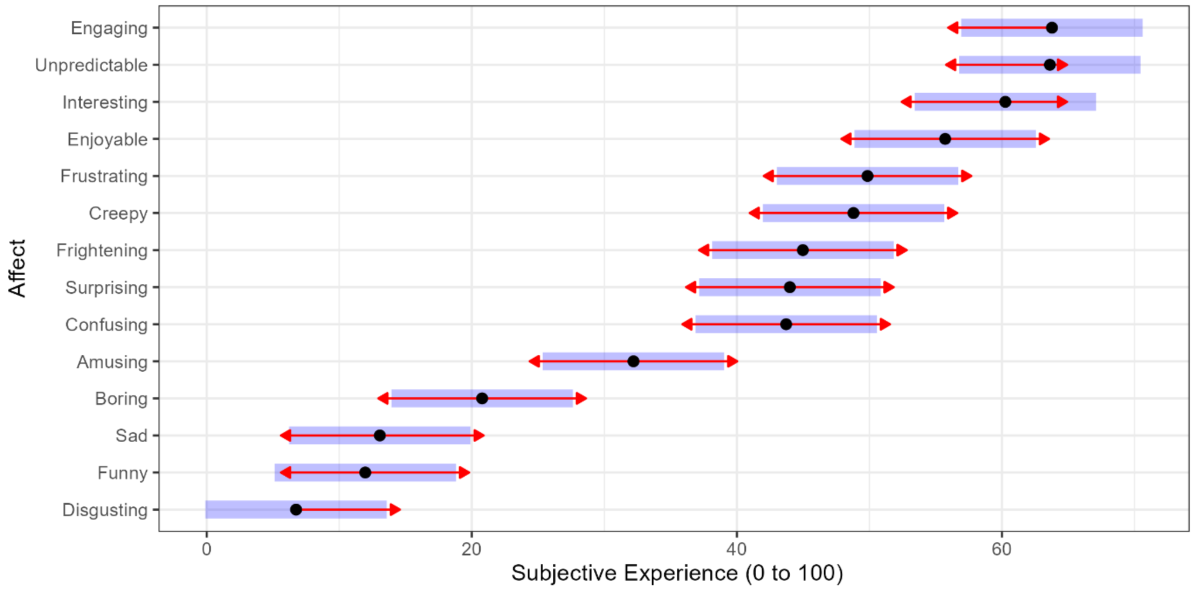
Affect	Estimate	DF	t.ratio	P
Frightening	**−36.34**	**1178.667**	**−7.46007**	**<.001**
Creepy	**−31.7**	**1178.667**	**−6.50755**	**<.001**
Frustrating	−8	1178.667	−1.64228	.758
Unpredictable	−12.72	1178.667	−2.61123	.118
Engaging	−13.42	1178.667	−2.75493	.079
Confusing	−5.62	1178.667	−1.1537	.979
Enjoyable	−2.56	1178.667	−0.52553	.999
Boring	9.02	1178.667	1.851675	.590
Funny	5.92	1178.667	1.21529	.967
Amusing	−2.32	1178.667	−0.47626	.999
Sad	−10.02	1178.667	−2.05696	.422
Disgusting	−2.48	1178.667	−0.50911	.999
Interesting	−11.12	1178.667	−2.28277	.267
Surprising	**−16.66**	**1178.667**	**−3.42006**	**.009**

Note. The MVT correction was applied to adjust for multivariate comparisons.

Measures of affect (14 categories, scaled from 0 to 100) were used from all participants that completed the VRIGT in the threat condition. Means (black dots), CIs (purple line) for ratings of the overall experience of the VRIGT. A LMM was used to analyse the relationship between the measures of affect used by participants to report their overall feeling during the VRIGT. The model predicted rating with a fixed effect of affect category (e.g., "frightening"). Intercepts were allowed to vary as a random factor at the level of the individual. The affect category with the lowest average rating score ("disgusting") was used as a reference level. A significant effect of affect category was found ($F(13, 637) = 35.93$, $p < .001$).

Supplementary Figure 4

Ratings of affect

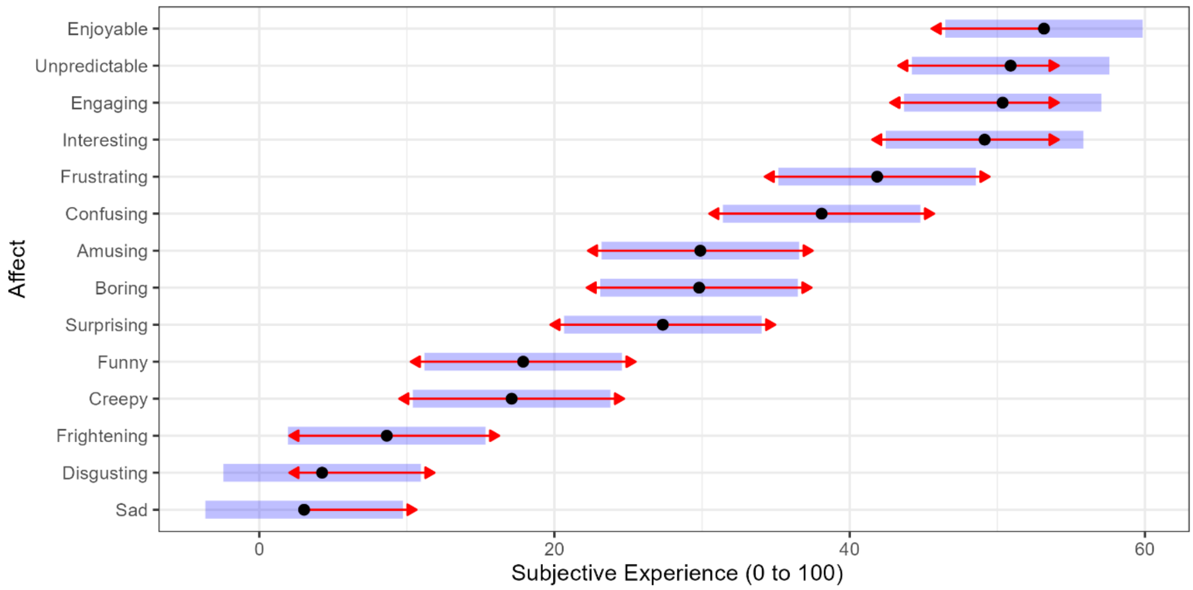


Note. Pairwise comparisons are reported using the Tukey correction for p-values highlighted significant contrasts between affect categories. The red arrows highlight the comparisons between means. Overlapping arrows between two means indicate a difference is not significant based on the adjusted p-values.

Measures of affect (14 categories, scaled from 0 to 100) were used from all participants that completed the VRIGT in the nonthreat condition. Means (black dots), CIs (purple line) for ratings of the overall experience of the VRIGT. A LMM was used to analyse the relationship between the measures of affect used by participants to report their overall feeling during the VRIGT. The model predicted rating with a fixed effect of affect category (e.g., "frightening"). Intercepts were allowed to vary as a random factor at the level of the individual. The affect category with the lowest average rating score ("sad") was used as a reference level. A significant effect of affect category was found ($F(13, 637) = 31.02$, $p < .001$).

Supplementary Figure 5

Ratings of affect



Note. Pairwise comparisons are reported using the Tukey correction for p-values highlighted significant contrasts between affect categories. The red arrows highlight the comparisons between means. Overlapping arrows between two means indicate a difference is not significant based on the adjusted p-values.

User Experience Scale

In Study 2, we only collected data on focused attention using the UES: FA, threat ($M = 3.71$, $SD = .75$), nonthreat ($M=3.62$, $SD = .71$). Each of these averaged reported scores were above the threshold of 3 (scaling midpoint," neither agree or disagree"), suggesting some degree of immersion was experienced by participants. However, we found no significant difference between conditions, $t(98) = -.594$, $p = .554$.

Tension

In the threat condition, n= 2 [4%] participants reported the tension as flat (with no change over time), n= 1 [2%] reported tension increasing over time, n= 14 [28%] decreased over time, and n= 33 [66%] that tension dynamically changed over time. While, In the nonthreat condition, n= 11 [22%] participants reported the tension as flat (with no change over time), n= 5 [10%] reported tension increasing over time, n= 16 [32%] decreased over time, and n= 18 [36%] that tension dynamically changed over time.

Performance

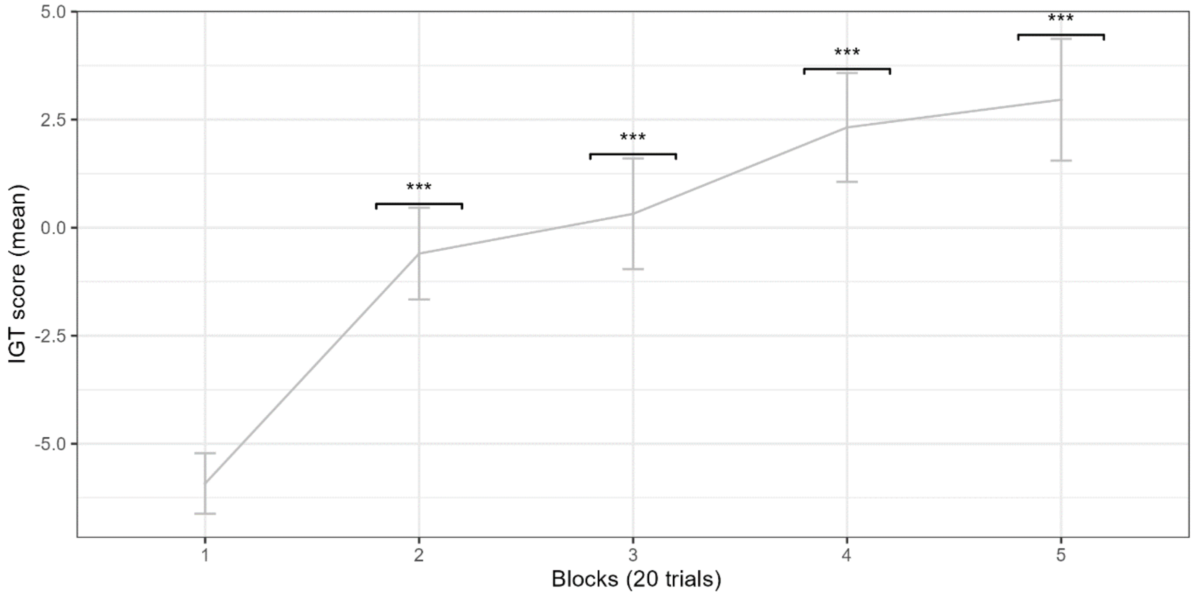
Supplementary Table 5

LMM Results: Performance over time by condition

Predictors	Estimates	CI	p
(Intercept)	−5.92	−8.23 – −3.61	<.001
Blocks 2	5.32	2.65 – 7.99	<.001
Blocks 3	6.24	3.57 – 8.91	<.001
Blocks 4	8.24	5.57 – 10.91	<.001
Blocks 5	8.88	6.21 – 11.55	<.001
Condition	0.80	−2.46 – 4.06	.630
Blocks 2 * Condition	−2.76	−6.54 – 1.02	.152
Blocks 3 * Condition	−5.52	−9.30 – −1.74	.004
Blocks 4 * Condition	−3.08	−6.86 – 0.70	.110
Blocks 5 * Condition	−4.76	−8.54 – −0.98	.014
Random Effects			
σ^2	46.29		
τ_{00} PPT	22.53		
ICC	.33		
N PPT	100		
Observations	500		
Marginal R^2 / Conditional R^2	.108 / .400		

Supplementary Figure 6

Nonthreat condition, Performance score over time



Note. Performance score over time in the nonthreat condition. Note. Error bars represent –/+ standard error. *Indicates differences from baseline from with a significant p value *< .050, ** < .01, *** <.001.

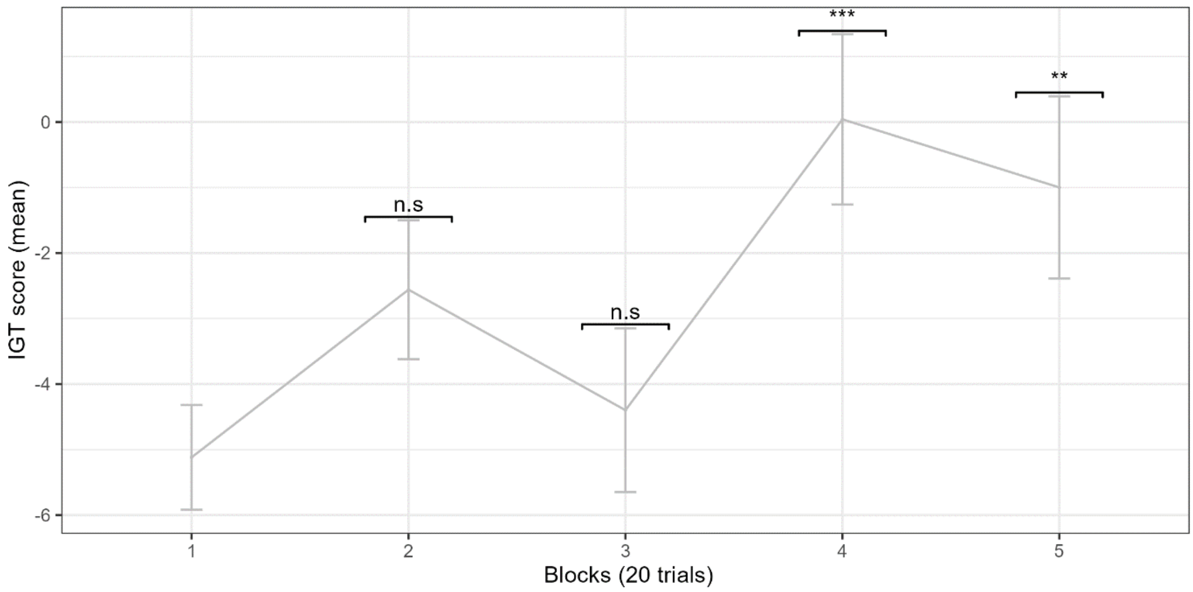
Supplementary Table 6

LMM Results: Performance over time in nonthreatening condition

Predictors	Estimates	CI	p
(Intercept)	−5.92	−8.22 – −3.62	<.001
Blocks 2	5.32	2.59 – 8.05	<.001
Blocks 3	6.24	3.51 – 8.97	<.001
Blocks 4	8.24	5.51 – 10.97	<.001
Blocks 5	8.88	6.15 – 11.61	<.001
Random Effects			
σ2	48.06		
τ00 PPT	20.02		
ICC	.29		
N PPT	50		
Observations	250		
Marginal R2 / Conditional R2	.127 / .384		

Supplementary Figure 7

Threat condition, Performance score over time



Note. Performance score over time in the threat condition. Error bars represent −/+ standard error. *Indicates differences from baseline from with a significant p value *< .050, ** < .01, *** <.001.

Supplementary Table 7

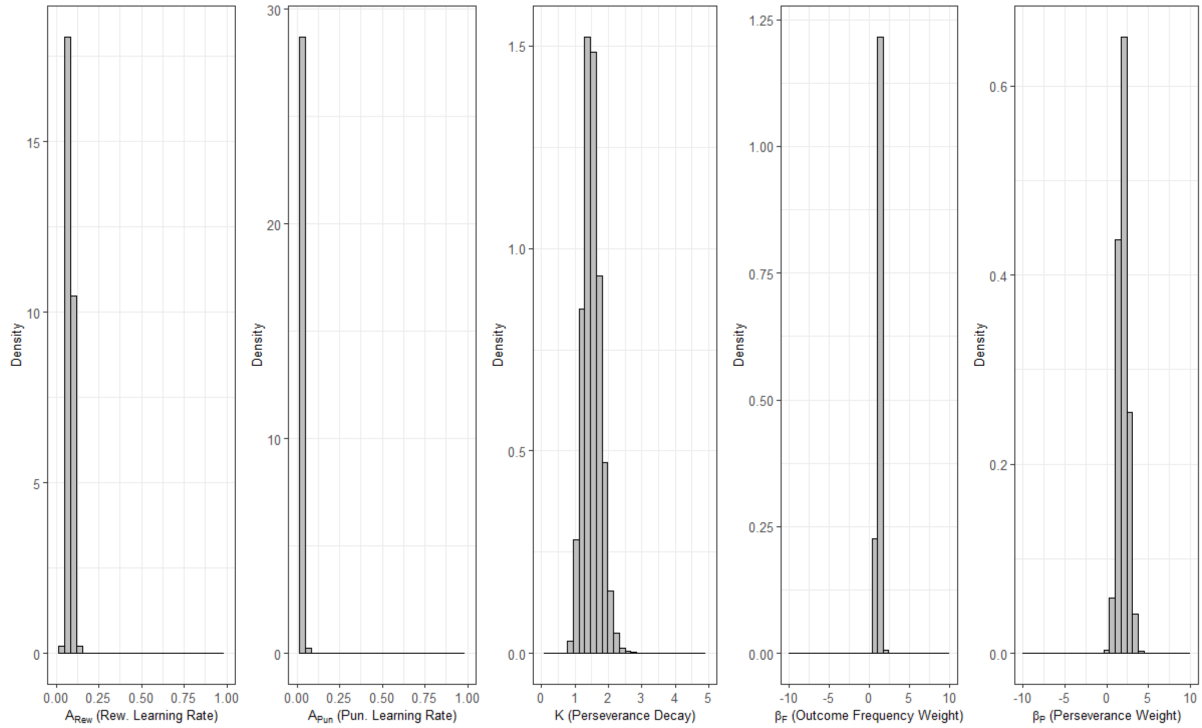
LMM Results: Performance over time in threatening condition

Predictors	Estimates	CI	p
(Intercept)	−5.12	−7.44 – −2.80	<.001
Block 2	2.56	−0.07 – 5.19	.056
Block 3	0.72	−1.91 – 3.35	.590
Block 4	5.16	2.53 – 7.79	<.001
Block 5	4.12	1.49 – 6.75	.002
Random Effects			
σ^2	44.52		
τ_{00} PPT	25.03		
ICC	.36		
N PPT	50		
Observations	250		
Marginal R^2 / Conditional R^2	.052 / .393		

Computational modelling

Supplementary Figure 8

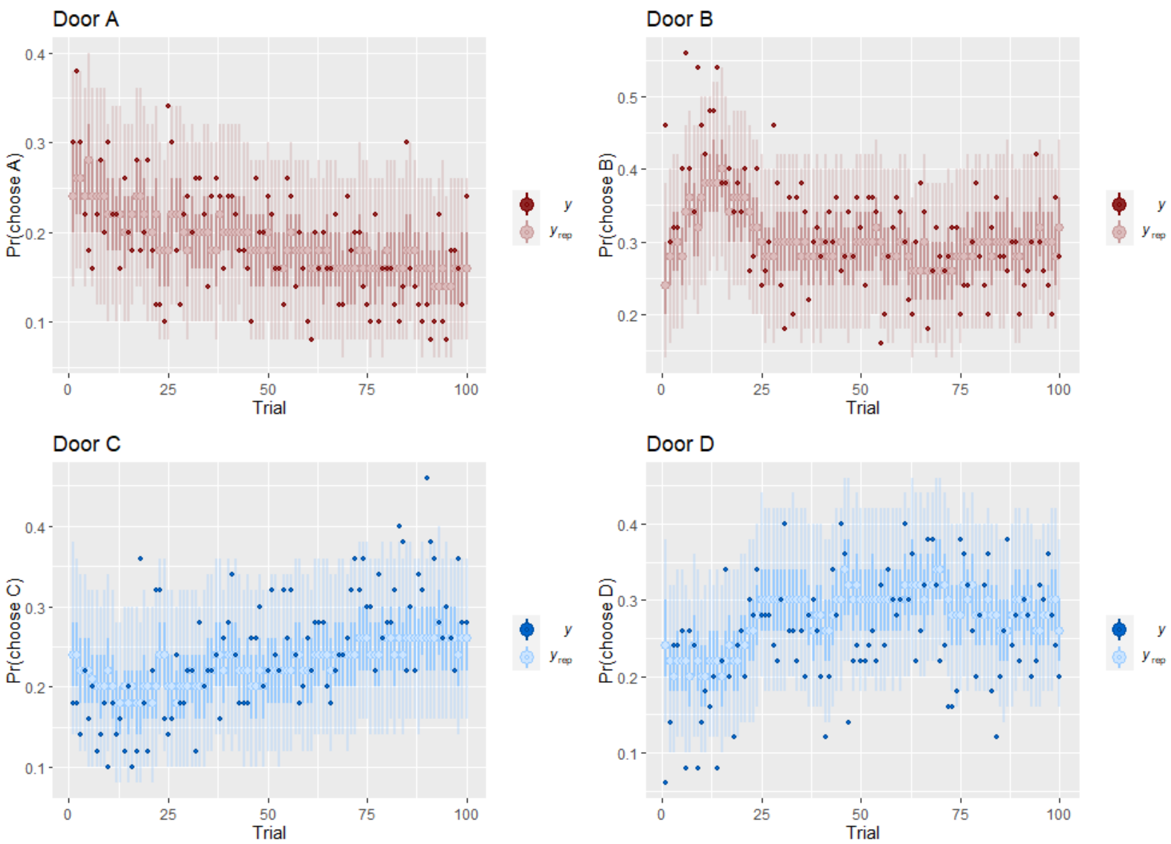
Nonthreat condition, posterior distributions of the hyper (group) parameters



Note. Arew (Reward sensitivity), Apun (Loss sensitivity), K (Forgetfulness), betaF (Frequency sensitivity), betaP (Choice perseveration).

Supplementary Figure 9

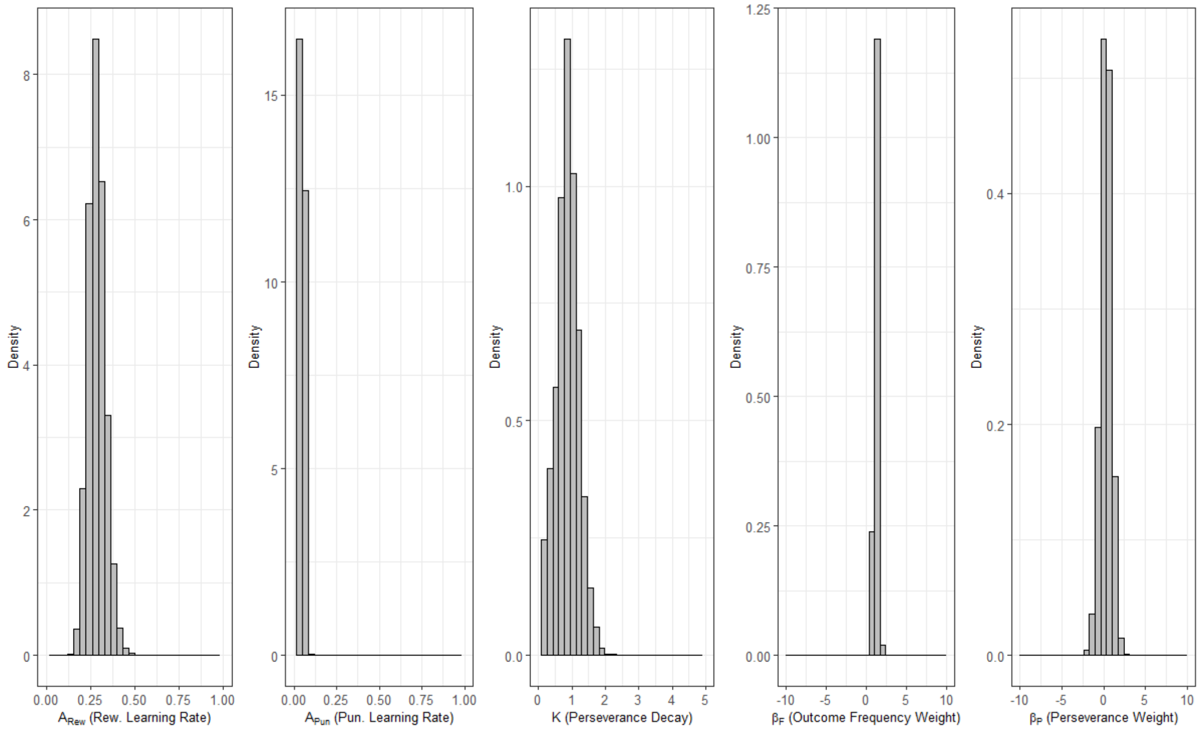
Nonthreat condition, posterior predictive checks



Note. Observed versus the posterior predicted probability of choosing each deck over trials. To illustrate performance trends “disadvantageous” doors are presented as red and “advantageous” doors in blue. The x-axis represents trial number, and y-axis the proportion of participants who selected the deck on each trial (Y), along with predictions (Yrep) computed from draws from the posterior predictive distribution of the same values. 50% (dark) and 95% (light) prediction intervals illustrate uncertainty.

Supplementary Figure 10

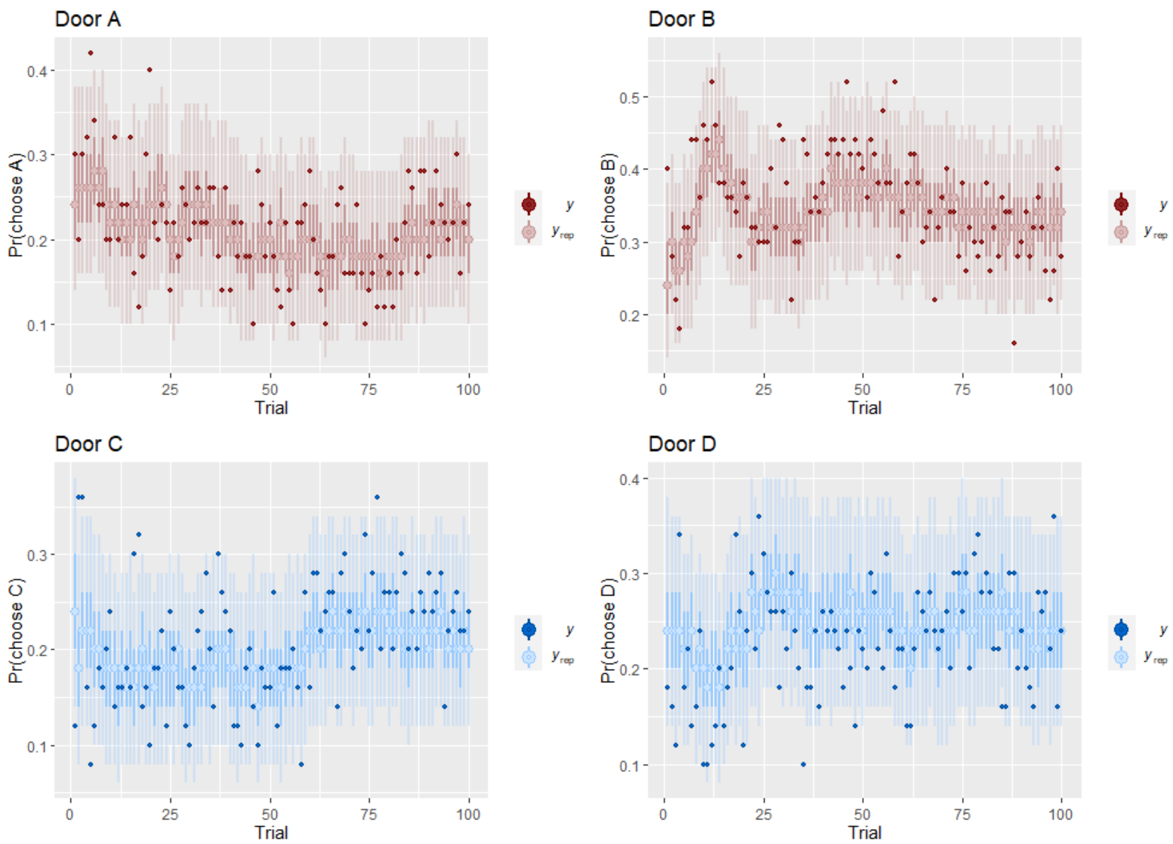
Threat condition, posterior distributions of the hyper (group) parameters



Note. Arew (Reward sensitivity), Apun (Loss sensitivity), K (Forgetfulness), betaF (Frequency sensitivity), betaP (Choice perseveration).

Supplementary Figure 11

Threat condition, posterior predictive checks



Note. Observed versus the posterior predicted probability of choosing each deck over trials. To illustrate performance trends "disadvantageous" doors are presented as red and "advantageous" doors in blue. The x-axis represents trial number, and y-axis the proportion of participants who selected the deck on each trial (Y), along with predictions (Yrep) computed from draws from the posterior predictive distribution of the same values. 50% (dark) and 95% (light) prediction intervals illustrate uncertainty.

Supplementary Materials: Chapter Three

Linear Mixed Model Outputs

Supplementary Materials Table 1

Pairwise comparisons: Affect between conditions. Nonthreat – Threat

Affect	Estimate	SE	DF	t.ratio	p.value
Amusing	−1.64	3.138135	2673	−0.5226	.999
Boring	**16.86**	**3.138135**	**2673**	**5.372618**	**<.001**
Confusing	−5.91	3.138135	2673	−1.88328	.578
Creepy	**−26.73**	**3.138135**	**2673**	**−8.5178**	**<.001**
Disgusting	−5.07	3.138135	2673	−1.61561	.792
Engaging	**−19.24**	**3.138135**	**2673**	**−6.13103**	**<.001**
Enjoyable	−2.95	3.138135	2673	−0.94005	.997
Frightening	**−34.65**	**3.138135**	**2673**	**−11.0416**	**<.001**
Frustrating	−2.64	3.138135	2673	−0.84126	.999
Funny	0.38	3.138135	2673	0.121091	1
Interesting	**−13.53**	**3.138135**	**2673**	**−4.31148**	**<.001**
Sad	−3.3	3.138135	2673	−1.05158	.992
Surprising	**−24.19**	**3.138135**	**2673**	**−7.7084**	**<.001**
Unpredictable	**−11.93**	**3.138135**	**2673**	**−3.80162**	**<.001**

Note. The MVT correction was applied to adjust for multivariate comparisons.

Supplementary Materials Table 2

Order and Overall Performance

	Score		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	53.90	51.72 – 56.08	**<.001**
Condition	1.90	−0.26 – 4.06	.084
Order	0.64	−2.45 – 3.73	.683
Condition * Order	−4.90	−7.96 – −1.84	**.002**
Random Effects			
σ^2	30.00		
$\tau_{00\ ID}$	31.23		
ICC	0.51		
N_{ID}	100		
Observations	200		
Marginal R^2 / Conditional R^2	0.038 / 0.529		

Supplementary Materials Table 3

Performance Over time

	Score		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	13.68	12.85 – 14.51	**<.001**
Block 2	−0.54	−1.57 – 0.49	.304
Block 3	−0.10	−1.13 – 0.93	.849
Block 4	−0.18	−1.21 – 0.85	.732
Condition	0.60	−0.43 – 1.63	.254
Order	0.70	−0.47 – 1.87	.242
Block 2 * Condition	−0.46	−1.92 – 1.00	.536
Block 3 * Condition	0.14	−1.32 – 1.60	.851
Block 4 * Condition	−0.18	−1.64 – 1.28	.809
Block 2 * Order	−0.92	−2.38 – 0.54	.216
Block 3 * Order	−0.12	−1.58 – 1.34	.872
Block 4 * order	−1.12	−2.58 – 0.34	.132
Condition * Order	−0.92	−2.38 – 0.54	.216
Block 2 * Condition * Order	−0.24	−2.30 – 1.82	.819

Block 3 * Condition * Order	−1.22	−3.28 – 0.84	.246
Block 4 * Condition * Order	0.24	−1.82 – 2.30	.819
Random Effects			
σ^2	6.89		
$\tau_{00\ ID}$	2.03		
ICC	0.23		
N_{ID}	100		
Observations	800		
Marginal R^2 / Conditional R^2	0.049 / 0.265		

Individual differences

We investigated the possibility that measures of individual difference (Intolerance of uncertainty and need for closure) predicted performance on the task.

Questionnaires

Intolerance of uncertainty (Carleton et al., 2007)

The Intolerance of Uncertainty Short Form (IoU; Carleton et al., 2007) assesses an individual's valenced associations to uncertainty and ambiguous situations. IoU is a 12-item condensed version of the original 27-item Intolerance of Uncertainty Scale (Freeston et al., 1994). A sample item is "I can't stand being taken by surprise". Responses are rated on a 5-point Likert scale, from 1 =" not at all characteristic of me" to 5 =" entirely characteristic of me"). Higher scores indicate greater IoU.

Need for closure (Roets & Van Hiel, 2011)

The Need for Closure Scale was developed to index one's motivational drive to seek and maintain certainty when dealing with ambiguous informational constraints (Roets & Van Hiel, 2007; Webster & Kruglanski, 1994). Here, we used a brief version (NfC; Roets & Van Hiel, 2011). This scale contains 15 items that are rated on a 5-point Likert scale (from 1 =" not at all characteristic of me" to 5 =" entirely characteristic of me"). A sample item is "I would quickly become impatient and irritated if I would not find a solution to a problem immediately". Higher scores indicate greater need for closure.

Analysis

We assessed the relationship between variables using a series of Pearson's correlation coefficients. Effect sizes were interpreted following empirical guidelines (Funder & Ozer, 2019): $r = .10$ (small), .20 (medium), and >.30 (large). Additionally, we report bias-corrected and accelerated bootstrapped 95% confidence intervals, calculated from 9,999 bootstrapped samples.

We also ran a series of linear mixed models to assess the effect of individual differences on overall performance, with condition, individual differences (either IoU or NfC), and their interaction as fixed factors. The overall score was used as a dependent variable. Intercepts were allowed to vary as a random factor at the level of the individual. Continuous fixed factors were standardised (z score) to aid model convergence.

Results

Results demonstrated, Intolerance of uncertainty did not significantly predict overall performance, threat = $r(98) = -.095$, $p = .347$, Cis = −.29, 0.12; nonthreat = $r(98) = 0.03$, $p = 0.742$, Cis = −.16, .23. Again, no significant relationship was found between measures of need for closure, and overall performance: threat = $r(98) = -.099$, $p = .325$, Cis = −.31, .13; nonthreat = $r(98) = -.102$, $p = .311$, Cis = −.29, .10. Here, the large spread in bootstrapped intervals is concerning and suggests we are underpowered, not surprising given small estimate of effect reported.

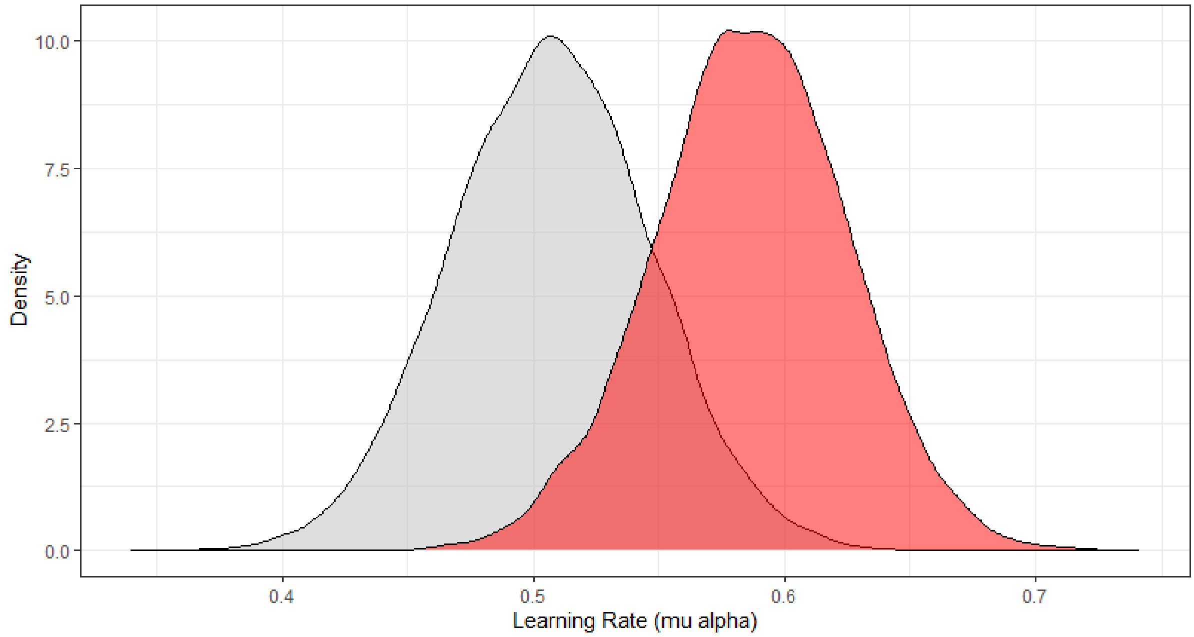
However, interestingly, the distributions of bootstrapped intervals do not appear to be equally spread across the range of possible values. Estimates tend to lean disproportionately towards the direction of a negative correlation, thus both measures of individual difference (which are highly correlated, $r(98) = .70$, $p < .001$) are more likely to result in an estimate of negative correlation with overall performance. Although this is nothing more than a fragile clue that should not be interpreted, it may warrant further investigation with an appropriately powered sample ($n > 781$, based on power analysis for a Pearson's correlation, $r = .1$, power = .80, $\alpha = .05$).

Again, results from our linear mixed models revealed no significant main or interaction effects for either individual difference measure on overall performance. Specifically, in the model including IoU, there was no significant effect of condition ($F(1, 98) = 0.47$, $p = .496$), IoU ($F(1, 98) = 0.16$, $p = .693$), or their interaction ($F(1, 98) = 1.65$, $p = .202$). Similarly, in the model including NfC, no significant effects were found for condition ($F(1, 98) = 0.46$, $p = .500$), NfC ($F(1, 98) = 1.36$, $p = .246$), or their interaction ($F(1, 98) = 0.01$, $p = .936$).

Posterior distribution of group-level parameter estimates

Supplementary Materials Figure 1

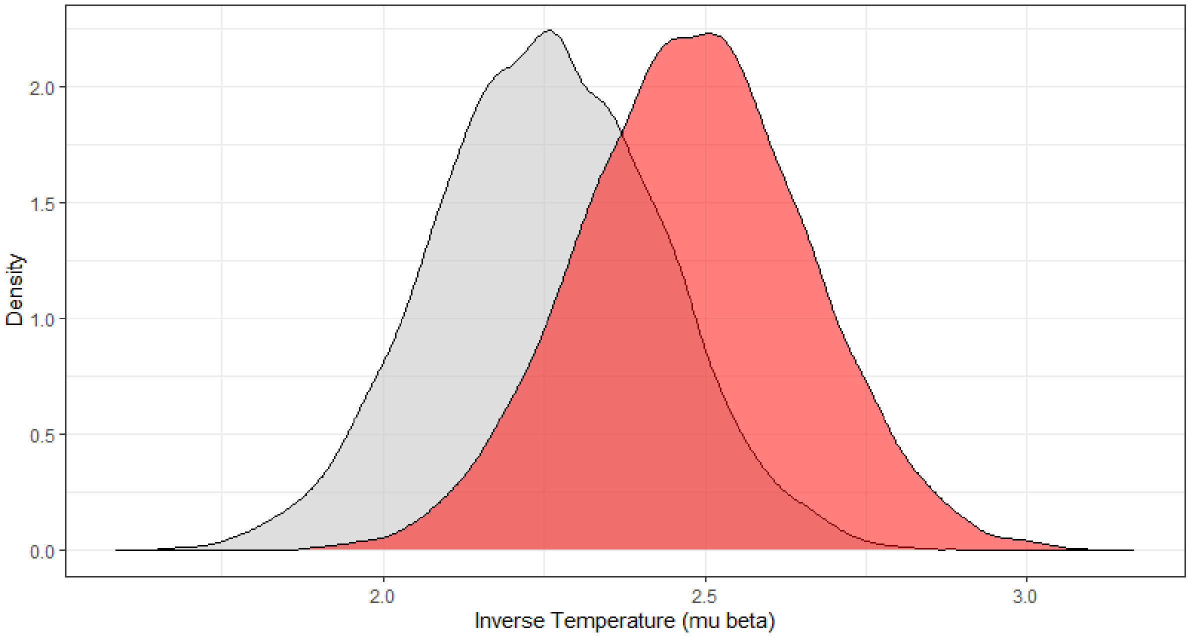
Learning Rate



Note. Posterior distribution of group-level parameter estimates for mu alfa (Learning Rate). Threat= red, Nonthreat = Gray.

Supplementary Materials Figure 2

Inverse temperature

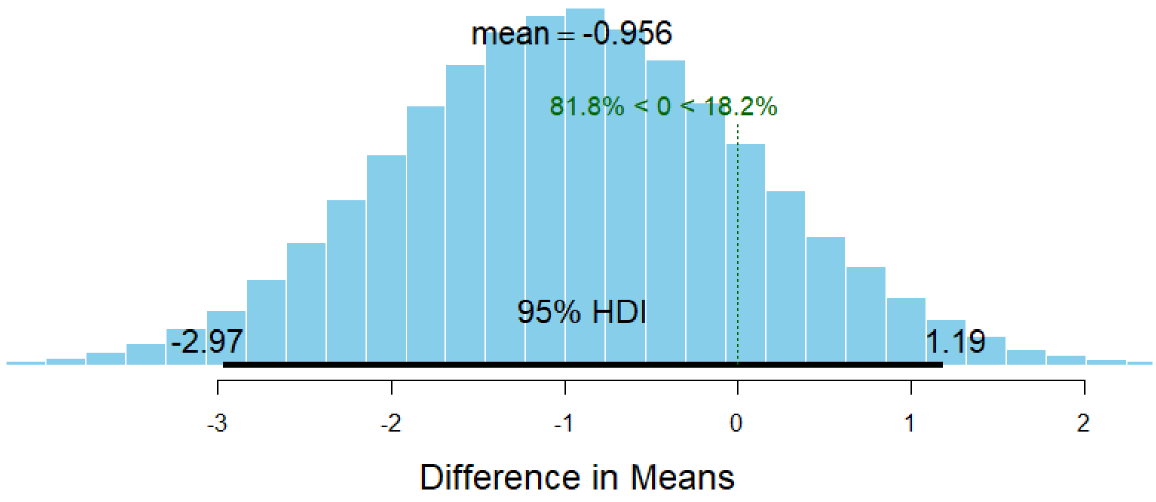


Note. Posterior distribution of group-level parameter estimates for mu beta (Inverse temperature). Threat= red, Nonthreat = Gray.

Bayesian estimation of the posterior distribution of differences in means

Supplementary Materials Figure 3

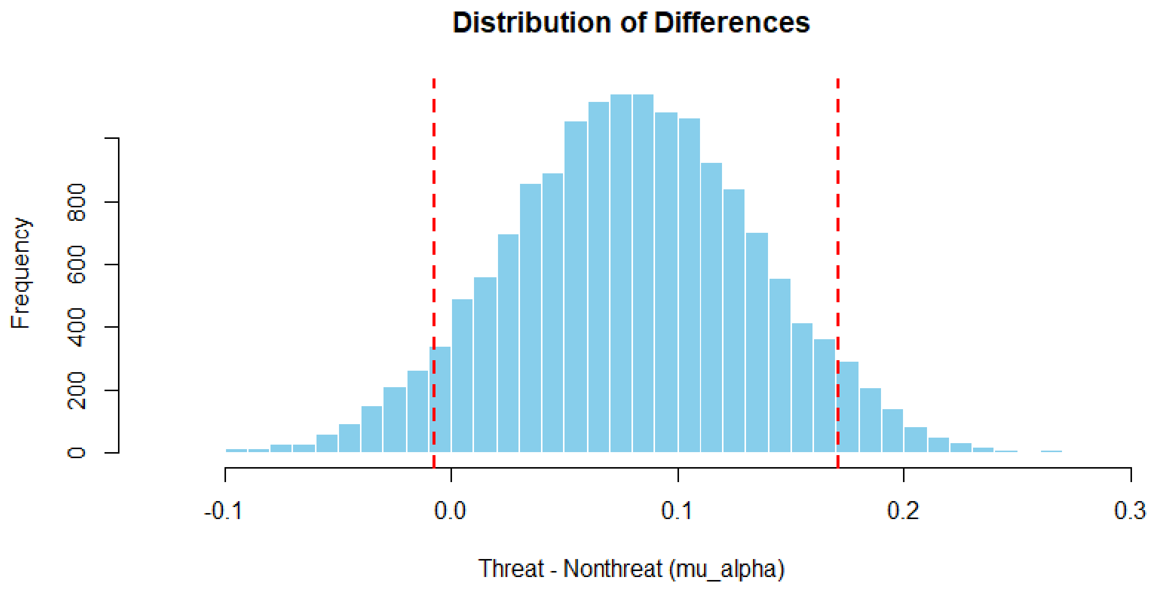
Bayesian estimation of posterior distribution: Overall Performance



Note. Bayesian estimation of the posterior distribution of differences in means (threat – nonthreat). This analysis was conducted using the BEST package in R (Kruschke & Meredith, 2021).

Supplementary Materials Figure 4

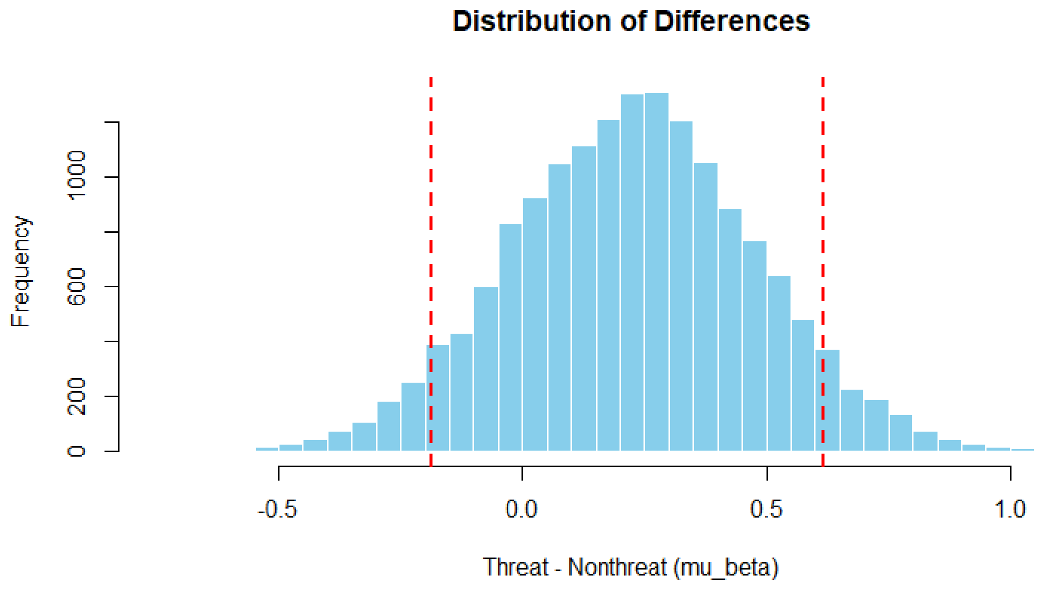
Learning Rate



Note. This Bayesian framework models the uncertainty in parameter values as a probability distribution (Kruschke, 2018). Differences in means (Learning Rate) between conditions are represented along the x axis. Vertical dotted lines represent the 89% HDI.

Supplementary Materials Figure 5

Inverse Temperature



Note. Differences in means (Inverse Temperature) between conditions are represented along the x axis. Vertical dotted lines represent the 89% HDI.

Pre-scene instructional scripts

Threat condition

You are trying to escape a building that is slowly collapsing.

Your goal is to get as far away from the danger zone as you can. To achieve this, open as many doors as possible. This will allow you through the building.

In each room, there are three colour coded doors to choose from, but not all doors will open. Some doors open more regularly but this can change. Your progress will be tracked so that you can always view the number of doors that you have opened. I won't tell you how long the task will take. You must keep choosing doors until the computer stops. Now get ready to start your escape.

Good luck.

Nonthreat condition

You are in a building trying to find the exit to meet your friends.

Your goal is to get as close to the outside car park as you can. To achieve this, open as many doors as possible. This will allow you through the building.

In each room, there are three colour coded doors to choose from, but not all doors will open. Some doors open more regularly but this can change. Your progress will be tracked so that you can always view the number of doors that you have opened. I won't tell you how long the task will take. You must keep choosing doors until the computer stops. Now get ready to find your friends.

Good luck.

Supplementary Materials: Chapter Four

Study 1 and 2

Supplementary Table 1

Means, standard deviations, and non-parametric Spearman's correlations

	Study	Mean (SD)	FnE	PD	IoU
1.FnE	1	2.45 (.62)			
	2	2.52 (.52)			
	3	2.66 (.59)			
	P	2.55 (.59)			
2.PD	1	1.62 (.65)	.23***		
	2	1.55 (.49)	.14		
	3	1.62 (.67)	.28***		
	P	1.61 (.63)	.23***		
3.IoU	1	2.26 (.62)	.52***	.34***	
	2	2.03 (.48)	.52***	.05	
	3	2.74 (.76)	.52***	.25***	
	P	2.40 (.72)	.52***	.23***	
4.NfC	1	2.49 (.64)	.41***	.35***	.77***
	2	2.38 (.47)	.38***	.29*	.57***
	3	2.97 (.67)	.47***	.28***	.74***
	P	2.66 (.67)	.45***	.27***	.77***

Note. *Indicates correlations with a significant P value < .050, ** < .01, *** <.001. Degrees of freedom (n): Study 1 T0= 209 (211), Study 2 T0= 99 (101), Study 3= 197 (199), Pooled (P) = 509 (511). Borders highlight associations within (evaluation sensitivity = solid, complexity intolerance= dashed), and between (dotted) theorised latent constructs.

Study 3

Gender

A series of Welch Two Sample t-test (two tailed) found no significant differences between genders (male and female) across all measures: IoU = t (195.79) = 0.15, p = .883; NfC = t (194.79) = −1.95, p = .053; FnE = t (196) = −1.24, p = .216; PD = t (188.8) = 1.16, p = .246; AL = t (195.99) = 1.37, p = .172. This analysis did not include participants that identified as non-binary (n = 1).

Linear regressions

To investigate we carried out series of multiple linear regressions to test if the interaction between measures of evaluation sensitivity and ratings of authoritarian leadership and, significantly predicted complexity tolerance.

First, we modelled the effect on FnE, AL and their interaction on IoU. The model accounted for a significant amount of the variance in IoU ratings, R^2 = .27, $F(3, 195)$ = 25.1, $p < .001$. FnE significantly predicted IoU (β = .87, SE = .22, $p < .001$), indicating as previously reported that higher FnE was associated with higher rating of IoU. There was no significant effect of AL on IoU (β = .29, SE = .27, p = .293), and no evidence of an interaction between FnE and AL predicting IoU was found (β = −.10, SE = .10, p = .342).

Next, we modelled the effect on FnE, Al and their interaction on NfC. Again, the overall model accounted for a significant amount of the variance in NfC ratings, R^2 = .24, $F(3, 195)$ = 21.76, $p < .001$. FnE significantly predicted NfC (β = .72, SE = .20, $p < 0.01$), but no effect of AL on NfC was observed (β = .31, SE = .25, p = .090). No evidence of an interaction between FnE and AL predicted NfC was found (β = .09, SE = .09, p = .308).

A similar pattern of results was found with power distance, with effect on PD, Al and their interaction on IoU, accounting for a significant amount of the variance in IoU ratings, R^2 = .07, $F(3, 195)$ = 6.25, $p < .001$. Here a significant effect of PD on IoU was found (β = .49, SE = .25, p = .046), but no significant effect of AL in predicting IoU (β = .16, SE = .16, p = .325). Further, there was no evidence that the interaction between PD and AL predicted IoU, β = −.07, SE = .09, p = .417.

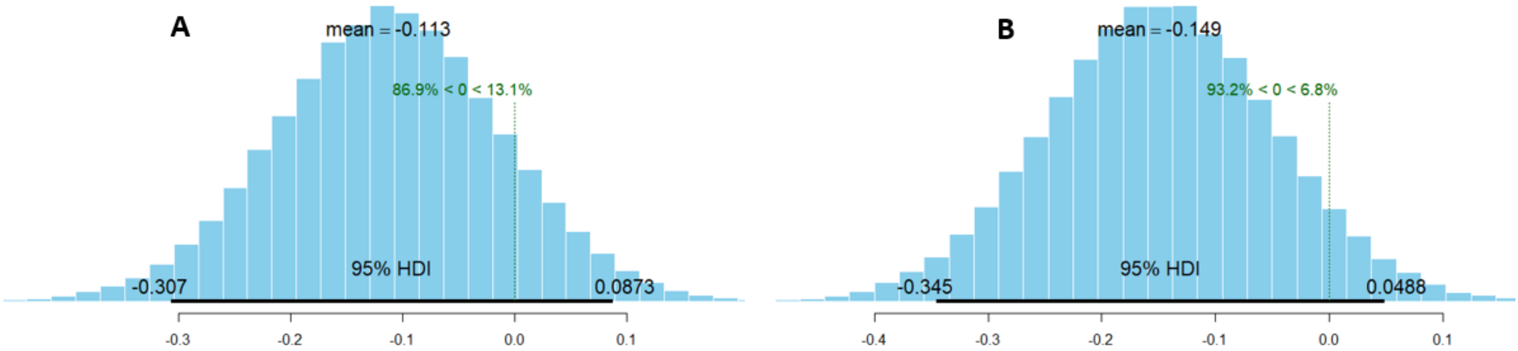
Finally, we modelled the effect on PD, AL and their interaction on NfC, again with the overall model explaining a significant amount of the variance in NfC ratings, $R^2 = .09$, $F(3, 195) = 7.76$, $p < .001$. However, the effect of PD on NfC marginally fell below the threshold of significance following the inclusion of AL into the model ($\beta = .42$, $SE = .21$, $p = .051$), suggesting that in civilian populations the relationship between PD and NfC may not hold when controlling for an individual's working environment. Yet again, there was no significant effect of AL on NfC, $\beta = .15$, $SE = .14$, $p = .270$, and no evidence of an interaction between PD and AL predicting NfC, $\beta = .09$, $SE = .09$, $p = .308$.

Thus, no support was found that AL moderates the reported between-construct relationship between evaluation sensitivity and complexity tolerance.

Study 4

Supplementary Figure 1

Histograms of Bayesian Estimation of Posterior Distribution of Differences in Means

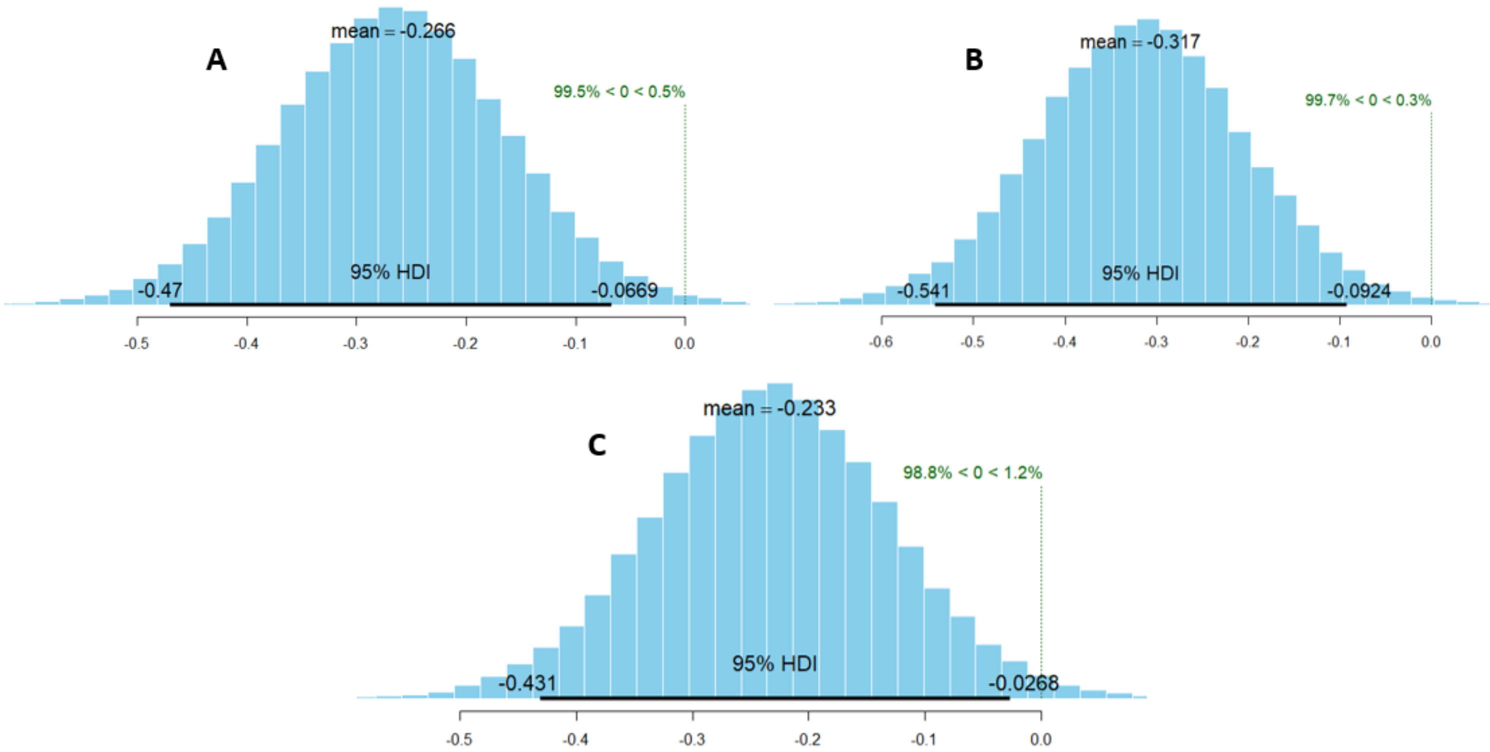


Note. This Bayesian framework models the uncertainty in parameter values as a probability distribution (Kruschke, 2018). Differences in means (T1−T0) following the operational command course are represented along the x axis. All $\hat{R}$ values where <1.1 suggesting adequate model convergence. The means of the posterior probability represent a point estimate (most likely value) of the true difference in means given the data. A) shows the results of IoU, mean = -.113, with an 86.9% probability that the true of a reduction in IoU is below 0. B) shows the results of NfC, mean =-.149, with an 93.2% probability that the true reduction in NfC is below 0.

Study 5

Supplementary Figure 2

Histograms of Bayesian Estimation of Posterior Distribution of Differences in Means



Note. A series of Bayesian estimations of posterior distribution of differences in means following training in Cohort B. Differences in means (T1−T0) following training are represented along the x axis. All $\hat{R}$ values where <1.1 suggesting adequate model convergence. A) shows the results of IoU, mean = −.266, with an 99.5% probability that the true of a reduction in IoU Is below 0. B) shows the results of NfC, mean =−.317, with an 99.7% probability that the true reduction in NfC is below 0. Finally, C) shows the results of PD, mean = −.223, with an 98.8% probability that the true of a reduction in PD Is below 0.

References

Ahn, W.-Y., Busemeyer, J. R., Wagenmakers, E.-J., & Stout, J. C. (2008). Comparison of Decision Learning Models Using the Generalization Criterion Method. *Cognitive Science*, *32*(8), 1376–1402. https://doi.org/10.1080/03640210802352992

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Computational Psychiatry*, *1*(0), Article 0. https://doi.org/10.1162/CPSY_a_00002

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2011). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Journal of Neuroscience, Psychology, and Economics*, *4*(2), 95–110. https://doi.org/10.1037/a0020684

Ahn, W.-Y., Vasilev, G., Lee, S.-H., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00849

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2016). The Trier Social Stress Test: Principles and practice. *Neurobiology of Stress*, *6*, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001

Almaatouq, A., Alsobay, M., Yin, M., & Watts, D. J. (2021). Task complexity moderates group synergy. *Proceedings of the National Academy of Sciences*, *118*(36), e2101062118. https://doi.org/10.1073/pnas.2101062118

Angelidis, A., Solis, E., Lautenbach, F., Does, W. van der, & Putman, P. (2019). I'm going to fail! Acute cognitive performance anxiety increases threat-interference and impairs

WM performance. *PLOS ONE*, *14*(2), e0210824. https://doi.org/10.1371/journal.pone.0210824

Anselme, P. (2010). The uncertainty processing theory of motivation. *Behavioural Brain Research*, *208*(2), 291–310. https://doi.org/10.1016/j.bbr.2009.12.020

Aram, S., Levy, L., Patel, J. B., Anderson, A. A., Zaragoza, R., Dashtestani, H., Chowdhry, F. A., Gandjbakhche, A., & Tracy, J. K. (2019). The Iowa Gambling Task: A Review of the Historical Evolution, Scientific Basis, and Use in Functional Neuroimaging. *SAGE Open*, *9*(3), 2158244019856911. https://doi.org/10.1177/2158244019856911

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410–422. https://doi.org/10.1038/nrn2648

Arnsten, Mazure, & Sinha. (2012). This Is Your Brain in Meltdown. *Scientific American*, *306*(4), 48–53.

Asutay, E., & Västfjäll, D. (2022). The continuous and changing impact of affect on risky decision-making. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-14810-w

Asutay, E., & Västfjäll, D. (2024). Affective integration in experience, judgment, and decision-making. *Communications Psychology*, *2*(1), 126. https://doi.org/10.1038/s44271-024-00178-2

Azriel, O., & Bar-Haim, Y. (2020). Attention bias. In *Clinical handbook of fear and anxiety: Maintenance processes and treatment mechanisms* (pp. 203–218). American Psychological Association. https://doi.org/10.1037/0000150-012

Baddeley, A. (1972). Selective attention and performance in dangerous environments. *British Journal of Psychology*, *63*(4), 537–546.

Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). OuP Oxford.

Bagneux, V., Thomassin, N., Gonthier, C., & Roulin, J.-L. (2013). Working memory in the processing of the Iowa Gambling Task: An individual differences approach. *PloS One*, *8*(11), e81498.

Bailey, M. R., Simpson, E. H., & Balsam, P. D. (2016). Neural substrates underlying effort, time, and risk-based decision making in motivated behavior. *Neurobiology of Learning and Memory*, *133*, 233–256. https://doi.org/10.1016/j.nlm.2016.07.015

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

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Bartoszek, G., Ranney, R. M., Curanovic, I., Costello, S. J., & Behar, E. (2022). Intolerance of uncertainty and information-seeking behavior: Experimental manipulation of threat relevance. *Behaviour Research and Therapy*, *154*, 104125. https://doi.org/10.1016/j.brat.2022.104125

Bates, D. (2006, May 19). *[R] lmer, p-values and all that*. https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bavolar, J., Kacmar, P., Hricova, M., Schrötter, J., Kovacova-Holevova, B., Köverova, M., & Raczova, B. (2021). Intolerance of uncertainty and reactions to the COVID-19 pandemic. *The Journal of General Psychology*, *0*(0), 1–28. https://doi.org/10.1080/00221309.2021.1922346

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences*, *9*(4), 159–162. https://doi.org/10.1016/j.tics.2005.02.002

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, *123*(11), 2189–2202. https://doi.org/10.1093/brain/123.11.2189

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Ben Hassen, N., Molins, F., Paz, M., & Serrano, M.-Á. (2023). Later stages of acute stress impair reinforcement-learning and feedback sensitivity in decision making. *Biological Psychology*, *180*, 108585. https://doi.org/10.1016/j.biopsycho.2023.108585

Ben-Asher, N., & Gonzalez, C. (2015). Effects of cyber security knowledge on attack detection. *Computers in Human Behavior*, *48*, 51–61.

Ben-Zur, H., & Zeidner, M. (2009). Threat to Life and Risk-Taking Behaviors: A Review of Empirical Findings and Explanatory Models. *Personality and Social Psychology Review*, *13*(2), 109–128. https://doi.org/10.1177/1088868308330104

Berenbaum, H., Bredemeier, K., & Thompson, R. J. (2008). Intolerance of uncertainty: Exploring its dimensionality and associations with need for cognitive closure, psychopathology, and personality. *Journal of Anxiety Disorders*, *22*(1), 117–125. https://doi.org/10.1016/j.janxdis.2007.01.004

Berghorst, L. H., Bogdan, R., Frank, M. J., & Pizzagalli, D. A. (2013). Acute stress selectively reduces reward sensitivity. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00133

Berkun, M. M., Bialek, H. M., Kern, R. P., & Yagi, K. (1962). Experimental studies of psychological stress in man. *Psychological Monographs: General and Applied*, *76*(15), 1.

Bienertova-Vasku, J., Lenart, P., & Scheringer, M. (2020). Eustress and Distress: Neither Good Nor Bad, but Rather the Same? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *42*(7), e1900238. https://doi.org/10.1002/bies.201900238

Binsch, O., Bottenheft, C., Landman, A., Roijendijk, L., & Vermetten, E. H. G. J. M. (2021). Testing the applicability of a virtual reality simulation platform for stress training of first responders. *Military Psychology*, *33*(3), 182–196. https://doi.org/10.1080/08995605.2021.1897494

Birrell, J., Meares, K., Wilkinson, A., & Freeston, M. (2011). Toward a definition of intolerance of uncertainty: A review of factor analytical studies of the Intolerance of Uncertainty Scale. *Clinical Psychology Review*, *31*(7), 1198–1208. https://doi.org/10.1016/j.cpr.2011.07.009

Blacker, K. J., Hamilton, J., Roush, G., Pettijohn, K. A., & Biggs, A. T. (2019). Cognitive Training for Military Application: A Review of the Literature and Practical Guide. *Journal of Cognitive Enhancement*, *3*(1), 30–51. https://doi.org/10.1007/s41465-018-0076-1

Bland, A. R., & Schaefer, A. (2012). Different Varieties of Uncertainty in Human Decision-Making. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00085

Bogdan, R., & Pizzagalli, D. A. (2006). Acute Stress Reduces Reward Responsiveness: Implications for Depression. *Biological Psychiatry*, *60*(10), 1147–1154. https://doi.org/10.1016/j.biopsych.2006.03.037

Bogdan, R., Santesso, D. L., Fagerness, J., Perlis, R. H., & Pizzagalli, D. A. (2011). Corticotropin-Releasing Hormone Receptor Type 1 (CRHR1) Genetic Variation and Stress Interact to Influence Reward Learning. *Journal of Neuroscience*, *31*(37), 13246–13254. https://doi.org/10.1523/JNEUROSCI.2661-11.2011

Bonetto, E., Pichot, N., Pavani, J.-B., & Adam-Troïan, J. (2020). Creative Individuals are Social Risk-takers: Relationships between Creativity, Social Risk-Taking and Fear of Negative Evaluations. *Creativity. Theories – Research – Applications*, *7*(2), 309–320. https://doi.org/10.2478/ctra-2020-0016

Bonetto, E., Pichot, N., Pavani, J.-B., & Adam-Troïan, J. (2021). The paradox of creativity. *New Ideas in Psychology*, *60*, 100820. https://doi.org/10.1016/j.newideapsych.2020.100820

Bottesi, G., Marchetti, I., Sica, C., & Ghisi, M. (2020). What is the internal structure of intolerance of uncertainty? A network analysis approach. *Journal of Anxiety Disorders*, *75*, 102293. https://doi.org/10.1016/j.janxdis.2020.102293

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*(3), 211–241. https://doi.org/10.1016/0001-6918(92)90019-A

Brehmer, B. (2005). Micro-worlds and the circular relation between people and their environment. *Theoretical Issues in Ergonomics Science*, *6*(1), 73–93. https://doi.org/10.1080/14639220512331311580

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S. E., & Rosen, B. R. (1996). Response and Habituation of the Human Amygdala during Visual Processing of Facial Expression. *Neuron*, *17*(5), 875–887. https://doi.org/10.1016/S0896-6273(00)80219-6

Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health*, *33*(6), 436–457. https://doi.org/10.1016/S1054-139X(03)00052-1

Brevers, D., Bechara, A., Cleeremans, A., & Noel, X. (2013). Iowa Gambling Task (IGT): Twenty years after – gambling disorder and IGT. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00665

Broadbent, D. E. (1971). *Decision and stress* (pp. xiv, 522). Academic Press.

Broadbent, D. E. (1984). Performance and its measurement. *British Journal of Clinical Pharmacology*, *18*(Suppl 1), 5S–9S. https://doi.org/10.1111/j.1365-2125.1984.tb02577.x

Brookes, J., Hall, S., Frühholz, S., & Bach, D. R. (2024). Immersive VR for investigating threat avoidance: The VRthreat toolkit for Unity. *Behavior Research Methods*, *56*(5), 5040-5054. https://doi.org/10.3758/s13428-023-02241-y

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Brun, C., Zerhouni, O., Akinyemi, A., Houtin, L., Monvoisin, R., & Pinsault, N. (2023). Impact of uncertainty intolerance on clinical reasoning: A scoping review of the 21st-century literature. *Journal of Evaluation in Clinical Practice*, *29*(3), 539–553. https://doi.org/10.1111/jep.13756

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Bublatzky, F., Alpers, G. W., & Pittig, A. (2017). From avoidance to approach: The influence of threat-of-shock on reward-based decision making. *Behaviour Research and Therapy*, *96*, 47–56. https://doi.org/10.1016/j.brat.2017.01.003

Buelow, M. T., & Cayton, C. (2020). Relationships between the big five personality characteristics and performance on behavioral decision making tasks. *Personality and Individual Differences*, *160*, 109931.

Buelow, M. T., & Suhr, J. A. (2009). Construct Validity of the Iowa Gambling Task. *Neuropsychology Review*, *19*(1), 102–114. https://doi.org/10.1007/s11065-009-9083-4

Buhr, K., & Dugas, M. J. (2002). The intolerance of uncertainty scale: Psychometric properties of the English version. *Behaviour Research and Therapy*, *40*(8), 931–945. https://doi.org/10.1016/S0005-7967(01)00092-4

Bull, P. N., Tippett, L. J., & Addis, D. R. (2015). Decision making in healthy participants on the Iowa Gambling Task: New insights from an operant approach. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00391

Busch, C. A., Wiesenthal, N. J., Mohammed, T. F., Anderson, S., Barstow, M., Custalow, C., Gajewski, J., Garcia, K., Gilabert, C. K., Hughes, J., Jenkins, A., Johnson, M., Kasper, C., Perez, I., Robnett, B., Tillett, K., Tsefrekas, L., Goodwin, E. C., & Cooper, K. M. (2023). The Disproportionate Impact of Fear of Negative Evaluation on First-Generation College Students, LGBTQ+ Students, and Students with Disabilities in College Science Courses. *CBE Life Sciences Education*, *22*(3), ar31. https://doi.org/10.1187/cbe.22-10-0195

Butter, C. M. (1969). Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in *Macaca mulatta*. *Physiology & Behavior*, *4*(2), 163–171. https://doi.org/10.1016/0031-9384(69)90075-4

Byrne, K. A., Cornwall, A. C., & Worthy, D. A. (2019). Acute stress improves long-term reward maximization in decision-making under uncertainty. *Brain and Cognition*, *133*, 84–93. https://doi.org/10.1016/j.bandc.2019.02.005

Byrne, K. A., Peters, C., Willis, H. C., Phan, D., Cornwall, A., & Worthy, D. A. (2020). Acute stress enhances tolerance of uncertainty during decision-making. *Cognition*, *205*, 104448. https://doi.org/10.1016/j.cognition.2020.104448

Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. *Current Directions in Psychological Science*, *7*(3), 83–87. https://doi.org/10.1111/1467-8721.ep10773005

Carleton, R. N., Duranceau, S., Shulman, E. P., Zerff, M., Gonzales, J., & Mishra, S. (2016). Self-reported intolerance of uncertainty and behavioural decisions. *Journal of Behavior Therapy and Experimental Psychiatry*, *51*, 58–65. https://doi.org/10.1016/j.jbtep.2015.12.004

Carleton, R. N., Norton, M. P. J., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105–117.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*, 1–32. https://doi.org/10.18637/jss.v076.i01

Castellà, J., Boned, J., Méndez-Ulrich, J. L., & Sanz, A. (2020). Jump and free fall! Memory, attention, and decision-making processes in an extreme sport. *Cognition and Emotion*, *34*(2), 262–272. https://doi.org/10.1080/02699931.2019.1617675

Champely, S. (2020). *pwr: Basic Functions for Power Analysis*. https://CRAN.R-project.org/package=pwr

Chen, J. T.-H., & Lovibond, P. F. (2016). Intolerance of uncertainty is associated with increased threat appraisal and negative affect under ambiguity but not uncertainty. *Behavior Therapy*, *47*(1), 42-53.

Cheng, B.-S., Chou, L.-F., Wu, T.-Y., Huang, M.-P., & Farh, J.-L. (2004). Paternalistic leadership and subordinate responses: Establishing a leadership model in Chinese organizations. *Asian Journal of Social Psychology*, *7*(1), 89-117. https://doi.org/10.1111/j.1467-839X.2004.00137.x

Chiu, Y.-C., Huang, J.-T., Lee, W.-K., Lin, C.-J., & Lin, C.-H. (2022). Reanalyzing the Maia and McClelland (2004) Empirical Data: How Do Participants Really Behave in the Iowa Gambling Task? *Frontiers in Psychiatry*, *13*, 788456. https://doi.org/10.3389/fpsyt.2022.788456

Choi, J. M., Padmala, S., & Pessoa, L. (2015). Counteracting effect of threat on reward enhancements during working memory. *Cognition and Emotion*, *29*(8), 1517-1526. https://doi.org/10.1080/02699931.2014.993596

Chou, K.-P., Wilson, R. C., & Smith, R. (2024). The influence of anxiety on exploration: A review of computational modeling studies. *Neuroscience & Biobehavioral Reviews*, *167*, 105940. https://doi.org/10.1016/j.neubiorev.2024.105940

Clauss, K., Gorday, J. Y., & Bardeen, J. R. (2022). Eye tracking evidence of threat-related attentional bias in anxiety- and fear-related disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *93*, 102142. https://doi.org/10.1016/j.cpr.2022.102142

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Comfort, L. K., & Wukich, C. (2013). Developing Decision-Making Skills for Uncertain Conditions: The Challenge of Educating Effective Emergency Managers. *Journal of Public Affairs Education*, *19*(1), 53–71. https://doi.org/10.1080/15236803.2013.12001720

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal Learning and Dopamine: A Bayesian Perspective. *Journal of Neuroscience*, *35*(6), 2407–2416. https://doi.org/10.1523/JNEUROSCI.1989-14.2015

Damasio, A. R. (1994). Descartes' error and the future of human life. *Scientific American*, *271*(4), 144–144.

Damasio, A. R., Everitt, B. J., Bishop, D., Roberts, A. C., Robbins, T. W., & Weiskrantz, L. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

Daniel-Watanabe, L., & Fletcher, P. C. (2022). Are Fear and Anxiety Truly Distinct? *Biological Psychiatry Global Open Science*, *2*(4), 341–349. https://doi.org/10.1016/j.bpsgos.2021.09.006

Davis, M., Walker, D. L., Miles, L., & Grillon, C. (2010). Phasic vs Sustained Fear in Rats and Humans: Role of the Extended Amygdala in Fear vs Anxiety. *Neuropsychopharmacology*, *35*(1), 105–135. https://doi.org/10.1038/npp.2009.109

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), Article 7095. https://doi.org/10.1038/nature04766

D'Cruz, A.-M., Ragozzino, M. E., Mosconi, M. W., Pavuluri, M. N., & Sweeney, J. A. (2011). Human reversal learning under conditions of certain versus uncertain outcomes. *NeuroImage*, *56*(1), 315–322. https://doi.org/10.1016/j.neuroimage.2011.01.068

Delacre, M., & Klein, O. (2019). *Routliers: Robust Outliers Detection*. https://CRAN.R-project.org/package=Routliers

Delahaij, R., Gaillard, A. W. K., & Soeters, J. (2006). Stress training and the new military environment. *Human Dimensions in Military Operations-Military Leaders' Strategies for Addressing Stress and Psychological Support*, 17A – 1.

Dennis, T. A., Chen, C.-C., & McCandliss, B. D. (2008). Threat-related attentional biases: An analysis of three attention systems. *Depression and Anxiety*, *25*(6), E1–E10. https://doi.org/10.1002/da.20308

den Ouden, H. E. M., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., Franke, B., & Cools, R. (2013). Dissociable Effects of Dopamine and Serotonin on Reversal Learning. *Neuron*, *80*(4), 1090–1100. https://doi.org/10.1016/j.neuron.2013.08.030

Derryberry, D., & Tucker, D. M. (1994). *Motivating the focus of attention.*

Dickerson, J. A., & Kosko, B. (1994). Virtual worlds as fuzzy cognitive maps. *Presence: Teleoperators & Virtual Environments*, *3*(2), 173–189.

Disatnik, D., & Steinhart, Y. (2015). Need for Cognitive Closure, Risk Aversion, Uncertainty Changes, and Their Effects on Investment Decisions. *Journal of Marketing Research*, *52*(3), 349–359. https://doi.org/10.1509/jmr.13.0529

Djikic, M., Oatley, K., & Moldoveanu, M. C. (2013). Opening the Closed Mind: The Effect of Exposure to Literature on the Need for Closure. *Creativity Research Journal*, *25*(2), 149–154. https://doi.org/10.1080/10400419.2013.783735

Dogaheh, E. R., Mohammadkhani, P., & Dolatshahi, B. (2011). Comparison of Group and Individual Cognitive-Behavioral Therapy in Reducing Fear of Negative Evaluation. *Psychological Reports*, *108*(3), 955–962. https://doi.org/10.2466/02.21.PR0.108.3.955-962

Dörner, D., & Funke, J. (2017). Complex Problem Solving: What It Is and What It Is Not. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01153

Dörner, D., Stäudel, T., & Strohschneider, S. (1986). *Moro-Programmdokumentation*.

Driskell, J. E., & Salas, E. (1991). Group decision making under stress. *Journal of Applied Psychology*, *76*(3), 473–478. https://doi.org/10.1037/0021-9010.76.3.473

Dubois, M., & Hauser, T. U. (2022). Value-free random exploration is linked to impulsivity. *Nature Communications*, *13*(1), 4542. https://doi.org/10.1038/s41467-022-31918-9

Duea, S. R., Zimmerman, E. B., Vaughn, L. M., Dias, S., & Harris, J. (2022). A Guide to Selecting Participatory Research Methods Based on Project and Partnership Goals. *Journal of Participatory Research Methods*, *3*(1), 10.35844/001c.32605. https://doi.org/10.35844/001c.32605

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. SAGE Publications.

Eccles, D. W., & and Arsal, G. (2017). The think aloud method: What is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, *9*(4), 514–531. https://doi.org/10.1080/2159676X.2017.1331501

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, *98*(24), 13763–13768. https://doi.org/10.1073/pnas.231499798

Eichler, P., Seine, R., Khanina, E., & Schön, A. (2006). Astronaut training for the European ISS contributions Columbus module and ATV. *Acta Astronautica*, *59*(12), 1146–1152.

Elliot, A. J., & Maier, M. A. (2007). Color and Psychological Functioning. *Current Directions in Psychological Science*, *16*(5), 250–254. https://doi.org/10.1111/j.1467-8721.2007.00514.x

Ernst, M., Grant, S. J., London, E. D., Contoreggi, C. S., Kimes, A. S., & Spurgeon, L. (2003). Decision Making in Adolescents With Behavior Disorders and Adults With Substance Abuse. *American Journal of Psychiatry*, *160*(1), 33–40. https://doi.org/10.1176/appi.ajp.160.1.33

Estes, W. K., & Maddox, T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*(3), 403–408.

Eva, K. W. (2009). Diagnostic error in medical education: Where wrongs can make rights. *Advances in Health Sciences Education*, *14*(1), 71–81. https://doi.org/10.1007/s10459-009-9188-9

Evans, D. A., Stempel, A. V., Vale, R., & Branco, T. (2019). Cognitive Control of Escape Behaviour. *Trends in Cognitive Sciences*, *23*(4), 334–348. https://doi.org/10.1016/j.tics.2019.01.012

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336.

Faber, N. S., & Häusser, J. A. (2022). Why stress and hunger both increase and decrease prosocial behaviour. *Current Opinion in Psychology*, *44*, 49–57. https://doi.org/10.1016/j.copsyc.2021.08.023

Fanselow, M. S. (2018). The role of learning in threat imminence and defensive behaviors. *Current Opinion in Behavioral Sciences*, *24*, 44–49. https://doi.org/10.1016/j.cobeha.2018.03.003

Fantino, E., & Stolarz-Fantino, S. (2005). Decision-making: Context matters. *Behavioural Processes*, *69*(2), 165–171. https://doi.org/10.1016/j.beproc.2005.02.002

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(1), 58–63. https://doi.org/10.1093/cercor/bhh108

Fernie, G., & Tunney, R. (2013). Learning on the IGT follows emergence of knowledge but not differential somatic activity. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00687

Field, A. (2000). *Discovering Statistics Using SPSS for Windows: Advanced Techniques for Beginners Text*.

Finn. (2019). *Jaywick proposal: Developing Agile Decision Makers*.

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

Fosse, T. H., Skogstad, A., Einarsen, S. V., & Martinussen, M. (2019). Active and passive forms of destructive leadership in a military context: A systematic review and meta-analysis. *European Journal of Work and Organizational Psychology*, *28*(5), 708–722. https://doi.org/10.1080/1359432X.2019.1634550

Franken, I. H., & Muris, P. (2005). Individual differences in decision-making. *Personality and Individual Differences*, *39*(5), 991–998.

Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry? *Personality and Individual Differences*, *17*(6), 791–802. https://doi.org/10.1016/0191-8869(94)90048-5

Frey, R., Hertwig, R., & Rieskamp, J. (2014). Fear shapes information acquisition in decisions from experience. *Cognition*, *132*(1), 90–99. https://doi.org/10.1016/j.cognition.2014.03.009

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Funke, J. (2012). Complex problem solving. *Encyclopedia of the Sciences of Learning (682–685). Heidelberg: Springer*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2233373

Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2025). *cmdstanr: R Interface to 'CmdStan'*. https://github.com/stan-dev/cmdstanr

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *J. R. Stat. Soc. A*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Gansler, D. A., Jerram, M. W., Vannorsdall, T. D., & Schretlen, D. J. (2011). Comparing alternative metrics to assess performance on the Iowa Gambling Task. *Journal of Clinical and Experimental Neuropsychology*, *33*(9), 1040–1048. https://doi.org/10.1080/13803395.2011.596820

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Geng, B., Li, Q., & Varshney, P. K. (2022). Human Decision Making with Bounded Rationality. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5493–5497. https://doi.org/10.1109/ICASSP43922.2022.9747866

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the Future. *Science*, *317*(5843), 1351–1354. https://doi.org/10.1126/science.1144161

Gilhooly, K. J. (1988). *Thinking: Directed, undirected and creative*. Academic Press.

Giovanniello, J., Bravo-Rivera, C., Rosenkranz, A., & Matthew Lattal, K. (2023). Stress, associative learning, and decision-making. *Neurobiology of Learning and Memory*, *204*, 107812. https://doi.org/10.1016/j.nlm.2023.107812

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related

Decision Making. *Cerebral Cortex*, *19*(2), 483–495. https://doi.org/10.1093/cercor/bhn098

Globig, L. K., Witte, K., Feng, G., & Sharot, T. (2021). Under Threat, Weaker Evidence Is Required to Reach Undesirable Conclusions. *Journal of Neuroscience*, *41*(30), 6502–6510. https://doi.org/10.1523/JNEUROSCI.3194-20.2021

Gonzalez, C. (2013). The boundaries of Instance-Based Learning Theory for explaining decisions from experience. *Progress in Brain Research*, *202*, 73–98.

Gonzalez, C. (2022). Learning and Dynamic Decision Making. *Topics in Cognitive Science*, *14*(1), 14–30. https://doi.org/10.1111/tops.12581

Gonzalez, C., Aggarwal, P., Cranford, E. A., & Lebiere, C. (2020). Design of Dynamic and Personalized Deception: A Research Framework and New Insights for Cyberdefense. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, *1834*.

Gonzalez, C., & Dutt, V. (2016). Exploration and exploitation during information search and experimential choice. *Journal of Dynamic Decision Making*, *2*, 2–2.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273–286. https://doi.org/10.1016/j.chb.2004.02.014

Gore, J., Flin, R., Stanton, N., & Wong, B. L. (2015). Applications for naturalistic decision-making. *Journal of Occupational and Organizational Psychology*, *88*(2).

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait-multimethod study of assessment instruments for complex problem solving. *Intelligence*, *41*(5), 579-596. https://doi.org/10.1016/j.intell.2013.07.012

Grillon, C., & Charney, D. R. (2011). In the face of fear: Anxiety sensitizes defensive responses to fearful faces. *Psychophysiology*, *48*(12), 1745-1752. https://doi.org/10.1111/j.1469-8986.2011.01268.x

Grillon, C., Pellowski, M., Merikangas, K. R., & Davis, M. (1997). Darkness facilitates the acoustic startle reflex in humans. *Biological Psychiatry*, *42*(6), Article 6. https://doi.org/10.1016/S0006-3223(96)00466-0

Grogans, S. E., Bliss-Moreau, E., Buss, K. A., Clark, L. A., Fox, A. S., Keltner, D., Cowen, A. S., Kim, J. J., Kragel, P. A., MacLeod, C., Mobbs, D., Naragon-Gainey, K., Fullana, M. A., & Shackman, A. J. (2023). The nature and neurobiology of fear and anxiety: State of the science and opportunities for accelerating discovery. *Neuroscience & Biobehavioral Reviews*, *151*, 105237. https://doi.org/10.1016/j.neubiorev.2023.105237

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*(7), Article 7. https://doi.org/10.1038/nrn3524

Haines, N., Vassileva, J., & Ahn, W.-Y. (2018). The Outcome-Representation Learning Model: A Novel Reinforcement Learning Model of the Iowa Gambling Task. *Cognitive Science*, *42*(8), 2534-2561. https://doi.org/10.1111/cogs.12688

Halfmann, K., Hedgcock, W., Bechara, A., & Denburg, N. L. (2014). Functional neuroimaging of the Iowa Gambling Task in older adults. *Neuropsychology*, *28*(6), 870–880. https://doi.org/10.1037/neu0000120

Hall, L. K. (2012). The Importance of Understanding Military Culture. In *Advances in Social Work Practice with the Military*. Routledge.

Hampton, A. N., Adolphs, R., Tyszka, J. M., & O'Doherty, J. P. (2007). Contributions of the Amygdala to Reward Expectancy and Choice Signals in Human Prefrontal Cortex. *Neuron*, *55*(4), 545–555. https://doi.org/10.1016/j.neuron.2007.07.022

Han, D., Duhachek, A., & Rucker, D. D. (2015). Distinct threats, common remedies: How consumers cope with psychological threat. *Journal of Consumer Psychology*, *25*(4), 531–545.

Han, S., Lerner, J. S., & Keltner, D. (2007). Feelings and consumer decision making: The appraisal-tendency framework. *Journal of Consumer Psychology*, *17*(3), 158–168.

Hannah, S. T., & Sowde, W. J. (2012). Leadership in the Profession of Arms. In M. G. Rumsey (Ed.), *The Oxford Handbook of Leadership* (1st ed., pp. 291–310). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398793.013.0017

Harman, J. L., Zhang, D., & Greening, S. G. (2019). Basic Processes in Dynamic Decision Making: How Experimental Findings About Risk, Uncertainty, and Emotion Can Contribute to Police Decision Making. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02140

Harrell, M. C., & Berglass, N. (2012). Employing America's veterans. *Center for a New American Security*. https://www.jstor.org/stable/pdf/resrep06356.pdf

Harris, K. R., Eccles ,David W., Freeman ,Carlos, & and Ward, P. (2017). 'Gun! Gun! Gun!': An exploration of law enforcement officers' decision-making and coping under stress during actual events. *Ergonomics*, *60*(8), 1112–1122. https://doi.org/10.1080/00140139.2016.1260165

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539.

Highgate, Q., & Schenk, S. (2021). Cognitive flexibility in humans and other laboratory animals. *Journal of the Royal Society of New Zealand*, *51*(1), 97–127. https://doi.org/10.1080/03036758.2020.1784240

Hildebrandt, L. K., McCall, C., Engen, H. G., & Singer, T. (2016). Cognitive flexibility, heart rate variability, and resilience predict fine-grained regulation of arousal during prolonged threat. *Psychophysiology*, *53*(6), 880–890. https://doi.org/10.1111/psyp.12632

Hofstede, G. (1984). *Culture's Consequences: International Differences in Work-Related Values*. SAGE.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hong, R. Y., & Cheung, M. W.-L. (2015). The Structure of Cognitive Vulnerabilities to Depression and Anxiety: Evidence for a Common Core Etiologic Process Based on a Meta-Analytic Review. *Clinical Psychological Science*, *3*(6), 892–912. https://doi.org/10.1177/2167702614553789

Huang, H., Thompson, W., & Paulus, M. P. (2017). Computational Dysfunctions in Anxiety: Failure to Differentiate Signal From Noise. *Biological Psychiatry*, *82*(6), 440–446. https://doi.org/10.1016/j.biopsych.2017.07.007

Hurtubise, J. L., & Howland, J. G. (2017). Effects of stress on behavioral flexibility in rodents. *Neuroscience*, *345*, 176–192. https://doi.org/10.1016/j.neuroscience.2016.04.007

Idzikowski, C., & Baddeley, A. (1983). Waiting in the wings Apprehension, public speaking and performance. *Ergonomics*, *26*(6), 575–583. https://doi.org/10.1080/00140138308963376

Idzikowski, C., & Baddeley, A. (1987). Fear and performance in novice parachutists. *Ergonomics*, *30*(10), 1463–1474. https://doi.org/10.1080/00140138708966039

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, *345*, 12–26. https://doi.org/10.1016/j.neuroscience.2016.03.021

Izquierdo, A., & Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology*, *219*(2), 607–620. https://doi.org/10.1007/s00213-011-2579-7

Jensen, D., Kind, A. J., Morrison, A. S., & Heimberg, R. G. (2014). Intolerance of uncertainty and immediate decision-making in high-risk situations. *Journal of Experimental Psychopathology*, *5*(2), 178–190.

Jones, G. (2008). Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments. *Journal of Economic Behavior & Organization*, *68*(3). https://doi.org/10.1016/j.jebo.2008.06.010

Kahneman, D. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 278.

Kamphuis, W., Gaillard, A. W. K., & Vogelaar, A. L. W. (2011). The Effects of Physical Threat on Team Processes During Complex Task Performance. *Small Group Research*, *42*(6), 700–729. https://doi.org/10.1177/1046496411407522

Kamphuis, W., Gaillard, T., & Vogelaar, A. (2008). *Threat-rigidity effects on planning and decision making in teams: (518442013-261)* [Dataset]. https://doi.org/10.1037/e518442013-261

Kane, H. S., McCall, C., Collins, N. L., & Blascovich, J. (2012). Mere presence is not enough: Responsive support in a virtual world. *Journal of Experimental Social Psychology*, *48*(1), 37–44.

Kark, R., Karazi-Presler, T., & Tubi, S. (2016). Paradox and Challenges in Military Leadership. In C. Peus, S. Braun, & B. Schyns (Eds.), *Monographs in Leadership and Management* (Vol. 8, pp. 157–187). Emerald Group Publishing Limited. https://doi.org/10.1108/S1479-357120160000008031

Kassam, K. S., Koslov, K., & Mendes, W. B. (2009). Decisions under distress: Stress profiles influence anchoring and adjustment. *Psychological Science*, *20*(11), 1394–1399.

Keinan, G. (1987). Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats. *Journal of Personality and Social Psychology*, *52*(3), 639.

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.

Kihlstrom, J. F. (2021). Ecological Validity and “Ecological Validity”. *Perspectives on Psychological Science*, *16*(2), 466–471. https://doi.org/10.1177/1745691620966791

Kim, A. J., Lee, D. S., & Anderson, B. A. (2021). The influence of threat on the efficiency of goal-directed attentional control. *Psychological Research*, *85*, 980–986.

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.

Klein, G. (2008). Naturalistic Decision Making. *Human Factors*, *50*(3), 456–460. https://doi.org/10.1518/001872008X288385

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.

Kornilov, S., Krasnov, E., Kornilova, T., & Chumakova, M. (2015). Individual Differences in Performance on Iowa Gambling Task are Predicted by Tolerance and Intolerance for Uncertainty. *EAPCogSci*, 728–731.

Kossowska, M. (2007). Motivation towards closure and cognitive processes: An individual differences approach. *Personality and Individual Differences*, *43*(8), 2149–2158. https://doi.org/10.1016/j.paid.2007.06.027

Kothgassner, O. D., & Felnhofer, A. (2020). Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? *Annals of the International Communication Association*, *44*(3), 210–218. https://doi.org/10.1080/23808985.2020.1792790

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Krems, J. F. (1995). Cognitive Flexibility and Complex Problem Solving. In *Complex Problem Solving*. Psychology Press.

Kruglanski, A. W. (1990). *Motivations for judging and knowing: Implications for causal attribution.*

Kruglanski, A. W. (2013). *The psychology of closed mindedness*. Psychology Press.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated Closing of the Mind: 'Seizing' and 'Freezing'. *Psychological Review*, *103*(2), 263–283. https://doi.org/10.1037/0033-295X.103.2.263

Krupić, D., Žuro, B., & Corr, P. J. (2021). Anxiety and threat magnification in subjective and physiological responses of fear of heights induced by virtual reality. *Personality and Individual Differences*, *169*, 109720. https://doi.org/10.1016/j.paid.2019.109720

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*.

Kruschke, J. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Kruschke, J., & Meredith, M. (2021). *BEST: Bayesian Estimation Supersedes the t-Test*. https://CRAN.R-project.org/package=BEST

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Ladouceur, R., Gosselin, P., & Dugas, M. J. (2000). Experimental manipulation of intolerance of uncertainty: A study of a theoretical model of worry. *Behaviour Research and Therapy*, *38*(9), 933–941. https://doi.org/10.1016/S0005-7967(99)00133-3

Laposa, J. M., Katz, D. E., Lisi, D. M., Hawley, L. L., Quigley, L., & Rector, N. A. (2022). Longitudinal changes in intolerance of uncertainty and worry severity during CBT for generalized anxiety disorder. *Journal of Anxiety Disorders*, *91*, 102623. https://doi.org/10.1016/j.janxdis.2022.102623

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*, 319–338. https://doi.org/10.1002/smj.2221

Laycock, A., Schofield, G., & McCall, C. (2024). The effects of threat on complex decision-making: Evidence from a virtual environment. *Scientific Reports*, *14*(1), 22637. https://doi.org/10.1038/s41598-024-72812-2

Leary, M. R. (1983). A Brief Version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*(3), 371–375. https://doi.org/10.1177/0146167283093007

LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annual Review of Psychology*, *46*(1), 209–235.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7. https://doi.org/10.1016/j.jmp.2010.08.013

Lehne, M., & Koelsch, S. (2015). Toward a general psychological model of tension and suspense. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00079

Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463. https://doi.org/10.1016/j.neuron.2016.12.040

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*(1).

Levine, S. (2005). Developmental determinants of sensitivity and resistance to stress. *Psychoneuroendocrinology*, *30*(10), 939–946. https://doi.org/10.1016/j.psyneuen.2005.03.013

Levy, I., & Schiller, D. (2021). Neural Computations of Threat. *Trends in Cognitive Sciences*, *25*(2), 151–171. https://doi.org/10.1016/j.tics.2020.11.007

Levy, I., & Schiller, D. (2023). Uncertainty in learning and decision-making: Introduction to the special issue. *Cognitive, Affective & Behavioral Neuroscience*, *23*(3), 459–464. https://doi.org/10.3758/s13415-023-01111-6

Lewinsohn, S., & Mano, H. (1993). Multi-attribute choice and affect: The influence of naturally occurring and manipulated moods on choice processes. *Journal of Behavioral Decision Making*, *6*(1), 33–51.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lieberman, H. R., Bathalon, G. P., Falco, C. M., Morgan, C. A., Niro, P. J., & Tharion, W. J. (2005). The fog of war: Decrements in cognitive performance and mood associated with combat-like stress. *Aviation, Space, and Environmental Medicine*, *76*(7), C7–C14.

Ligneul, R. (2019). Sequential exploration in the Iowa gambling task: Validation of a new computational model in a large dataset of young and old healthy participants. *PLoS Computational Biology*, *15*(6), e1006989. https://doi.org/10.1371/journal.pcbi.1006989

Lindquist, K. A., & Barrett, L. F. (2008). Constructing Emotion: The Experience of Fear as a Conceptual Act. *Psychological Science*, *19*(9), 898–903.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*(5), 331–352. https://doi.org/10.1002/bdm.381

Lipshitz, R., & Strauss, O. (1997). Coping with Uncertainty: A Naturalistic Decision-Making Analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163. https://doi.org/10.1006/obhd.1997.2679

Luce, M. F., Bettman, J. R., & Payne, J. W. (1997). Choice processing in emotionally difficult decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 384.

Luhmann, C. C., Ishida, K., & Hajcak, G. (2011). Intolerance of uncertainty and decisions about delayed, probabilistic rewards. *Behavior Therapy*, *42*(3), 378–386.

Luxem, K., Sun, J. J., Bradley, S. P., Krishnan, K., Yttri, E., Zimmermann, J., Pereira, T. D., & Laubach, M. (2023). Open-source tools for behavioral video analysis: Setup, methods, and best practices. *eLife*, *12*, e79305. https://doi.org/10.7554/eLife.79305

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(45), 16075–16080. https://doi.org/10.1073/pnas.0406666101

Maner, J., Richey, J., Cromer, K., & Mallott, M. (2007). Dispositional anxiety and risk-avoidant decision-making. *Personality and Individual Differences*, *42*(4), Article 4.

Männiste, T., Pedaste, M., & Schimanski, R. (2019). Review of instruments measuring decision making performance in military tactical level battle situation context. *Military Psychology*, *31*(5), 397–411. https://doi.org/10.1080/08995605.2019.1645538

Matthews, G., Davies, D. R., Stammers, R. B., & Westerman, S. J. (2000). *Human performance: Cognition, stress, and individual differences*. Psychology Press.

Mayer, M. (2023). *confintr: Confidence Intervals*. https://CRAN.R-project.org/package=confintr

Mazzei, M. J., DeBode, J., Gangloff, K. A., & Song, R. (2025). Old Habits Die Hard: A Review and Assessment of the Threat-Rigidity Literature. *Journal of Management*, *51*(6), 2154–2181. https://doi.org/10.1177/01492063241286493

McCall, C., & Blascovich, J. (2009). How, When, and Why to Use Digital Experimental Virtual Environments to Study Social Behavior. *Social and Personality Psychology Compass*, *3*(5), 744–758. https://doi.org/10.1111/j.1751-9004.2009.00195.x

McCall, C., Hildebrandt, L. K., Bornemann, B., & Singer, T. (2015). Physiophenomenology in retrospect: Memory reliably reflects physiological arousal during a prior threatening experience. *Consciousness and Cognition*, *38*, 60–70. https://doi.org/10.1016/j.concog.2015.09.011

McCall, C., & Laycock, A. (2024). Moving Targets: Experiences of threat and uncertainty on the front line. *The University of York. In Review*.

McCall, C., Schofield, G., Halgarth, D., Blyth, G., Laycock, A., & Palombo, D. J. (2022). The underwood project: A virtual environment for eliciting ambiguous threat. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-02002-3

McCloskey, M., & Cohen, N. J. (1989). Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 24, pp. 109–165). Academic Press. https://doi.org/10.1016/S0079-7421(08)60536-8

McElreath, R. (2016). rethinking: Statistical Rethinking book package. *R Package Version*, *1*, 59.

McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315372495

McEvoy, P. M., & Erceg-Hurn, D. M. (2016). The search for universal transdiagnostic and trans-therapy change processes: Evidence for intolerance of uncertainty. *Journal of Anxiety Disorders*, *41*, 96–107.

McEvoy, P. M., Hyett, M. P., Shihata, S., Price, J. E., & Strachan, L. (2019). The impact of methodological and measurement factors on transdiagnostic associations with intolerance of uncertainty: A meta-analysis. *Clinical Psychology Review*, *73*, 101778. https://doi.org/10.1016/j.cpr.2019.101778

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*, 191-215. https://doi.org/10.1037/dec0000033

Meredith, M., & Kruschke, J. (2021). *Bayesian Estimation Supersedes the t-test*.

Miller, G. A. (1981). Trends and debates in cognitive psychology. *Cognition*, *10*(1), 215–225. https://doi.org/10.1016/0010-0277(81)90049-4

Milosevic, I., & McCabe, R. (2015). *The Psychology of Irrational Fear, The Fight or Flight Response*. Bloomsbury Publishing USA.

Mkrtchian, A., Valton, V., & Roiser, J. P. (2021). Reliability of decision-making and reinforcement learning computational parameters. *BioRxiv*, 2021–06.

Mobbs, D., Adolphs, R., Fanselow, M. S., Barrett, L. F., LeDoux, J. E., Ressler, K., & Tye, K. M. (2019). Viewpoints: Approaches to defining and investigating fear. *Nature Neuroscience*, *22*(8), 1205–1216. https://doi.org/10.1038/s41593-019-0456-6

Mobbs, D., Marchant, J. L., Hassabis, D., Seymour, B., Tan, G., Gray, M., Petrovic, P., Dolan, R. J., & Frith, C. D. (2009). From Threat to Fear: The Neural Organization of Defensive Fear Systems in Humans. *Journal of Neuroscience*, *29*(39), 12236–12243. https://doi.org/10.1523/JNEUROSCI.2378-09.2009

Molins, F., Ayuso, C., & Serrano, M. Á. (2021). Emotional stress & decision-making: An emotional stressor significantly reduces loss aversion. *Stress*, *24*(6), 780–786.

Molins, F., Serrano, M. Á., & Alacreu-Crespo, A. (2021). Early stages of the acute physical stress response increase loss aversion and learning on decision making: A Bayesian approach. *Physiology & Behavior*, *237*, 113459. https://doi.org/10.1016/j.physbeh.2021.113459

Moreno-Padilla, M., Alacreu-Crespo, A., Guillaume, S., & Reyes del Paso, G. A. (2022). The outcome-representation learning model: Impairments in decision-making in adolescents with excess weight. *Current Psychology*. https://doi.org/10.1007/s12144-022-03299-1

Morgado, P., Sousa, N., & Cerqueira, J. j. (2015). The impact of stress in decision making in the context of uncertainty. *Journal of Neuroscience Research*, *93*(6), 839–847. https://doi.org/10.1002/jnr.23521

Moser, I., Chiquet, S., Strahm, S. K., Mast, F. W., & Bergamin, P. (2020). Group Decision-Making in Multi-User Immersive Virtual Reality. *Cyberpsychology, Behavior, and Social Networking*, *23*(12), 846–853. https://doi.org/10.1089/cyber.2020.0065

Muris, P., Mayer, B., van Lint, C., & Hofman, S. (2008). Attentional control and psychopathological symptoms in children. *Personality and Individual Differences*, *44*(7), 1495–1505.

Murphy, F. C., Michael, A., Robbins, T. W., & Sahakian, B. J. (2003). Neuropsychological impairment in patients with major depressive disorder: The effects of feedback on task performance. *Psychological Medicine*, *33*(3), 455–467. https://doi.org/10.1017/S0033291702007018

Mussey, J. L., Travers, B. G., Klinger, L. G., & Klinger, M. R. (2015). Decision-Making Skills in ASD: Performance on the Iowa Gambling Task. *Autism Research*, *8*(1), 105–114. https://doi.org/10.1002/aur.1429

Nan, J., Jaiswal, S., Ramanathan, D., Withers, M. C., & Mishra, J. (2025). Climate trauma from wildfire exposure impacts cognitive decision-making. *Scientific Reports*, *15*(1), 11992. https://doi.org/10.1038/s41598-025-94672-0

Noonan, M. P., Walton, M. E., Behrens, T. E. J., Sallet, J., Buckley, M. J., & Rushworth, M. F. S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, *107*(47), 20547–20552. https://doi.org/10.1073/pnas.1012246107

Nowacki, J., Heekeren, H. R., Deuter, C. E., Joerißen, J. D., Schröder, A., Otte, C., & Wingenfeld, K. (2019). Decision making in response to physiological and combined physiological and psychosocial stress. *Behavioral Neuroscience*, *133*(1), 59–67. https://doi.org/10.1037/bne0000288

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory New York. *NY: McGraw-Hill*.

Oberauer, K. (2019). Working memory and attention—A conceptual analysis and review. *Journal of Cognition*, *2*. https://doi.org/10.5334/joc.58

Oberdörfer, S., Heidrich, D., Birnstiel, S., & Latoschik, M. E. (2021). Enchanted by Your Surrounding? Measuring the Effects of Immersion and Design of Virtual Environments on Decision-Making. *Frontiers in Virtual Reality*, *2*. https://www.frontiersin.org/articles/10.3389/frvir.2021.679277

O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form.

International Journal of Human-Computer Studies, *112*, 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

Okada, I. (2020). A Review of Theoretical Studies on Indirect Reciprocity. *Games*, *11*(3), 27. https://doi.org/10.3390/g11030027

Omodei, M. M., & Wearing, A. J. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments, & Computers*, *27*(3), 303–316. https://doi.org/10.3758/BF03200423

Pahng, P. H., & Kang, S. M. (2023). Voice vs. silence: The role of cognitive appraisal of and emotional response to stressors. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1079244

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

Paret, C., & Bublatzky, F. (2020). Threat rapidly disrupts reward reversal learning. *Behaviour Research and Therapy*, *131*, 103636. https://doi.org/10.1016/j.brat.2020.103636

Park, L. E., Naidu, E., Lemay, E. P., Canning, E. A., Ward, D. E., Panlilio, Z., & Vessels, V. (2023). Chapter Three—Social evaluative threat across individual, relational, and collective selves. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (Vol. 68, pp. 139–222). Academic Press. https://doi.org/10.1016/bs.aesp.2023.03.001

Pasion, R., Gonçalves, A. R., Fernandes, C., Ferreira-Santos, F., Barbosa, F., & Marques-Teixeira, J. (2017). Meta-Analytic Evidence for a Reversal Learning Effect on the Iowa

Gambling Task in Older Adults. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.01785

Pattni, N., Arzola, C., Malavade, A., Varmani, S., Krimus, L., & Friedman, Z. (2019). Challenging authority and speaking up in the operating room environment: A narrative synthesis. *British Journal of Anaesthesia*, *122*(2), 233–244. https://doi.org/10.1016/j.bja.2018.10.056

Pelaccia, T., Tardif, J., Triby, E., Ammirati, C., Bertrand, C., Dory, V., & Charlin, B. (2014). How and When Do Expert Emergency Physicians Generate and Evaluate Diagnostic Hypotheses? A Qualitative Study Using Head-Mounted Video Cued-Recall Interviews. *Annals of Emergency Medicine*, *64*(6), 575–585. https://doi.org/10.1016/j.annemergmed.2014.05.003

Penney, G., Launder, D., Cuthbertson, J., & Thompson, M. B. (2022). Threat assessment, sense making, and critical decision-making in police, military, ambulance, and fire services. *Cognition, Technology & Work*, *24*(3), 423–439. https://doi.org/10.1007/s10111-022-00694-3

Perandrés-Gómez, A., Navas, J. F., van Timmeren, T., & Perales, J. C. (2021). Decision-making (in)flexibility in gambling disorder. *Addictive Behaviors*, *112*, 106534. https://doi.org/10.1016/j.addbeh.2020.106534

Peterson, S. M., Furuichi, E., & Ferris, D. P. (2018). Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading. *PLOS ONE*, *13*(7), e0200306. https://doi.org/10.1371/journal.pone.0200306

Pezzulo, G., & Rigoli, F. (2011). The value of foresight: How prospection affects decision-making. *Frontiers in Neuroscience*, *5*, 79.

Philpot, R., & Levine, M. (2022). Evacuation Behavior in a Subway Train Emergency: A Video-based Analysis. *Environment and Behavior*, *54*(2), 383–411. https://doi.org/10.1177/00139165211031193

Philpot, R., Levine, M., Betts, C., Nelson, M., Duck, E., & Power, N. (2025). Observing interoperability: A behavioural framework and analysis of multi-agency interactions in the UK emergency services. *Safety Science*, *191*, 106947. https://doi.org/10.1016/j.ssci.2025.106947

Pilin, M. A. (2021). The past of predicting the future: A review of the multidisciplinary history of affective forecasting. *History of the Human Sciences*, *34*(3–4), 290–306. https://doi.org/10.1177/0952695120976330

Prather, R. W., Benitez, V. L., Brooks, L. K., Dancy, C. L., Dilworth-Bart, J., Dutra, N. B., Faison, M. O., Figueroa, M., Holden, L. R., Johnson, C., Medrano, J., Miller-Cotto, D., Matthews, P. G., Manly, J. J., & Thomas, A. K. (2022). What Can Cognitive Science Do for People? *Cognitive Science*, *46*(6), e13167. https://doi.org/10.1111/cogs.13167

Preston, S. D., Buchanan, T. W., Stansfield, R. B., & Bechara, A. (2007). Effects of anticipatory stress on decision making in a gambling task. *Behavioral Neuroscience*, *121*, 257–263. https://doi.org/10.1037/0735-7044.121.2.257

Prolific. (2014). *Prolific. London, UK. Version: 2022.* https://www.prolific.com/

Putwain, D. (2007). Researching academic stress and anxiety in students: Some methodological considerations. *British Educational Research Journal*, *33*(2), 207–219. https://doi.org/10.1080/01411920701208258

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, *23*(1), 1–14. https://doi.org/10.1002/bdm.681

Ranes, B., Long, C. P., Traynham, S., & Hayes, A. (2017). The Influence of Combat Experience on Psychologically Healthy Soldiers' Attentiveness to Environmental Threats. *Military Medicine*, *182*(7), e1787–e1793. https://doi.org/10.7205/MILMED-D-16-00261

Rescorla, R. A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. *Classical Conditioning, Current Research and Theory*, *2*, 64–69.

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. https://CRAN.R-project.org/package=psych

Reverberi, C., Pischedda, D., Mantovani, M., Haynes, J.-D., & Rustichini, A. (2022). Strategic complexity and cognitive skills affect brain response in interactive decision-making. *Scientific Reports*, *12*(1), 15896. https://doi.org/10.1038/s41598-022-17951-0

Rickles, D., Hawe, P., & Shiell, A. (2007). A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health*, *61*(11), 933–937. https://doi.org/10.1136/jech.2006.054254

Robinson, O., Vytal, K., Cornwell, B., & Grillon, C. (2013). The impact of anxiety upon cognition: Perspectives from human threat of shock studies. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00203

Roets, A., & Hiel, A. van. (2008). Why some hate to dilly-dally and others do not: The arousal-invoking capacity of decision-making for low-and high-scoring need for closure individuals. *Social Cognition*, *26*(3), 333–346.

Roets, A., Kruglanski, A. W., Kossowska, M., Pierro, A., & Hong, Y. (2015). The motivated gatekeeper of our minds: New directions in need for closure theory and research. In *Advances in experimental social psychology* (Vol. 52, pp. 221–283). Elsevier.

Roets, A., & Van Hiel, A. (2007). Separating Ability From Need: Clarifying the Dimensional Structure of the Need for Closure Scale. *Personality and Social Psychology Bulletin*, *33*(2), 266–280. https://doi.org/10.1177/0146167206294744

Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and Individual Differences*, *50*(1), 90–94.

Rolls, E. T. (2007). *Memory, attention, and decision-making: A unifying computational neuroscience approach*.

Rosen, J. B., & Schulkin, J. (1998). From normal fear to pathological anxiety. *Psychological Review*, *105*(2), 325–350. https://doi.org/10.1037/0033-295X.105.2.325

Rovetta, A. (2020). Raiders of the Lost Correlation: A Guide on Using Pearson and Spearman Coefficients to Detect Hidden Correlations in Medical Sciences. *Cureus*, *12*(11), e11794. https://doi.org/10.7759/cureus.11794

Runyon, R. P., Coleman, K. A., & Pittenger, D. J. (2000). *Fundamentals of behavioral statistics, 9th ed* (pp. xxii, 634). McGraw-Hill.

Sankaran, S., Szumowska, E., & Kossowska, M. (2017). "When the going gets tough, the tough get going": Motivation towards closure and effort investment in the performance of cognitive tasks. *Motivation and Emotion*, *41*(3), 308–321. https://doi.org/10.1007/s11031-017-9613-y

Savine, A. C., Beck, S. M., Edwards, B. G., Chiew, K. S., & Braver, T. S. (2010). Enhancement of cognitive control by approach and avoidance motivational states. *Cognition and Emotion*, *24*(2), 338–356.

Sayfulina, K. E., Kozunova, G. L., Medvedev, V. A., Rytikova, A. M., & Chernyshev, B. V. (2020). Decision making under uncertainty: Exploration and exploitation. *Journal of Modern Foreign Psychology*, *9*(2), 93–106.

Schmitt, J., & Klein, G. (1999). *A recognitional planning model*. Klein Associates Inc Fairborn OH.

Schneider, B., & Koenigs, M. (2017). Human lesion studies of ventromedial prefrontal cortex. *Neuropsychologia*, *107*, 84–93. https://doi.org/10.1016/j.neuropsychologia.2017.09.035

Schulz, E., Franklin, N. T., & Gershman, S. J. (2020). Finding structure in multi-armed bandits. *Cognitive Psychology*, *119*, 101261. https://doi.org/10.1016/j.cogpsych.2019.101261

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, *55*, 7–14.

Schwabe, L., Oitzl, M. S., Philippsen, C., Richter, S., Bohringer, A., Wippich, W., & Schachinger, H. (2007). Stress modulates the use of spatial versus stimulus-response learning strategies in humans. *Learning & Memory*, *14*(1-2), 109–116.

Schwabe, L., & Wolf, O. T. (2011). Stress-induced modulation of instrumental behavior: From goal-directed to habitual control of action. *Behavioural Brain Research*, *219*(2), 321–328. https://doi.org/10.1016/j.bbr.2010.12.038

Senftleben, U., Schoemann, M., Rudolf, M., & Scherbaum, S. (2021). To stay or not to stay: The stability of choice perseveration in value-based decision making. *Quarterly Journal of Experimental Psychology*, *74*(1), 199–217.

Sergiou, C.-S., Gerstner, D., Nee, C., Elffers, H., & van Gelder, J.-L. (2024). Virtual reality-based retrospective think aloud (VR-RTA): A novel method for studying offender decision-making. *Crime Science*, *13*(1), 39. https://doi.org/10.1186/s40163-024-00236-4

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248–1284.

Shortland, N., Alison, L., & Barrett-Pink, C. (2018). Military (in)decision-making process: A psychological framework to examine decision inertia in military operations. *Theoretical Issues in Ergonomics Science*, *19*(6), 752–772. https://doi.org/10.1080/1463922X.2018.1497726

Shortland, N., Alison, L., & Thompson, L. (2020). Military maximizers: Examining the effect of individual differences in maximization on military decision-making. *Personality and Individual Differences*, *163*, 110051. https://doi.org/10.1016/j.paid.2020.110051

Sidis, B. (1922). *Nervous ills: Their cause and cure*. RG Badger.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology* (pp. x, 428). Basic Books.

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review*, *3*(3), 314–321. https://doi.org/10.3758/BF03210755

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, 103535. https://doi.org/10.1016/j.artint.2021.103535

Simoens, J., Verguts, T., & Braem, S. (2024). Learning environment-specific learning rates. *PLOS Computational Biology*, *20*(3), e1011978. https://doi.org/10.1371/journal.pcbi.1011978

Simon, H. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Simon, L., Jiryis, T., & Admon, R. (2021). Now or Later? Stress-Induced Increase and Decrease in Choice Impulsivity Are Both Associated with Elevated Affective and Endocrine Responses. *Brain Sciences*, *11*(9), Article 9. https://doi.org/10.3390/brainsci11091148

Simonetti, L. C. R. M., Marine, I. A., & Tripodi, P. (2020). Automation and the Future of Command and Control: The End of Auftragstaktik? *Journal of Advanced Military Studies*, *11*(1), 127–146.

Simonovic, B., Stupple, E. J., Gale, M., & Sheffield, D. (2017). Stress and risky decision making: Cognitive reflection, emotional learning or both. *Journal of Behavioral Decision Making*, *30*(2), 658–665.

Simonovic, B., Stupple, E. J. N., Gale, M., & Sheffield, D. (2018). Performance Under Stress: An Eye-Tracking Investigation of the Iowa Gambling Task (IGT). *Frontiers in Behavioral Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnbeh.2018.00217

Simpkin, A., & Schwartzstein, R. (2016). Tolerating uncertainty—The next medical revolution? *New England Journal of Medicine*, *375*(18). https://ora.ox.ac.uk/objects/uuid:29cef5e2-0680-40ea-b511-4911723b7db2/files/m3d9b465139a0af33ebfad4679dd08a92

Sladky, R., Baldinger, P., Kranz, G. S., Tröstl, J., Höflich, A., Lanzenberger, R., Moser, E., & Windischberger, C. (2013). High-resolution functional MRI of the human amygdala at 7T. *European Journal of Radiology*, *82*(5), 728–733. https://doi.org/10.1016/j.ejrad.2011.09.025

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557.

Smith, M. (2020). Educating risk: How fear of failure is stifling creative practice within Higher Education. *EDUCATING ANIMATORS Academic Conference 2019 Teaching the World's Most Expressive Art Form*. https://lau.repository.guildhe.ac.uk/id/eprint/17684/

Smith, R., & Sarason, I. G. (1975). Social anxiety and the evaluation of negative interpersonal feedback. *Journal of Consulting and Clinical Psychology*, *43*(3), 429.

Soeters, J. (2018). Organizational Cultures in the Military. In G. Caforio & M. Nuciari (Eds.), *Handbook of the Sociology of the Military* (pp. 251–272). Springer International Publishing. https://doi.org/10.1007/978-3-319-71602-2_13

Sokol-Hessner, P., Raio, C. M., Gottesman, S. P., Lackovic, S. F., & Phelps, E. A. (2016). Acute stress does not affect risky monetary decision-making. *Neurobiology of Stress*, *5*, 19–25. https://doi.org/10.1016/j.ynstr.2016.10.003

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, *20*(10), Article 10. https://doi.org/10.1038/s41583-019-0180-y

Sporrer, J. K., Brookes, J., Hall, S., Zabbah, S., Hernandez, U. D. S., & Bach, D. R. (2023). Functional sophistication in human escape. *iScience*, *26*(11). https://doi.org/10.1016/j.isci.2023.108240

Stan Development Team. (2023). *Stan Modeling Language Users Guide and Reference Manual, Version 2.32*. https://mc-stan.org/docs/2_32/reference-manual/index.html

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1228–1248. https://doi.org/10.1016/j.neubiorev.2012.02.003

Starcke, K., & Brand, M. (2016). Effects of stress on decisions under uncertainty: A meta-analysis. *Psychological Bulletin*, *142*, 909–933. https://doi.org/10.1037/bul0000060

Starita, F., Kroes, M. C. W., Davachi, L., Phelps, E. A., & Dunsmoor, J. E. (2019). Threat learning promotes generalization of episodic memory. *Journal of Experimental Psychology: General*, *148*, 1426–1434. https://doi.org/10.1037/xge0000551

Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat Rigidity Effects in Organizational Behavior: A Multilevel Analysis. *Administrative Science Quarterly*, *26*(4), 501–524. https://doi.org/10.2307/2392337

Steimer, T. (2002). The biology of fear- and anxiety-related behaviors. *Dialogues in Clinical Neuroscience*, *4*(3), 231–249.

Steingroever, H., Wetzels, R., Horstmann, A., Neumann, J., & Wagenmakers, E.-J. (2013). Performance of healthy participants on the Iowa Gambling Task. *Psychological Assessment*, *25*(1), 180.

Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2013). Validating the PVL-Delta model for the Iowa gambling task. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00898

Suarez-Jimenez, B., Balderston, N. L., Bisby, J. A., Leshin, J., Hsiung, A., King, J. A., Pine, D. S., Burgess, N., Grillon, C., & Ernst, M. (2021). Location-dependent threat and associated neural abnormalities in clinical anxiety. *Communications Biology*, *4*(1), Article 1. https://doi.org/10.1038/s42003-021-02775-x

Sullivan-Toole, H., Haines, N., Dale, K., & Olino, T. M. (2022). Enhancing the Psychometric Properties of the Iowa Gambling Task Using Full Generative Modeling. *Computational Psychiatry*, *6*(1), Article 1. https://doi.org/10.5334/cpsy.89

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, Not Consistency, Characterizes Flashbulb Memories. *Psychological Science*, *14*(5), 455–461.

Tashjian, S. M., Fedrigo, V., Molapour, T., Mobbs, D., & Camerer, C. F. (2022). Physiological Responses to a Haunted-House Threat Experience: Distinct Tonic and Phasic Effects. *Psychological Science*, *33*(2), 236–248. https://doi.org/10.1177/09567976211032231

Thomas, A. K., & Wulff, A. N. (2024). What the Acute Stress Response Suggests about Memory. *Topics in Cognitive Science*, *16*(4), 691–706. https://doi.org/10.1111/tops.12664

Ting, C.-C., Palminteri, S., Lebreton, M., & Engelmann, J. B. (2022). The elusive effects of incidental anxiety on reinforcement-learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(5), 619–642. https://doi.org/10.1037/xlm0001033

Tipples, J., Young, A. W., Quinlan, P., Broks, P., & Ellis, A. W. (2002). Searching for threat. *The Quarterly Journal of Experimental Psychology Section A*, *55*(3), 1007–1026. https://doi.org/10.1080/02724980143000659

Toda, M. (1962). The Design of a Fungus-Eater: A Model of Human Behavior in an Unsophisticated Environment. *Behavioral Science*, *7*(2), 164–183.

Turkle, S. (1984). *The Second Self: Computers and the Human Spirit. Granada*. London, Simon & Schuster, New York.

Tuzsus, D., Pappas, I., & Peters, J. (2024). A meta reinforcement learning account of behavioral adaptation to volatility in recurrent neural networks. *bioRxiv*, 2024–05.

Tyszka, T., Nancy, E., Janusz, R., & Ervin, S. (1989). Information and evaluation processes in decision making. *Social and Moral Values: Individual and Societal Perspectives*, 175–193.

Uddin, L. Q. (2021). Cognitive and behavioural flexibility: Neural mechanisms and clinical considerations. *Nature Reviews Neuroscience*, *22*(3), Article 3. https://doi.org/10.1038/s41583-021-00428-w

Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience*, *10*(6), Article 6. https://doi.org/10.1038/nrn2647

van Ast, V. A., Spicer, J., Smith, E. E., Schmer-Galunder, S., Liberzon, I., Abelson, J. L., & Wager, T. D. (2016). Brain Mechanisms of Social Threat Effects on Working Memory. *Cerebral Cortex*, *26*(2), 544–556. https://doi.org/10.1093/cercor/bhu206

Van Den Heuvel, C., Alison, L., & Crego, J. (2012). How Uncertainty and Accountability can Derail Strategic 'Save Life' Decisions in Counter-Terrorism Simulations: A Descriptive Model of Choice Deferral and Omission Bias. *Journal of Behavioral Decision Making*, *25*(2), 165–187. https://doi.org/10.1002/bdm.723

van der Heiden, C., Muris, P., & van der Molen, H. T. (2012). Randomized controlled trial on the effectiveness of metacognitive therapy and intolerance-of-uncertainty therapy for generalized anxiety disorder. *Behaviour Research and Therapy*, *50*(2), 100–109. https://doi.org/10.1016/j.brat.2011.12.005

Västfjäll, D., Slovic, P., Burns, W. J., Erlandsson, A., Koppel, L., Asutay, E., & Tinghög, G. (2016). The Arithmetic of Emotion: Integration of Incidental and Integral Affect in Judgments and Decisions. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2016.00325

Vytal, K., Cornwell, B., Arkin, N., & Grillon, C. (2012). Describing the interplay between anxiety and cognition: From impaired performance under low cognitive load to reduced anxiety under high load. *Psychophysiology*, *49*(6), 842–852. https://doi.org/10.1111/j.1469-8986.2012.01358.x

Wake, S., Wormwood, J., & Satpute, A. B. (2020). The influence of fear on risk taking: A meta-analysis. *Cognition and Emotion*, *34*(6), 1143–1159. https://doi.org/10.1080/02699931.2020.1731428

Walia, Y., & Gupta, R. K. (2025). The role of threat sensitivity as a mediator in the relationship between cognitive conflict and risk-taking behavior in young adults. *Cognitive Research: Principles and Implications*, *10*(1), 62. https://doi.org/10.1186/s41235-025-00673-y

Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H., & Rushworth, M. F. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron*, *65*(6), 927–939.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*(4), 448–457. https://doi.org/10.1037/h0027806

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049.

Weiss, E. O., Kruppa, J. A., Fink, G. R., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2021). Developmental differences in probabilistic reversal learning: A computational modeling approach. *Frontiers in Neuroscience*, *14*, 536596.

Wemm, S. E., & Wulfert, E. (2017). Effects of Acute Stress on Decision Making. *Applied Psychophysiology and Biofeedback*, *42*(1), 1–12. https://doi.org/10.1007/s10484-016-9347-8

White, H. A. (2022). Need for cognitive closure predicts stress and anxiety of college students during COVID-19 pandemic. *Personality and Individual Differences*, *187*, 111393. https://doi.org/10.1016/j.paid.2021.111393

Wichary, S., Mata, R., & Rieskamp, J. (2016). Probabilistic Inferences Under Emotional Stress: How Arousal Affects Decision Processes. *Journal of Behavioral Decision Making*, *29*(5), 525–538. https://doi.org/10.1002/bdm.1896

Williams, B. (2010). Heuristics and biases in military decision making. *Military Review*, *90*(5), 40.

Williams, B., Rodriguez-Sobstel, C., FitzGibbon, L., Morriss, J., & Christakou, A. (2024). *The influence of trait intolerance of uncertainty on behavioural flexibility* (p. 2024.10.30.621039). bioRxiv. https://doi.org/10.1101/2024.10.30.621039

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Wittmann, M. K., Scheuplein, M., Gibbons, S. G., & Noonan, M. P. (2023). Local and global reward learning in the lateral frontal cortex show differential development during human adolescence. *PLOS Biology*, *21*(3), e3002010. https://doi.org/10.1371/journal.pbio.3002010

Worthy, D. A., Pang, B., & Byrne, K. A. (2013). Decomposing the roles of perseveration and expected value representation in models of the Iowa gambling task. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00640

Xu, C. Y., Dan, O., Jia, R., Wertheimer, E., Chawla, M., Fuhrmann-Alpert, G., Fried, T., & Levy, I. (2024). Quantitative vs. Qualitative Outcomes: A Longitudinal Study of Risk and Ambiguity in Monetary and Medical Decision-Making. *Research Square*, rs.3.rs-4249490. https://doi.org/10.21203/rs.3.rs-4249490/v1

Yechiam, E. (2020). Robust consistency of choice switching in decisions from experience. *Judgment and Decision Making*, *15*(1), 74–81.

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science*, *16*(12), 973–978.

Yoo, B., Donthu, N., & Lenartowicz, T. (2011). Measuring Hofstede's Five Dimensions of Cultural Values at the Individual Level: Development and Validation of CVSCALE. *Journal of International Consumer Marketing*, *23*(3–4), 193–210. https://doi.org/10.1080/08961530.2011.578059

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

Zeif, D., & Yechiam, E. (2020). Autism is not associated with poor or enhanced performance on the Iowa Gambling Task: A Meta-Analysis. *Neuroscience & Biobehavioral Reviews*, *113*, 440–447. https://doi.org/10.1016/j.neubiorev.2020.04.016

Zhang, Y., Li, Y., & Mai, X. (2023). Fear of negative evaluation modulates the processing of social evaluative feedback with different valence and contexts. *Cerebral Cortex*, *33*(8), 4927–4938. https://doi.org/10.1093/cercor/bhac390