The Cognitive and Computational Mechanisms of Risk-taking Behavior

by

Juliana Katharina Sporrer

A dissertation submitted for the degree of

Doctor of Philosophy (PhD)

in

Computational Psychiatry

Max Planck UCL Centre for Computational Psychiatry
and Ageing Research, & the Wellcome Centre for Human Neuroimaging
UCL Queen Square Institute of Neurology
University College London

March, 2025

Declaration

I, Juliana Katharina Sporrer, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

March, 2025

Abstract

Every day, individuals make risky decisions—whether ignoring a chest pain, investing in a friend's venture, or stepping outside during a storm. These decisions raise important questions: How do individuals behave under risk? How much risk are they willing to accept? And what factors drive their choices? In this thesis, I investigate the cognitive and computational characteristics of risk-taking behavior using increasingly complex and naturalistic paradigms, focusing on how internal traits and external contexts modulate decisions. In the first study (Chapter 2), I demonstrate that risk aversion is enhanced in anxious depressed patients after a worry induction, compared to baseline or depressed patients. This finding underscores the transient nature of risk preferences in generalized anxiety and suggests that these preferences may be driven by anxiety symptoms rather than causing them. The second study (Chapter 3), conducted with large online samples, reveals that transdiagnostic compulsivity is the primary predictor of cautious behavior in an approach-avoidance conflict task set in a risky foraging scenario. This challenges traditional views on these tasks, which have been extensively used to assess the effects of anxiolytic agents. The final study (Chapter 4) utilizes fully immersive virtual reality to examine naturalistic escape decisions under risk of predation by bio-realistic threats. The results show that escape decisions can be dynamically updated, depend on personal and threat characteristics; and are implemented to optimize secondary goals. I also demonstrate that different types of threat-related behavior rely on distinct computational mechanisms. These findings indicate that escape decisions are not instinctive but depend on flexible computational mechanisms that integrate both internal and

external factors. Together, these studies converge to highlight the complexity and flexibility of human risk-taking behavior. By bridging clinical, computational, and ecologically valid approaches, this thesis advances our mechanistic understanding of how humans navigate risk and offers new insights into the cognitive processes that underpin these decisions.

Impact Statement

Risk-taking is a core aspect of human decision-making, influencing domains as diverse as mental health or economical stability. Understanding its drivers is essential for developing interventions that encourage adaptive behaviors and reduce maladaptive ones. This thesis examines how internal traits and external contexts influence risky decision-making, integrating insights from clinical and computational psychiatry, neuroscience, and behavioral economics.

Scientific Impact

In a first approach, we employ a computational prospect-theoretic model to assess how induced worry and anxiety symptoms influence risk preferences in an economic decision-making task (cf. Chapter 2). Our results indicated that risk aversion was heightened in anxious-depressed patients after worry induction, but not at baseline or in patients with depression alone. This highlights the importance of considering state-dependent processes in risk-taking behavior.

In a second approach, we use a cross-species validated approach-avoidance conflict (AAC) task to explore how transdiagnostic compulsivity and other psychiatric traits influence cautious behavior (cf. Chapter 3). Transdiagnostic compulsivity emerges as the strongest predictor, emphasizing the need to prioritize transdiagnostic dimensions over traditional psychiatric categories for understanding approach-avoidance behaviors.

In a third approach, we utilize immersive virtual reality to investigate human escape responses to biologically relevant threats (cf. Chapter 4). The results demonstrate how external threat characteristics and internal factors shape escape behaviors, providing a more nuanced understanding of defensive responses. The incorporation of virtual reality (VR) underscores the potential of innovative tools to investigate real-world decision-making in controlled, dynamic environments.

Clinical Impact

This thesis has significant implications for mental health understanding and intervention.

Understanding the distinct impacts of anxiety and depression on risk-taking (cf. Chapter 2) provides valuable insights for developing therapeutic strategies. For instance, interventions for generalized anxiety disorder might focus on reducing sensitivity to uncertainty rather than solely addressing fear of negative outcomes. In contrast, such approaches may not be universally effective for major depressive disorder.

On the other hand, the identification of compulsivity as a key driver of cautious behavior in AAC tasks (cf. Chapter 3) indicates that therapeutic strategies should also consider compulsivity when addressing avoidance behaviors. Although AAC tests have been extensively used to characterize the effects of anxiolytic agents and probe neural circuitry related to anxiety, they might not specifically relate to self-reported anxiety. Therefore, their validity for etiology research related to anxiety disorder in healthy humans should be questioned.

Finally, the association between spider phobia and avoidance behaviors in VR environments (cf. Chapter 4) highlights the potential of VR-based exposure therapies. By simulating realistic threats in a controlled setting, VR could enhance patient engagement, offering a safe space to confront and overcome fears.

Academic Contribution

This thesis aligns with open, reproducible science, making all data and code available online. By replicating prior studies and expanding on established findings, the research contributes to a robust and cumulative body of knowledge. It deepens our understanding of how internal and external factors converge to shape risk-related behaviors across diverse contexts, offering a reliable foundation for future exploration.

Acknowledgements

Words cannot express how sincerely thankful I am to my supervisor, *Doktorvater*, Professor Dominik Bach. His unwavering commitment to nurturing intellectual curiosity and his exceptional dedication to scientific precision have been pivotal in my academic growth. Beyond the knowledge I have gained, I am profoundly grateful for his consistent encouragement, support, and wisdom, which have shaped not only my academic journey but also my personal growth. I truly could not have asked for a better mentor.

Dominik's remarkable ability to foster an inspiring and collaborative lab environment—despite the challenges posed by a global pandemic and a lab relocation—has left a lasting impression on me. I feel incredibly fortunate to have worked with such supportive and kind lab mates, whose generosity, both scientifically and personally, has been invaluable. I would especially like to thank Sajjad, Josie, Ulises, Ted, and Marina for the many memorable years we shared at UCL. I am equally grateful to every lab member at the University of Bonn, who, despite the physical distance, have always felt like an integral part of the team. Additionally, I am deeply thankful to the wonderful people I have met through the Max Planck Centre, the FIL, and UCL, many of whom have become valued friends.

I extend my gratitude to my previous supervisors, with special thanks to Dr. Marion Rouault and Dr. Matilde Vaghi, who have been truly inspirational figures in my scientific development. I am also profoundly grateful to Dr. Amir-Homayoun Javadi, who trusted me as an undergraduate student and provided me with the foundational skills that launched my academic career. Going further back, I owe immense thanks to my high school mathematics teacher, Monsieur Sébastien Hamon, who introduced me to coding. Ironically, while I initially dismissed his suggestion to combine coding with psychology, this very fusion has since become the cornerstone of my career. Even more amusing, my first coding project—a predator escape game—seems to have set the stage for what was to come.

My appreciation also goes to Sofinnova Partners, and in particular to Dr. Javier Nunez-Vicandi and Dr. Anta Gkelou, for granting me an incredible internship opportunity. This experience was rich in learning, growth, and discovery, and I will always treasure the time I spent there. I am equally grateful for the friendships I made along the way, which made the experience unforgettable.

I am also deeply thankful to my friends outside academia, who have brought balance, joy, and perspective to my life during these years in London. It was a privilege to never feel alone in such a big city, thanks to your unwavering support and companionship. You made London feel like home, despite being 600 miles away. A special thank-you to Ali for sharing so many adventures with me—you truly made these years extraordinary.

I could not have accomplished this without the unwavering support of my parents, who have always encouraged me to pursue my passions without ever imposing pressure. Your love, faith in me, and steady guidance have been my anchor throughout this journey. Thank you especially for helping me navigate so many moves across countries—I truly could not have done it without you.

Last but certainly not least, my deepest gratitude goes to Johanna, who has been my partner in crime, my best friend, and my greatest source of strength. I have always looked up to you, and together we have tackled every challenge, always supporting each other, even in the most uncertain times. Your love, joy, and matcha lattes have been a constant source of endless comfort and motivation. And finally, thank you to Mirabelle, who is undeniably the best emotional support animal—her presence has made everything just a little brighter (and louder).

Publications during the PhD

The content of Chapter 2 is based on the results published as a research article in Sporrer, J. K., Johann, A., Chumbley, J., Robinson, O.J., & Bach, D. R. (2024). Induced Worry Increases Risk Aversion in Patients with Generalized Anxiety. *Behavioural Brain Research*. doi.org/10.1016/j.bbr.2024.115192.

The content of Chapter 3 is based on the results published as a preprint in Sporrer, J. K., Melinscak, F., & Bach, D. R. (2024). Transdiagnostic psychiatric symptom dimensions predict behavioral cautiousness. *PsyArxiv*. doi.org/10.31234/osf.io/3u47k.

The content of Chapter 4 is based on the results published as a research article in Sporrer*, J. K., Brookes*, J., Hall, S., Zabbah, S., Hernandez, U. D. S., & Bach, D. R. (2023). Functional sophistication in human escape. *Iscience*. doi.org/10.1016/j.isci.2023.108240.

Other publications related to the PhD but not further discussed include:

Bach, D. R., Sporrer, J. K., Abend, R., Beckers, T., Dunsmoor, J. E., Fullana, M. A., Gamer, M., Gee, D. G., Hamm, A., Hartley, C. A., Herringa, R. J., Jovanovic, T., Kalisch, R., Knight, D. C., Lissek, S., Lonsdorf, T. B., Merz, C. J., Milad, M., Morriss, J., Phelps, E. A., Pine, D. S., Olsson, A., van Reekum, C. M., Schiller, D. (2023). Consensus design of a calibration experiment for human fear conditioning. *Neuroscience & Biobehavioral Reviews*, 148, 105146. doi.org/10.1016/j.neubiorev.2023.105146.

Mancinelli*, F., Sporrer*, J. K., Myrov, V., Melinscak, F., Zimmermann, J., Ted, H. L., & Bach, D. (2024). Dimensionality and optimal combination of autonomic threat learning measures in humans. *Behavior Research Methods.* 56, 6119–6129. doi.org/10.3758/s13428-024-02341-3.

Other publication unrelated to the PhD and not further discussed include:

Sefcikova, V., Sporrer, J. K., Juvekar, P., Golby, A., & Samandouras, G. (2022). Converting sounds to meaning with ventral semantic language networks: integration of interdisciplinary data on brain connectivity, direct electrical stimulation and clinical disconnection syndromes. *Brain Structure and Function*, 227(5), 1545-1564. doi.org/10.1007/s00429-021-02438-x.

The symbol * represents co-first authors.

Commitment to Open Science

Throughout my PhD, I remained committed to the principles of open science, ensuring that all the code, data, and materials used in every research project are publicly accessible on the Open Science Framework (OSF) for transparency, reproducibility, and replicability.

For Chapter 2, all R codes utilized in the analysis, along with the curated data, are available on OSF (osf.io/74n6c/).

For Chapter 3, the task presentation code, developed using the open-source JsPsych JavaScript library, is shared on OSF (osf.io/r8bdn/). Additionally, curated CSV data and all Matlab and R codes for pre-processing and analysis can be found on OSF (osf.io/r8bdn/). The pre-registration is available at osf.io/5hmgk/.

For Chapter 4, the compiled versions of the two VR games are available for download on OSF (osf.io/2b3k7/). Moreover, the R codes used for analysis, along with curated data, including questionnaire summary scores, are accessible at osf.io/w4c73/.

Finally, the thesis is available on GitHub at github.com/jusporrer/PhD_thesis.

Table of Contents

1 Introduction

Life is filled with risky decisions, ranging from the trivial to the critical. We constantly encounter situations that require us to weigh options with varying levels of risk. At a social event, you might decide whether to approach a stranger, risking an awkward moment for the chance of a meaningful connection. When crossing a street, you may be tempted to dash across on a red light to save time, weighing the minor convenience against the potential danger. On poker night, you may place a large bet hoping for a red diamond to complete your flush. Even in recreation—skiing down a challenging slope for the thrill or choosing a safer path to avoid injury—we find ourselves constantly balancing risk with reward.

These examples underscore several important points. Risk taking spans all areas of life, from leisure to finance, social dynamics, health, and survival. Typically, higher-risk options offer the opportunity for greater rewards, but also come with more severe potential downsides compared to safer alternatives. In addition, the outcomes of riskier decisions are generally less predictable. In some cases, this predictability can be quantified, as in poker, while in others it can remain unknown.

Although these scenarios provide a glimpse into risk-taking behavior, they fail to capture the full depth and intricacy of real-life choices. In reality, decisions are rarely as straightforward as choosing between two options. Often, there are multiple factors at play, countless alternatives, and various ways to implement each choice. This complexity makes the process of evaluating and navigating risk much more challenging than what these simplified examples may suggest. Additionally, real-life decisions often involve interacting risks, where the outcome of one decision can influence the risks associated with another, further complicating the process. Understanding and probing the nuances of risk-taking behavior is a challenging endeavor.

The universality of risk might imply that all organisms respond to risk in the same way. However, risk-taking behavior is influenced both by the characteristics of the decision-maker and by the context of the decision. Internal factors include individual differences, such as personality traits that play a significant role in shaping how one responds to risky situations. For example, a person with high sensation-seeking tendencies may be more inclined to take greater risks for thrills (cf. Section 1.4 will dwell deeper into the effect of individual differences). External factors include environmental and situational factors, such as proximity to threats, which also shape risk-taking decisions. For instance, individuals may act more cautiously at greater distances from a threat, but as the danger draws nearer, they are pushed to make riskier decisions to avoid harm. These internal and external factors interact, leading to large individual variations in risk behavior. Thus, understanding risk-taking requires examining both the decision-maker's internal predispositions and the external circumstances they face.

1.1 Conceptualizing risk-taking behavior

When viewed from a higher level of abstraction, the term "risk" encompasses the inherent properties of the world. In everyday language, risk generally implies the chance of a negative event occurring. Scientifically, however, its definition remains elusive and varies considerably between disciplines such as psychology, economics, and biology. Some fields define risk in terms of probability, expected values, or utility function (Kahneman and Tversky, 1979; Markowitz, 1952), while others see it as the likelihood of undesirable events or dangers (Haynes, 1895), or even as an embodiment of uncertainty (Aven, 2012; Mishra, 2014; Payzan-LeNestour and Bossaerts, 2011). This divergence extends further, with some viewing risk as an epistemic subjective construct dependent on the observer's knowledge, while others regard it as an objective independent feature of the world (Aven, 2012; Aven and Renn, 2009; Aven et al., 2011).

Over the past two decades, the definition of risk has gradually evolved, moving away from narrowly probability-based frameworks toward more comprehensive and non-probabilistic models. One reason for this shift is the recognition that probability models, particularly frequentist approaches, are inadequate for capturing risk in unique or non-repeatable real-world events, failing to address the full complexity of risk in diverse scenarios (Aven, 2012).

The definition of risk that underpins this thesis is as follows: Risk represents uncertainty in outcomes regardless of whether the outcome involves a potential loss or gain. It is important to note that there are multiple types of uncertainty (Payzan-LeNestour and Bossaerts, 2011). Here, risk is treated as first-order uncertainty, indicating the irreducible uncertainty that persists even after optimal learning, reflecting the inherent stochasticity in the outcome-generating process itself. This is distinct from second-order uncertainty, which captures the likelihood of sudden environmental changes (Payzan-LeNestour and Bossaerts, 2011; Sandhu et al., 2023)—an area beyond the scope of this thesis.

1.2 Risk Preferences

People's responses to those properties of the world are captured by their risk preference that ranges from risk aversion, through risk neutrality, to a risk-seeking attitude (Frey et al., 2017). These risk preferences, in turn, drive risk-taking behavior and are shaped by the characteristics of the situation, the decision maker, and the dynamic and fluctuating interaction between the two. It should be noted that responses to risk also involve concepts such as risk perception (e.g. Slovic et al., 2016) and risk appraisal (e.g. Horvath and Zuckerman, 1993), but this goes beyond the scope of the thesis.

The nature of risk preferences remains contested on two key fronts. The first debate centers on whether risk preferences are stable traits—similar to personality traits in psychology (Anusic and Schimmack, 2016) or enduring tastes in economics (Stigler and Becker, 1977)—or if they are more dynamic, resembling transient states like emotions. This variability may be influenced by various factors; for instance, risk-sensitive foraging theory (Mishra, 2014) posits that metabolic needs shape risk preferences. The second issue concerns the structure of risk preferences: whether there is a single universal factor, comparable to the g factor in intelligence (Deary, 2012) or the p factor in psychopathology (Caspi et al., 2014), or if risk preferences are instead a multidimensional construct that varies between different domains of life.

A comprehensive study involving over 1500 participants and 39 risk-taking measures (Frey et al., 2017) supports the existence of a general r factor that captures both broad, overarching risk tendencies and more domain-specific elements. Crucially, this r factor has shown a strong and consistent link to risky behaviors in the real world and has been shown to be highly stable over time. These findings suggest that risk preferences may indeed function as a stable psychological trait, reflecting a fundamental aspect of human behavior in various contexts (Frey et al., 2017).

Nevertheless, the strength of the conclusions drawn about these two properties hinges on the quality of the measurement tools (Schildberg-Hörisch, 2018). The following Section (1.3) will delve deeper into the psychometric properties of these varied measures.

1.3 Assessing risk-taking behavior

The study of risk-taking behavior in animals is influenced by two distinct schools of thought: comparative psychology and ethology (Datta et al., 2019; Gomez-Marin et al., 2014). The first focuses on understanding how the brain generates behavior in response to rewards and punishments, often employing controlled experiments where animals are trained to respond to sensory stimuli with simple actions (Jazayeri and Afraz, 2017). In contrast, ethology emphasizes the observation of natural behaviors in their ecological settings, proposing that understanding these natural patterns can provide insights into how the brain generates behavior (Datta et al., 2019; Mobbs and Kim, 2015; Tinbergen, 1963). These traditions are not only foundational in animal research, but have also been extended to human studies, providing a basis for examining risk-taking behavior between species (Mobbs et al., 2021).

Building on insights from animal research, recent studies have begun to explore various positions along a continuum between controlled and naturalistic paradigms. This continuum enables a more comprehensive understanding of how risk-taking behaviors are shaped by the environment (Mobbs and Kim, 2015). Indeed, human research has increasingly emphasized the use of ecologically valid paradigms that present "rich, multimodal dynamic stimuli reflecting our daily lived experience" (Sonkusare et al., 2019) and the need to incorporate ecological challenges—problems that our brains have evolved to solve (Scholl and Klein-Flügge, 2018). Alongside these approaches, self-reported measures have also been employed to provide insight into how individuals assess their own risk preferences, adding another layer to understanding risk-taking behavior in both controlled and naturalistic contexts (Hertwig et al., 2019).

While there is a growing push towards more ecologically valid paradigms, this does not diminish the value of controlled studies. Controlled experiments remain essential for revealing critical distinctions that might otherwise be obscured by potential confounding factors or lack of granularity in naturalistic settings. Accordingly, this thesis utilizes a mixed-methods approach that integrates both controlled and naturalistic settings. This allows us to validate and confirm different aspects of behavior across varied contexts, providing a fuller understanding of the mechanisms driving risk-taking.

As such, this section provides an overview of various tasks commonly used to investigate human risk-taking behavior, ranging from controlled experimental setups to more naturalistic settings, as well as self-reported measures. I will also examine their reliability and validity. Notably, the thesis focuses specifically on risk-taking behavior driven by threats or other negative consequences, rather than on broader behavioral effects on attention, such as the prioritization of processing emotional stimuli. Fear conditioning will also not be covered in this overview, as it primarily addresses how defensive responses are generalized to other stimuli rather than focusing on innate responses themselves.

1.3.1 Self-reported Measures

The most straightforward way to assess risk-taking preferences is simply to ask people to report on their tendencies (Charness et al., 2013; Frey et al., 2017; Hertwig et al., 2019). This can be achieved using several types of questions.

Self-Assessment Questions

Self-assessment questions rely on the introspective abilities of the participants, asking them to self-assess their risk preferences, often using a Likert scale. Such questions can range from general inquiries about risk-taking tendencies (e.g. “Are you generally a risk-taking person or do you try to avoid risks?”) to more specific but hypothetical scenarios (e.g., "How likely would you be to go white-water rafting at high water in the spring?"; Blais and Weber, 2006; Dohmen et al., 2011; Josef et al., 2016; Mata et al., 2016; Weber et al., 2002). Some approaches also use multiple-choice formats for specific scenarios to further gauge risk preferences (e.g., "If there has been a natural disaster at your travel destination, would you still go?"), offering risky choices (e.g., "I travel to the destination") and safer but costly options (e.g., "I cancel the vacation"; Hockey et al., 2000; Mitte, 2007).

This section deliberately omits questionnaires on impulsivity (e.g. Patton et al. 1995), sensation seeking (e.g. Hoyle et al. 2002; Zuckerman 2007b), or daringness (e.g. Lahey et al. 2010) as these measures, while related, do not directly assess behavioral risk-taking but instead capture risk-related traits that may influence such behavior. For a detailed discussion of findings related to these personality traits, refer to Section 1.4.2.

Visual Scenario-Based Questions

To increase ecological validity and immersion, some questionnaires incorporate visual aids such as images or videos to simulate real-life scenarios. For example, the Vienna Risk-Taking Test (Arendasy et al., 2005; Hergovich et al., 2007), enhances the self-report experience by presenting participants with sequences of images depicting risky situations and asking them to indicate the point at which the situation becomes unsafe. In one scenario, a series of images show a car preparing to overtake another on a snowy road with oncoming traffic (cf. Figure 1.1.A.). Participants press a key to indicate when they believe the action is no longer safe, and these response times serve as a measure of risk preference.

Behavioral Frequency Questions

The behavioral frequency questions involve asking participants to report the frequency of their actual risky behaviors, such as "How many cigarettes do you smoke per day?" (Marsch, 2021) or how likely they would be to engage in specific behaviors like "Drinking heavily at a social function" or "Gambling a week's income at a casino." Unlike self-assessment questions, this method focuses on specific, observable behaviors, providing insight into real-life decisions that carry actual consequences. The underlying assumption is that these reported behaviors may better reflect true risk preferences because they are grounded in actual experiences rather than hypothetical situations.

Reliability and Validity

Overall, these questions are straightforward to administer, making them particularly useful for large-scale surveys. However, they primarily capture perceived rather than actual risk behavior. Self-reported data are subject to various biases, such as social desirability bias, where individuals may provide more favorable answers, or recall bias, where individuals may not accurately remember their past behaviors. Respondents may also lack the incentive to answer truthfully or possess varying levels of metacognitive ability to introspect accurately about their risk preferences. Nonetheless, self-reported propensity measures have been repeatedly found to have high test-retest reliability (Frey et al., 2017, 2021; Mata et al., 2018). They are also reliable predictors of naturalistic risk-taking behavior, such as investment in stocks, self-employment, and smoking, therefore demonstrating good external validity (Dohmen et al., 2011). In sum, while self-report measures of risk-taking seem to offer a reliable and valid insight into risk preferences, these do not allow access to behavioral risk-taking per se.

1.3.2 Formalized Tasks

This section explores two prominent categories of formalized tasks used to measure behavioral risk-taking (Frey et al., 2017; Hertwig et al., 2019; Pedroni et al., 2017; Schonberg et al., 2011). The first category is rooted in economics and focuses on lotteries or gambling tasks, with Chapter 2 employing one such economic task. The second type focuses on ethologically inspired risky foraging tasks to simulate the natural settings in which risk-taking behavior evolved, as utilized in the task featured in Chapter 3. These tasks are often incentivized, which can be both positive (e.g., monetary gain) and/or negative (e.g., monetary loss or threat of shock).

Economic Choices

The simplest way to assess overt risk-taking behavior involves two-outcome lottery choices. In each trial, participants choose between a guaranteed amount and a gamble with two equally likely outcomes, one of which always exceeds the guaranteed amount (cf. Figure 1.1.B.). There are three possible trials: the gamble offers either a positive gain or nothing (e.g. £1 certain versus £2/£0 gamble), a gain or a loss (e.g. £0 certain versus £3/-£3 gamble), or a loss or nothing (e.g. -£2 certain versus -£4/£0 gamble). These choices are often analyzed using a Prospect Theory model (cf. Section 1.5.3 for more details), though other models also exist (Markowitz, 1952). This task has been conducted on paper (Kahneman and Tversky, 1979), via computer (Baek et al., 2017; Charpentier et al., 2017, 2016a,b; Chumbley et al., 2014; Ernst et al., 2014; Klaus et al., 2020; Leahy et al., 2012; Schulreich et al., 2016; Smoski et al., 2008; Sokol-Hessner et al., 2013, 2009, 2016; Sokol-Hessner and Rutledge, 2019a; Tom et al., 2007; Walasek et al., 2018; Xu et al., 2020), in fMRI (Canessa et al., 2017, 2013; Martino et al., 2010), and on smartphones (Bedder et al., 2023; Rutledge et al., 2016). A version of this task is utilized in Chapter 2. Similar but less common iterations also include the price list method (Holt and Laury, 2002), a risky asset investment task (Gneezy and Potters, 1997), and a two-choice gambling task (Eckel and Grossman, 2002).

In real-life decision-making, it is rare to encounter situations where probabilities are objectively known, as they are in simple monetary gambles. To better capture this complexity, more sophisticated economic tasks have been developed, in which individuals need to learn the statistics of the task through experience (Hertwig et al., 2019). One of the most well-known examples is the Iowa Gambling Task (IGT; Bechara et al., 1994). In the IGT, participants choose cards from one of four decks in each trial. Two decks offer high rewards but come with larger losses, making them less profitable in the long run, while the other two decks provide smaller but more consistent rewards and better long-term outcomes. Through trial and error, participants gradually learn which decks are advantageous. Some versions of the IGT add further complexity by altering probabilities throughout the task, requiring continuous adaptation to changing contingencies. A similar task is the Columbia Card Task (CCT; Figner et al., 2009), where participants flip over 32 cards from a deck to accumulate points, but with each flip, they risk encountering a loss card that can end the round and cost them all accumulated points.

Risky Foraging Tasks

To better replicate the ecological conditions under which risk-taking behaviors evolved, researchers have developed ethologically inspired risky foraging tasks. The general objective is to collect rewards while avoiding a predator which would make you lose your rewards. This often involves an Approach-Avoidance Conflict (AAC), where an individual must integrate a variety of information to navigate a conflict between collecting a reward but risking harmful consequences or sacrificing rewards to avoid potential negative outcomes. Various iterations of these tasks exist, ranging from relatively simple designs to more intricate, complex scenarios. An example is the "Active Escape Paradigm" (AEP), where participants wait as a predator approaches them on a runway grid, accumulating money the longer they remain (cf. Figure 1.1.D.). At any point, they can choose to escape, but if the predator accelerates and catches them, they lose their accumulated reward and receive an electric shock (Fung et al., 2019; Qi et al., 2018).

Another task is the "AAC scoop and run" in which participants start in a safe place in a quadrant and can choose whether and when to approach a reward at the risk of getting caught by a predator with varying levels of threat probability and magnitude (cf. Figure 1.1.E.). The probability of threat is indicated by different grid colors and is learned through direct experience, while the potential loss is explicitly indicated as a possible monetary amount (Abivardi et al., 2020; Bach, 2015, 2017; Bach et al., 2019; Castegnetti et al., 2020; Khemka et al., 2017). A version of this task is utilized in Chapter 3. Unlike in the AEP, the threat probability is constant throughout the trial and is not contingent on the reward.

A similar task is the “AAC stay and play” where participants navigate a 24 x 16 open-field grid, aiming to collect as many dispersed monetary tokens as possible (cf. Figure 1.1.F.). They start in a safe corner, while a predator begins in another corner and chases them across the grid. If caught, participants lose all the tokens they have collected (Bach et al., 2014, 2020).

There are also variations of these foraging tasks that omit the explicit AAC element. For instance, some tasks focus on avoiding starvation without the presence of a predator (Korn and Bach, 2018), while others involve evading threats without the prospect of any reward (Mobbs et al., 2007).

Additionally, there are numerous AAC task variants with different cover stories beyond the context of risky foraging (Aupperle et al., 2011; Loh et al., 2017; O’Neil et al., 2015; Smith et al., 2021). One of the most commonly used is the “Balloon Analogue Risk Task” (BART; Lejuez et al., 2002). In the BART, participants inflate a virtual balloon to earn money, with each pump increasing both the potential reward and the risk of the balloon bursting (cf. Figure 1.1.C.). They must decide when to stop pumping and cash out to avoid losing their earnings if the balloon bursts. Although the BART does not involve a predator, it shares similarities with the AEP since the decision to "cash out" mirrors the decision to escape.

Reliability and Validity

Only a few formalized tasks report reliability, and it ranges from weak to excellent, depending on the task. Low reliability can be attributed to various factors, such as measurement error, response bias, temporal instability, or low internal consistency. In economic choices, the parameters of the Prospect Theory model have 2-week good-to-excellent test-retest reliability (Mkrtchian et al., 2023). Conversely, the reliability of the IGT has been consistently challenged, with evidence suggesting that it is weak (Buelow and Barnhart, 2018; Buelow and Suhr, 2009; Lin et al., 2013). For AAC tasks, the BART shows acceptable reliability over a 2-week period (Buelow and Barnhart, 2018; Weafer et al., 2013; White et al., 2008), while the "AAC stay and play" task demonstrates good 2-year test-retest reliability (Bach et al., 2020).

The evidence regarding the internal validity (or construct validity) of risk-taking tasks is mixed. For example, the IGT and BART did not share an underlying structure and appear to measure different aspects of risk-taking behavior (Buelow and Blaine, 2015). Another study did not find consistency in several economic tasks (Charness et al., 2020). However, in a large adolescent cohort, the propensity to choose a gamble over a certain reward parametrized by an economic risk preference model (Markowitz, 1952) was related, though weakly, to behavioral cautiousness in the "AAC scoop and run task" (Bach et al., 2020).

In terms of external validity (or predictive validity), the risk preferences elicited using the BART in adolescents were linked to real-world risky behavior such as gambling, drug use, and unprotected sex (Bornovalova et al., 2005; Charness et al., 2013; Lejuez et al., 2003, 2002). Lottery-based measurements were related to actual market behavior (Pennings and Smidts, 2000), although not consistently (Charness et al., 2020; Schonberg et al., 2011).

1.3.3 Virtual Reality Tasks

Immersive virtual reality (VR), coupled with emerging technical and methodological advances in motion tracking, is part of a new wave of experimental and behavioral research that represents a promising tool to study and reproduce naturalistic human actions. This is particularly useful when recreating risky scenarios as it gives participants a high feeling of immersion and VR allows for minimal risk but maximal experimental control and provides precise kinematic measurements of the defensive movements.

This technology has already shown its promise in threat research by permitting the ecologically valid translation of established simple AAC tests, including the open field test (Gromer et al., 2021; Kallai et al., 2007) and the elevated plus-maze (Biedermann et al., 2017; Rodrigues et al., 2020; Yilmaz Balban et al., 2021). Indeed, identical to what has been shown for rodents in the elevated plus-maze, humans spent more time on the safe fractions of the maze (i.e., center and closed arms) and avoided spending time on the aversive open arms fraction in a mixed-reality version of the elevated plus-maze (cf. Figure 1.1.G., Biedermann et al., 2017).

Recently, more complex and ethnologically inspired risky foraging tasks have emerged in VR. One significant advantage of using VR is that it provides more natural stimulus material, enhancing the ecological validity of experimental tasks. In one such study, similar to the formalized task by Korn and Bach (Korn and Bach, 2018), participants had to collect enough apples to survive 24 days (one trial) by moving around an environment (using a joystick) composed of three different naturalistic contexts—water, forest, and desert—each with varying levels of reward and punishment magnitude (cf. Figure 1.1.H., Kastrinogiannis and Lonsdorf, 2023).

An underutilized potential of VR lies in its ability to offer a more naturalistic action space, allowing participants to engage in continuous and dynamic decision-making processes that challenge simple decision algorithms—this goes beyond mere ecological validity. For instance, continuous decisions made in a VR environment cannot always be explained by categorical decision models. Chapter 4 introduces a novel risky foraging task where participants have to collect as many fruits as possible while avoiding bio-realistic threats that could chase them (cf. Figure 1.1.I.). They can run to a safe house or try to evade these threats in a large 5 x 10-meter room, engaging in continuous, naturalistic actions beyond simple categorical decisions.

Reliability and Validity

Due to its relatively recent adoption, the reliability and validity of VR tasks remain underexplored. The complexity of the resulting behavioral data further complicates this investigation. Nevertheless, the fact that VR tasks can replicate previously validated AAC tests indicates their potential and promise.

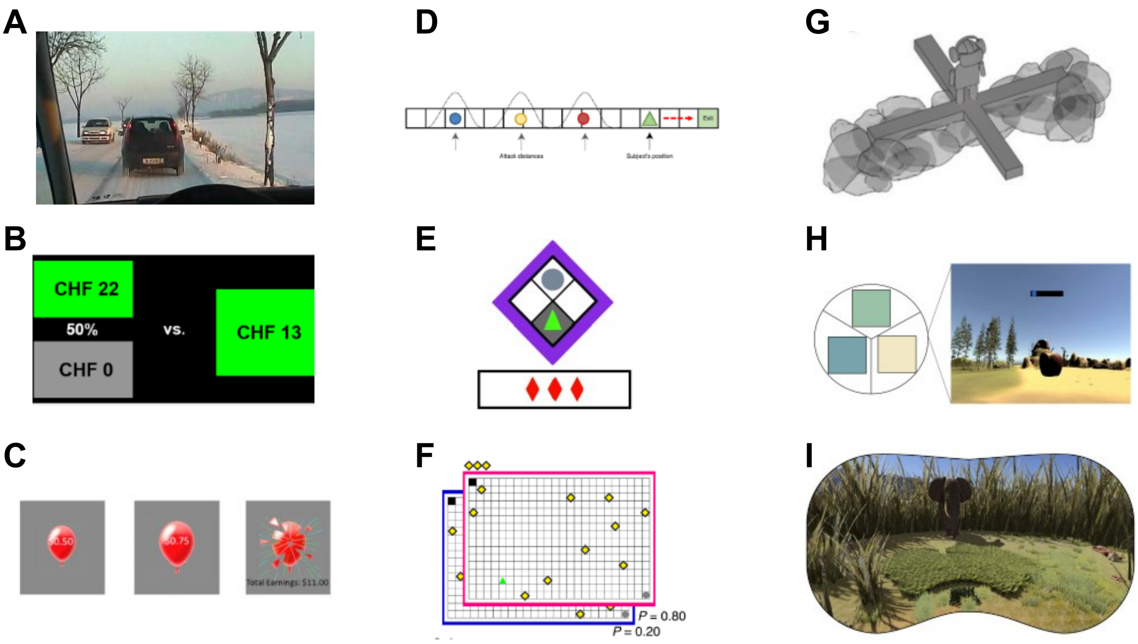

Figure 1.1: **Diverse experimental setup used to assess risk-taking behavior.** Visual representation of **(A)** the Vienna Risk-Taking Test (VRTT, Arendasy et al., 2005), **(B)** a two-outcome lottery choice used in Chapter 2, **(C)** the Balloon Analogue Risk Task (BART, Lejuez et al., 2002), **(D)** the Active Escape Paradigm (AEP, Qi et al., 2018), **(E)** the AAC scoop and run (Bach, 2015) used in Chapter 3, **(F)** the AAC stay and play (Bach et al., 2014), **(G)** the mixed reality elevated plus-maze (Biedermann et al., 2017), **(H)** a VR AAC risky foraging task (Kastrinogiannis and Lonsdorf, 2023), and **(I)** the VR risky foraging task used in Chapter 4.

1.4 Individual Differences in Risk-taking Behavior

While there is a lack of consensus on what constitutes risk, how to measure it, and how risk-taking behavior should be defined, it is widely accepted that there are large individual differences in risk-taking behavior. It is influenced by both the characteristics of the decision-maker and the context of the decision. Factors such as sex, impulsivity, and other personality traits, as well as the specific context or domain of the decision, play significant roles in shaping risk-taking behavior.

1.4.1 Anxiety and Fear

Maladaptive avoidance of risky situations is widely believed to be a core mechanism sustaining anxiety, and reducing it is the primary aim of most psychotherapies, such as Cognitive-Behavioral Therapy and Exposure Therapy (Aupperle and Paulus, 2010). Furthermore, the behavioral readout from AAC tests (cf. Section 1.3.2) often reflects the impact of anxiolytic drugs, making these tests central to pre-clinical anxiety research. Anxiety is thought to be a key driver behind reduced risk-taking behavior, as supported by studies showing that individuals with higher levels of anxiety, or those diagnosed with anxiety disorders, tend to exhibit greater risk avoidance (Giorgetta et al., 2012; Lejuez et al., 2002; Maner et al., 2007). For example, people with high anxiety spend less time in the open arms of a virtual reality version of the Elevated Plus Maze (Biedermann et al., 2017) and cashed-out earlier in the BART (Lejuez et al., 2002; Maner et al., 2007).

However, this association has not been consistent across all studies. Many large-scale investigations did not find significant effects, as demonstrated in the "AAC stay and play" task (Bach et al., 2020) or other risk-taking paradigms (Kallai et al., 2007; Struijs et al., 2018, 2017). Notably, a recent systematic review on interindividual differences in task-based approach-avoidance behavior found no consistent evidence of an effect of self-reported anxiety (Fricke and Vogel, 2020).

Some studies, however, suggest more specific differences rather than broad effects. Various factors, such as the context of the risk and individual-specific constructs, have been explored to explain these mixed findings. For instance, the domain and framing of risk seem to alter risk-taking behavior in anxious individuals. High-anxiety individuals are more likely to engage in health-related risks in contrast to other domain-specific risks (Nicholson et al., 2005), especially when framed negatively (Lauriola et al., 2005). On the other hand, the level of ambiguity involved in the likelihood of outcomes also seems to have an effect. Higher anxiety predicts less risk-taking in a high-ambiguity version of the BART (cf. Section 1.3.2), whereas anxiety is unrelated to risk-taking in low-ambiguity conditions (Smith et al., 2016). The effect of uncertainty is further supported by studies using economic tasks, where individuals with clinical anxiety show an increased risk aversion when faced with uncertain outcomes, but not necessarily a higher sensitivity to loss (Charpentier et al., 2017). This is consistent with evidence from other contexts, such as self-reports, where greater intolerance of uncertainty is reliably associated with increased anxiety (Dugas et al., 1997, 1998; Dugas and Ladouceur, 2000; Grupe and Nitschke, 2013; Mahoney and McEvoy, 2012; Mcevoy and Mahoney, 2011; Sandhu et al., 2023; Yook et al., 2010).

Additional factors may account for the inconsistencies found across studies. These include limitations such as small sample sizes, significant variability in the stimuli, tasks, or populations examined. Moreover, many studies did not assess whether the observed effects of anxiety are distinct from those of other clinically relevant traits, potentially capturing nonspecific effects that may not generalize across different tasks (this issue will be explored further in Chapter 3). Another possibility is that anxiety's impact on risk-taking behavior may be more state-dependent, influenced by temporary emotional conditions, rather than representing a stable, trait-like characteristic (this hypothesis will be explored further in Chapter 2). On the other hand, it is possible that (some) approach-avoidance tasks might not be as sensitive to individual differences as initially thought. Indeed, recent reports have suggested that using these tests to explore anxiety etiology in healthy populations needs to be queried and may turn out to be unrealistic (Bach, 2021).

Finally, it is also essential to consider that some risk-taking tasks widely used to assess anxiety may actually be measuring fear—a closely related but potentially distinct construct. Fear is typically triggered by an immediate and specific threat, whereas anxiety tends to arise from uncertain future threats (Mobbs et al., 2020). Using the AEP (cf. Section 1.3.2), individuals with higher trait anxiety escaped earlier from slow predators but not from fast predators (Fung et al., 2019). This can be interpreted in line with the threat imminence continuum model (Fanselow and Lester, 1988; Mobbs et al., 2020), where slow threats evoke anxiety-like behaviors due to their uncertain and distant nature, while fast threats trigger fear-driven, reflexive responses. The earlier escape in response to slow predators may reflect the heightened anticipatory avoidance typical of anxiety, whereas the absence of this behavior with fast predators aligns with the more immediate, fear-based reaction. Similarly, defensive behaviors in rodents vary depending on whether the threat is close or far away (Davis et al., 2010), with anxiolytics shown to influence reactions only to distant threats (Blanchard et al., 1993). However, the distinction between fear and anxiety can be ambiguous, as both constructs exist along a continuum and are supported by overlapping neural systems responsible for defensive behavior (Shackman and Fox, 2016).

1.4.2 Impulsivity, Sensation-seeking, and Daringness

Impulsivity and sensation-seeking are personality traits commonly believed to influence various forms of risk-taking (Mishra, 2014; Mishra and Lalumière, 2011; for a review, see Zuckerman, 2007a). Impulsivity refers to the tendency to act spontaneously without prior deliberation. Its exact nature is debated, with some considering it a heterogeneous cluster of lower-order traits (Cross et al., 2011; MacKillop et al., 2016). Sensation-seeking, on the other hand, involves a desire for diverse, novel, and intense experiences and sensations (e.g., physiological arousal), often accompanied by a willingness to take physical, social, and financial risks to attain such stimulation (Zuckerman, 1979). While some view sensation-seeking as a subcategory of impulsivity (Berg et al., 2015), others regard it as a distinct personality type (Cross et al., 2011; MacKillop et al., 2016).

Individuals high in impulsivity and sensation-seeking often engage in high-risk activities such as sexual risk-taking and reckless driving (Schwebel et al., 2006), smoking, alcohol, and drug use (Berg et al., 2015; Hoyle et al., 2002; Martin et al., 2002; Nicholson et al., 2005; Reynolds et al., 2019; Roberti et al., 2003; Stephenson et al., 2003; Zuckerman, 2007b). Furthermore, BART performance is associated with these traits (Charness et al., 2013; Lauriola et al., 2014; Lejuez et al., 2003, 2002), although not consistently (Mishra and Lalumière, 2011). Sensation-seeking is also associated with less avoidance on the open arms of an Elevated Plus Maze in VR (Biedermann et al., 2017).

Both traits also diverge in specific ways. Sensation-seeking is often associated with experimentation with drugs during adolescence but typically does not lead to maladaptive outcomes like substance abuse (Khurana et al., 2015, 2017, 2018). In contrast, impulsivity predicts the onset, progression, and relapse of drug abuse (Argyriou et al., 2018; Loree et al., 2015). Furthermore, while both traits are higher in men, sex differences are more pronounced in sensation-seeking (Cross et al., 2011, 2013; MacKillop et al., 2016).

In addition, daringness, a personality type closely related to sensation-seeking, has emerged as the best self-reported predictor in the "AAC stay and play" risky foraging task among young individuals (Bach et al., 2020). Whether daringness is truly distinct from sensation-seeking or simply a matter of terminology remains to be clarified. The literature on daringness is limited, but a comparison of the CADS daringness scale items (cf. Appendix Table B.4) reveals strong similarities to those of the sensation-seeking scale (cf. Appendix Table B.5).

1.4.3 Sex Differences

Males exhibiting greater risk-taking behavior is a characteristic observed across numerous animal species (Brand et al., 2023; Daly and Wilson, 1983; Habig et al., 2017; King et al., 2008; Orsini et al., 2016; Videlier et al., 2015). Evolutionary theories attribute this pattern to lower parental investment and greater variance in reproductive success among males. These factors drive males to engage in high-risk, high-reward behaviors to maximize their mating opportunities. Therefore, the differential costs and benefits associated with reproduction suggest that males are biologically predisposed to take more risks than females.

Extensive self-report studies in humans corroborate this hypothesis, revealing consistent sex differences in risky behaviors (Byrnes et al., 1999; Dohmen et al., 2011; Frey et al., 2021; Josef et al., 2016; Nicholson et al., 2005). For instance, men are more prone than women to engage in activities with a possibility of physical harm (Byrnes et al., 1999), take financial risks (Charness and Gneezy, 2012), and more risk-taking in mundane everyday situations, like catching a bus (Pawlowski et al., 2008). These sex differences are evident from adolescence, as national surveys demonstrate higher risk-taking behaviors among male teenagers (Eaton et al., 2008). Interestingly, the magnitude of these differences varies by domain; larger sex differences are observed in physical risk-taking compared to domains like smoking or social risks (Byrnes et al., 1999; Frey et al., 2021; Josef et al., 2016; Nicholson et al., 2005).

When examining behavioral risk-taking tasks, the findings are more nuanced and sometimes inconsistent. Some studies report overall sex differences in both simple (Dohmen et al., 2011) and complex gambling tasks (Lewis et al., 2022; Overman, 2004; Reavis and Overman, 2001; Weller et al., 2010). Conversely, other studies find no significant differences in tasks like the BART (Lauriola et al., 2014; Lighthall et al., 2009, 2012) and certain gambling paradigms (Frey et al., 2021; Josef et al., 2016). Some other studies find more specific differences rather than overall differences. In the "AAC stay and play" risky foraging task, males were less cautious when their potential loss was smaller but adapted their behavior to the same level as females when the potential loss was higher (Bach et al., 2020). Similarly, several studies found that women exhibit lower risk adjustment (Deakin et al., 2004; van den Bos et al., 2013, 2014). Therefore, it seems that men do not necessarily behave more recklessly in general, but adjust to risk more in low-risk situations (Lewis et al., 2022).

1.4.4 Interaction between Individual and Situational Differences

The interaction between individual differences and situational factors is a crucial aspect of understanding risk-taking behavior. Individuals often vary in their propensity for risk depending on the specific context or domain in which the decision is made. For example, personal skills and experiences in particular areas can significantly shape how one approaches risk. This might be especially evident when the individual possesses specialized skills relevant to the decision-making domain. For instance, the risk attitudes of company managers have been found to differ substantially depending on whether the risk is related to recreational activities or financial decisions (Maccrimmon and Wehrung, 1990). Similarly, Dreber et al. (2011) demonstrated that female tournament bridge players exhibited markedly different risk-taking propensities in the domain of bridge compared to financial decision-making. These examples highlight how expertise or familiarity with a particular domain can influence one's willingness to take risks.

In addition to personal skills and domain-specific expertise, cultural norms play a significant role in shaping risk perceptions and behaviors. Risk-taking is often evaluated differently across cultures, with some societies valorizing risk as a sign of bravery, leadership, or innovation, while others view it as reckless or irresponsible. This cultural lens can lead to substantial variability in risk behaviors between individuals from different cultural backgrounds. Research has documented between-country differences in risk-taking, often measured by participants' willingness to engage in monetary gambles or other forms of risk-related decision-making. For example, L'Haridon and Vieider (2019) and Rieger et al. (2015) found that individuals from countries with lower income per capita exhibited higher tolerance to risk compared to individuals from wealthier nations. This finding suggests that economic factors, such as poverty, may contribute to increased risk tolerance, potentially as a necessity or an adaptive response to uncertain living conditions. Although this thesis does not delve deeply into the complexities of socio-cultural influences on risk-taking, it is essential to acknowledge that individual differences in risk preferences may be shaped by broader societal structures, such as economic conditions, political stability, and social norms.

1.5 Frameworks for Analyzing Risk-Taking Behavior

Individual differences in the extremes of risk-taking play a significant role in the development and maintenance of psychiatric symptoms. Excessive risk aversion can result in debilitating avoidance behaviors, where individuals withdraw from opportunities or challenges due to fear of failure or harm, reinforcing anxiety, depression, and diminished quality of life. In contrast, excessive risk-seeking behaviors can lead to severe consequences, particularly among young adults. Risky decisions, such as substance abuse or neglecting safety measures like wearing seatbelts, contribute to unintentional injuries—the leading cause of death in this demographic (Eaton et al., 2008). These patterns underscore the role of risk-taking on mental health and highlight the importance of better understanding their impact. The following sections will explore frameworks that provide the tools to analyze, model, and quantify how individual differences, including psychiatric symptoms, influence risk-taking behaviors.

1.5.1 From Biological and Cognitive Psychiatry to Computational Psychiatry

Historically, various dominant schools of thought have emerged to better understand behavior. Biological psychiatry focuses on identifying the biological underpinnings of cognitive traits and psychiatric symptoms. It emphasizes the brain's role as the key organ behind mental processes, often viewing mental disorders as arising from dysfunctions in neurochemical or structural aspects of the brain. This perspective has driven significant advancements, particularly in psychotropic drug development, by targeting neurotransmitter systems, such as dopamine or serotonin, to treat conditions like depression, anxiety, and schizophrenia. However, this approach has limitations, as it often involves making broad conceptual leaps from molecular processes, such as receptor activation or synaptic plasticity, to complex psychological phenomena and behaviors (Montague et al., 2012).

In contrast, cognitive psychiatry emerged with a different emphasis, viewing psychiatric symptoms as the outcome of maladaptive cognitive processes (Shallice, 1988). Cognitive psychiatry draws on cognitive psychology and theories of learning and perception. It posits that mental disorders are often the result of distorted or dysfunctional cognitive operations—such as flawed information processing, erroneous belief systems, or maladaptive learning patterns. Early foundational work, such as the Little Albert experiment, demonstrated how emotional responses (such as fear) could be conditioned through learning processes, shaping behavior in ways that could lead to long-term psychological difficulties (Harris, 1979). This early cognitive behavior framework has been instrumental in understanding conditions such as depression and anxiety, where maladaptive thinking patterns (e.g., catastrophizing, obsessive thoughts) can be directly targeted and restructured through therapeutic interventions such as cognitive-behavioral therapy (CBT). This approach also has limitations in its lack of integration of biological functions, but these were partially addressed by the eventual development of cognitive neuropsychiatry (Halligan and David, 2001).

The rise of computational approaches in cognitive neuroscience allowed researchers to fill the gap left by a lack of intermediate descriptive frameworks (for reviews see Huys et al., 2011; Montague et al., 2012). This paradigm proposes that the brain functions as an information-processing system, continuously updating its internal model of the world with new data (Huys et al., 2021a). As such, computational models facilitate a mechanistic understanding that can operate across multiple levels of analysis (Maia and Frank, 2011; Wang and Krystal, 2014). This approach underpins the emerging field of computational psychiatry.

Computational psychiatry is more expansive than its name suggests. While the term "psychiatry" suggests a focus solely on mental disorders, the scope of the field extends across the entire spectrum of cognitive phenotype as the distinction between "normal" versus "aberrant" is often ambiguous. For example, avoidance is adaptive in response to real danger but is considered pathological when it occurs without a threat. Moreover, while this thesis primarily concentrates on modeling behavior and individual differences, it represents only a small portion of the broader field of computational psychiatry. It does not delve into other significant advancements, such as applying pattern recognition in neuroimaging (Wolfers et al., 2015) or digital phenotyping (Insel, 2017; Onnela and Rauch, 2016).

In summary, computational psychiatry can be characterized by its efforts to apply quantitative methods and computational modeling techniques to develop new insights and approaches for understanding biological or cognitive functions that psychiatric disorders may impact—integrating methods from multiple disciplines such as psychology, biology, neuroscience, mathematics, and artificial intelligence (Adams et al., 2016; Friston et al., 2014; Huys et al., 2021a, 2016, 2011; Montague et al., 2012; Moutoussis et al., 2018).

1.5.2 The Two Branches of Computational Psychiatry

Computational psychiatry has two main branches: a theory-driven approach and a data-driven approach (Huys et al., 2016). Theory-driven computational approaches employ formal models of psychiatric conditions to make explicit hypotheses, sometimes at multiple levels of analysis (Huys et al., 2021b; Maia and Frank, 2011; Wang and Krystal, 2014). There are many formal schemes ranging from neural network theory, reinforcement learning, or game theory to detailed biological models (Stephan et al., 2009). On the other hand, data-driven machine-learning approaches focus on developing agnostic models by leveraging big, high-dimensional data (Marsch, 2021; Rutledge et al., 2019). Researchers have applied this approach to high-dimensional neuroimaging data (DeBattista et al., 2011; Iosifescu et al., 2016; Wolfers et al., 2015) and large-scale clinical cohorts (Chekroud et al., 2016; Koutsouleris et al., 2016) to recover diagnostic information, predict treatment outcomes, or improve treatment selection. While the agnostic nature of data-driven approaches allows for unbiased flexibility, it can also present a downside by disregarding existing scientific theories and findings. Therefore, integrating data-driven methods with theory-based approaches can help mitigate this issue (Hauser et al., 2022; Huys et al., 2016). For example, the parameters from mechanistic models can be used as features to improve machine learning performance. This strategy has proven effective in reinforcement learning (Eshel et al., 2015; Steinberg et al., 2013).

This thesis employs both approaches across two chapters to explore clinically relevant individual differences in risk-taking behavior. Chapter 2 utilizes Prospect Theory as a framework for analyzing risk preferences, relying on computational measures rather than raw behavioral readouts to capture the effects of clinical differences. Chapter 3 adopts a dimensional, data-driven approach to identify psychiatric phenotypes that are associated with altered risk-taking behaviors.

1.5.3 Example of Theory-driven Approach: Prospect Theory

Descriptive behavioral readouts, such as summary statistics, often serve as distant proxies for underlying cognitive processes and may be more indicative of the experimental design than the processes themselves (Haines et al., 2020; Karvelis et al., 2023). In contrast, generative models provide a more robust approach by explicitly modeling cognitive processes, yielding computational measures that are more closely aligned with the theoretical constructs of interest (Huys et al., 2016). These measures can offer enhanced content validity, convergent validity, and reliability, though this is not true for all generative models (Karvelis et al., 2023). For example, parameters of the Prospect Theory model have good-to-excellent reliability, while those from a four-armed bandit reinforcement learning model ranged from poor to good (Mkrtchian et al., 2023).

Prospect Theory offers a valuable framework for understanding how individuals perceive and evaluate risk and reward (Ruggeri et al., 2020; Schonberg et al., 2011; Sokol-Hessner and Rutledge, 2019b; Tversky and Kahneman, 1979, 1992). Interestingly, this model, originally developed in economics, has been adapted as a cognitive-computational framework to explain clinical variability. It translates objective values into subjective ones according to the decision-maker's preferences (cf.

1.3.2 for examples of those choices). This behavioral economic theory posits that two stable individual factors (i.e., parameters in the model) explain heterogeneity in economic decisions such as the commonly observed inclination towards sure outcomes over risky gambles with potential higher returns (Tversky and Kahneman, 1992). First, risk aversion refers to the tendency to avoid uncertainty in outcomes regardless of whether the outcome is a potential loss or gain. Second, loss aversion refers to the tendency to overweigh potential losses relative to potential gains regardless of how uncertain they are. The concept of loss aversion is rooted in the idea that losing something carries a greater emotional impact than the equivalent gain. While related, these factors provide insight into two distinct processes relying on separate neural correlates. For example, amygdala damage in patients was linked to a dramatic reduction in loss aversion but left risk aversion relatively intact (Martino et al., 2010). It is worth noting that variations in risk preferences are not normatively good or bad per se but are contextually dependent. Loss aversion may prevent death in a survival context but could reduce returns in an investment context (Sokol-Hessner and Rutledge, 2019b).

Formally, the generative model is composed of two main components: the first set of equations translates objective utility into subjective utility (cf. Equation 1.1), while the second equation compares these subjective utilities and maps them into a probability space (cf. Equation 1.2).

$$v(x; \rho_i, \lambda_i) = \begin{cases} -\lambda_i(-x)^{\rho_i} & \text{for } x < 0 \\ x^{\rho_i} & \text{for } x \geq 0 \end{cases} \tag{1.1}$$

$$\text{Prob}[o_1] = \frac{1}{1 + e^{-\mu_i(V_{o_1} - V_{o_2})}} \tag{1.2}$$

This model allows for the estimation of three risk preference parameters, namely λ (loss aversion), ρ (risk aversion), and μ (choice consistency). The underlying assumption of the widely-used S-shaped power utility function is that individuals assess the value of payments with respect to a reference point, which is presumed to be zero. It suggests that based on typical parameter values, individuals tend to be risk-averse when dealing with positive payments (gains) and risk-seeking when facing negative payments (losses). The presence of an inflection point in the utility function at zero indicates loss aversion. This is expressed in Equation 1.1 where λ_i represents loss aversion, ρ_i risk aversion, and x is a monetary outcome relative to the reference point.

λ greater than 1 corresponds to loss aversion, as opposed to less than 1 which corresponds to loss tolerance. λ is usually around 2 as people tend to weigh losses about twice as much as gains (Brown et al., 2021; Chapman et al., 2018; Tversky and Kahneman, 1992). ρ represents the curvature of the utility function, capturing how outcomes are evaluated as amounts increase. Specifically, it determines sensitivity to diminishing returns: when $0 < \rho < 1$, the value function is concave for gains and convex for losses, indicating that people become less sensitive to changes in wealth as amounts grow. This curvature in turn gives rise to risk-related behavior: values of $\rho < 1$ are associated with risk aversion, whereas $\rho > 1$ corresponds to risk-seeking. Thus, while ρ does not directly measure risk aversion, it is conventionally labeled a "risk aversion parameter" in the literature. In keeping with this convention, and to ensure comparability with the studies in Chapter 2, we will continue to describe ρ in those terms. Then utility and choices are mapped following the SoftMax rule using the logit function. The logit function depends both on the utility difference between options o_1 and o_2 and the choice consistency parameter μ_i (cf. Equation 1.2). Higher value of μ represents greater consistency in choices, meaning choices that align more closely with the predictions of Prospect Theory of selecting the value-maximizing option.

Concretely, if you consider two options with magnitudes of a certain choice and a gamble, [+7, +14], the objective difference between them is +7. If risk sensitivity is set at 0.9, the subjective utilities for these values are [+5.6, +11.1], resulting in a difference of +5.5. On the other hand, if risk sensitivity is set at 1.1, the subjective utilities increase to [+9.2, +18.6], with a difference of +9.4. This demonstrates that as risk sensitivity increases, the difference between subjective utilities also increases. Although this does not alter which option is more attractive by default, it becomes relevant in experimental tasks where higher magnitudes are associated with lower probabilities. If the higher magnitude [+14] has a 60% chance of being chosen, while the lower magnitude [+7] is guaranteed, for a risk sensitivity of 1.1, the utilities are [+9.2, +8.5], with a small difference of -0.7. However, with a risk sensitivity of 0.9, the utilities are [+5.6, +4.3], and the difference becomes -1.3. A risk sensitivity of 0.9 reflects risk aversion, where the certain option becomes more attractive, whereas a risk sensitivity of 1.1 appears to encourage risk-seeking behavior.

In sum, the generative model of Prospect Theory allows the extraction of three risk preference parameters—λ, ρ, and μ. These parameters provide a computational alternative to the direct behavioral readouts of economic choices, such as the inclination to choose a gamble over a certain reward. Unlike these basic measures, which do not factor in the different values of gambles, these computational measures provide a solution by offering a more reliable and theoretically grounded approach (Mkrtchian et al., 2023).

1.5.4 Example of Data-driven Approach: Transdiagnostic Psychiatric Symptom Dimensions

The current process of diagnosing mental disorders is largely based on quantifying the types and severity of symptoms a person is experiencing, as well as the degree of associated distress or impairment. This method has the advantage of providing a common language for mental health professionals, as embodied in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) and the International Classification of Diseases (ICD-10; World Health Organization, 1992). These diagnostic systems were developed to offer a simple, reliable, and descriptive taxonomy of discrete psychiatric conditions at the expense of considering the complex underlying etiology and pathophysiology. There are many drawbacks to this traditional phenotyping (Hyman, 2007; Lilienfeld and Treadway, 2016; Stephan et al., 2016). Firstly, there is a high degree of heterogeneity within disorder categories and substantial symptom overlap between different conditions (i.e., comorbidity). Second, the DSM-ICD categories do not easily map onto findings from genetics, neuroimaging, or other modern cognitive neuroscience tools, making it challenging to integrate these findings into diagnosis. Additionally, these diagnostic frameworks also imply a rigid classification system in which an individual must either conform to a set category or be excluded from it, resulting in a binary label of "ill" versus healthy. This dichotomy fails to capture the reality that many symptoms are more accurately represented as continuous traits that span both 'healthy' and clinical populations to varying degrees (Cuthbert and Insel, 2013; Lilienfeld and Treadway, 2016). The National Institute of Mental Health (NIMH) has acknowledged these challenges and launched the Research Domain Criteria (RDoC) initiative to address them (Insel et al., 2010). This initiative focuses on bridging the gaps between various levels of analysis—genes, cells, molecules, behavior, and self-reports—to identify transdiagnostic markers that are biologically grounded and applicable across psychiatric disorders (Cuthbert, 2022; Cuthbert and Kozak, 2013; Insel et al., 2010).

The RDoC initiative has encouraged the development of new methods for studying psychiatric symptomatology (de Wit et al., 2014; Kendler et al., 2011; Schumann et al., 2010). This thesis focuses on one such emerging approach: dimensional analysis among healthy populations, as discussed in Chapter 3. This approach involves examining transdiagnostic symptom dimensions to identify specific psychiatric phenotypes associated with disruptions in decision-making (Gillan et al., 2016; Lee et al., 2023; Patzelt et al., 2019; Rouault et al., 2018; Seow et al., 2020, 2021; Seow and Gillan, 2020; Wise and Dolan, 2020). To achieve the large sample sizes needed methodically for machine learning applications and theoretically to capture the wide range of symptom presentations, this method leverages the benefits of large-scale online data collection (Gillan and Daw, 2016). Specifically, it uses factor analysis, an unsupervised machine learning algorithm applied to standard questionnaires covering nine diagnostic categories. This method serves two key purposes: to reduce the collinearity between the questionnaire scores and to explore the possibility that a more parsimonious latent transdiagnostic structure could explain deviations in behavior. Previous research revealed a robust 3-factor latent structure: 'Compulsive Behavior and Intrusive Thought,' 'Anxious-Depression,' and

'Social Withdrawal' (Gillan et al., 2016; Hopkins et al., 2022; Rouault et al., 2018). Interestingly, the transdiagnostic compulsivity factor is specifically linked to deficits in goal-directed control, which has a relatively well-established neurobiological basis (Dolan and Dayan, 2013). These insights underscore the importance of examining normal variability in psychopathology to better understand the neurocognitive basis of psychiatric dimensions that can span across different disorders.

1.6 Thesis outline

This thesis employs a multi-method approach to investigate the cognitive and computational characteristics of risk-taking behavior, combining controlled experimental settings with more ecologically valid methods such as virtual reality. Across the studies, the focus is on how internal (e.g. psychiatric traits and states) and external factors (e.g. framing, threat characteristics) shape different types of risk-related behaviors — specifically risk aversion, approach-avoidance conflicts, and escape responses. Chapter 2 explores the influence of a heightened anxious state on risk preferences in clinically anxious patients. Chapter 3 investigates the role of transdiagnostic psychiatric traits on approach-avoidance conflict. Chapter 4 shifts the focus to an innovative approach, using immersive virtual reality to study real-time decision-making in dynamic and naturalistic environments and investigate the effect of various internal and external factors on escape decisions. The concluding Chapter 5 synthesizes these findings, highlighting broader insights into risk-taking and outlining limitations as well as future research directions.

2
Risk Aversion

The content of this chapter has been published as a research article in Sporrer, J. K., Johann, A., Chumbley, J., Robinson, O.J., & Bach, D. R. (2024). Induced Worry Increases Risk Aversion in Patients with Generalized Anxiety. *Behavioural Brain Research*. doi.org/10.1016/j.bbr.2024.115192.

2.1 Introduction

Generalized Anxiety Disorder (GAD) is the most common anxiety disorder with lifetime prevalence estimates ranging between 6% (Kessler et al., 2012) and 14% (Moffitt et al., 2010) and causes substantial psychological, social, and economic costs (Beddington et al., 2008; Craske et al., 2017; Mohammadi et al., 2020). GAD remains a complex and challenging condition to treat effectively (Newman et al., 2013): only around half of the patients respond to pharmacological (Reinhold et al., 2011) or psychological (Westen and Morrison, 2001) treatment, and more than half of these do not reach full remission. This lack of effective treatment for many patients warrants further research into the specifics of these disorders with the potential to identify new therapeutic targets.

The defining feature of GAD is persistent, excessive, and uncontrollable worry. This occurs typically in the form of "worry chains," that is, chains of thoughts about potential negative outcomes that concern several life domains such as health, finances, relationships, work, and safety (Association, 2013). Worry is associated with feelings of anxiety but also with low mood and other depressive symptoms to a similar extent (Beard et al., 2016; Hong, 2007; Olatunji et al., 2010; Vîslă et al., 2022). Indeed, Major Depressive Disorder (MDD) and GAD are highly comorbid, and their genetic risk factors are closely correlated (Baxter et al., 2014; Kendler et al., 2007). Hence, worry might be an important core transdiagnostic process (Kertz et al., 2012; Olatunji et al., 2010) with somewhat distinct features in different diagnostic groups that may be linked to its uncontrollable and distressing nature (Hirsch et al., 2013; Ruscio and Borkovec, 2004).

Clinical observations suggest that generalized anxiety (GA) is associated with negative evaluation of the environment and a tendency to overestimate negative outcomes within worry chains (Beck et al., 1985). In laboratory studies, GAD and MDD patients consistently interpret ambiguous events or information as negative, which is known as the interpretation bias (Aylward et al., 2020; Hirsch and Mathews, 1997, 2012; Hirsch et al., 2016; MacLeod and Cohen, 1993). This contrasts with healthy people who favor benign (i.e. neutral or positive) interpretations of ambiguous events (Hirsch and Mathews, 2012). When GAD and MDD patients are encouraged to adopt more benign interpretations, their worry is reduced, suggesting that interpretational bias may contribute to the emergence and persistence of negative thought intrusions (Feng et al., 2022; Hayes et al., 2010; Hirsch et al., 2018). Similarly, highly anxious individuals, including GAD patients, tend to assign a higher probability and subjective cost of negative events in hypothetical everyday life scenarios compared to less anxious individuals (Berenbaum et al., 2007; Butler and Mathews, 1983, 1987; Stöber, 1997), although other studies yielded less clear results (Blair et al., 2017; Hockey et al., 2000; Mitte, 2007). In particular, Mitte (2007) found a relationship only when probabilities were described verbally (as in many other studies) but not when they were described numerically. Thus, individual differences in verbally describing costs and probabilities and a tendency to self-blame in everyday-life scenarios (Panayiotou et al., 2014) might have contributed to the aforementioned results.

To achieve consistent quantification, more recent works (Charpentier et al., 2017; Ernst et al., 2014; Xu et al., 2020) have relied on economic preference tests in which outcomes are expressed in terms of monetary or points-based payoffs, allowing the use of standard economic decision-making models such as Prospect Theory (cf. Section 1.5.3 for more details on this model). Within this framework, studies have demonstrated that GAD adults (Charpentier et al., 2017) and GAD adolescents (Ernst et al., 2014) exhibited similar heightened sensitivity to negative outcomes (loss aversion) as healthy controls. However, GAD patients exhibited a higher preference for smaller certain rewards over risky gambles with potentially higher returns (risk aversion) than controls (Charpentier et al., 2017). In other words, GAD patients were characterized by enhanced risk aversion but not loss aversion compared to healthy people (Charpentier et al., 2017).

This finding raises an important question, namely whether risk aversion is a stable and potential causal factor in the initiation of worry chains or whether it is a consequence of phases of heightened worry levels and thus variable over time. In the latter case, it might be triggered by worry chains like thoracic pain can be triggered by a cardiac stress test. Crucially, many GAD patients experience excessive worry intermittently over many hours but can be worry-free for parts of a day or even a few days. Indeed, triggering worry chains amplified over-expectation of negative events in a non-economic setup (Butler and Mathews, 1987). The study by Charpentier et al. (2017) used negative emotional priming on a proportion of their economic gambles which could have triggered risk aversion. The present study sought to disentangle state-like and trait-like uncertainty aversion by conducting an economic preference test (Wang et al., 2010) before and after a worry induction (similar to Ruscio and Borkovec, 2004). We hypothesized that risk aversion is heightened in MDD+GA after worry induction compared to baseline to a greater extent than in MDD-GA. This would suggest that enhanced risk aversion is a state-like feature of generalized anxiety triggered by worry rather than causing it. To gauge any potential influence of MDD on results, we additionally recruited an HC group.

2.2 Methods

2.2.1 Participant Details

Sixty-three participants (29 females; mean age ± SD = 42.17 ± 11.78) took part in the study. This included 40 in-patients from an affective disorder unit with partly remitted severe depressive episode (MDD), some with (MDD+GA; N = 16, 7 females; age = 41.20 ± 9.94) and others without (MDD-GA; N = 24, 12 females; age = 44.00 ± 11.50) GA symptoms. All MDD+GA patients met the DSM-IV criteria for GAD but are not classified as having GAD due to criterion F which specifies that symptoms should not be exclusively present during an MDD episode – a condition that cannot be confirmed during a depressive episode. The proportion of patients in these two subgroups reflects their prevalence in the recruiting clinical unit. Additionally, 23 healthy controls (HC; 10 females; age = 41.00 ± 13.40) were recruited from the general population using university advertisements and across personal networks to approximate age and gender distribution of the other groups.

Diagnosis and additional comorbid disorders of each participant, including HC, were assessed by a psychiatrist using the Structured Clinical Interview for DSM-IV (SCID-I/SCID-II; First and Gibbon, 2004) and recorded (cf. Appendix Table A.1).

All participants gave written informed consent before starting the experiment and received a monetary compensation. The experiment complied with all relevant ethical regulations and was approved by Governmental Ethics Committee (KEK-ZH 2013-0470).

2.2.2 Experimental Design

Participants completed the Dynamically Optimized Sequential Experimentation (DOSE) task (Wang et al., 2010), an adaptive task with 20 trials dynamically adjusted to estimate a participant's parameters in a prospect-theoretic decision model (cf. Data Analysis section 2.2.4). On each trial, participants chose between a guaranteed amount and an uncertain gamble with two equiprobable outcomes. Notably, one of these gamble outcomes always exceeded the guaranteed amount (cf. Appendix Figure A.1 for examples). In gain-only trials, the gamble offered either a positive gain or nothing, while mixed trials included both a gain and a loss. The amounts were adapted on each trial based on previous choices such as to maximize the expected gain in information about that participant's parameters (Wang et al., 2010). This adaptive procedure allowed to estimate a 3-parameter risk preference model (loss aversion, risk aversion, and choice consistency). This task was repeated after the worry induction (described below). For each of the two task sessions, the outcome of one randomly chosen decision was paid out to the participants to ensure that each trial is treated like a one-shot gamble. This amount was then added to a fixed monetary compensation contingent on the duration of the entire study.

As worry chains can last many hours and the assessments had to take place on the same day, the worry induction always followed the first instance of the DOSE. In the first phase of the worry induction (similar to Ruscio and Borkovec (2004)), the experimenter asked participants to enumerate for 2 minutes all the topics that worried them recently. Then they were asked to estimate the importance of these worries (from 1 unimportant to 7 extremely important) and how much time they spent worrying about them in the previous week (from 0 to 100%). The second phase aimed at catastrophizing their most time-consuming worrisome topic. Participants were prompted to explain each worry regarding that particular topic. Subsequently, they were asked what they would perceive as frightening or undesirable if the worrisome event actually happened. Ultimately, participants had to estimate the number of negative events or steps that would need to unfold in the catastrophic scenario. This was repeated until participants (a) refused to continue, (b) gave the same answer 3 times, (c) spent 10 minutes with the task. Afterwards, the experimenter read out their answers aloud and participants were asked to estimate the probability that each of these events would happen.

The most worrisome topics were work and education for HC and MDD-GA, and health for MDD+GA. Both patient groups exhibited higher levels of discomfort prior to the induction ($t(55) = 4.54, p < .001$), spent more time worrying ($t(58) = 3.65, p < .001$) than HC, and at a trend level indicated a higher number of steps for the disaster scenario to unfold ($t(61) = 1.84, p = .07$). There were no significant differences between the two patient groups in terms of these three variables (all $p > .05$, cf. Table 2.1). Similarly, no significant differences were found among the three groups regarding their estimated probability of the negative event occurring (all $p > .05$, cf. Table 2.1). Nevertheless, the three groups significantly differed in the perceived importance of these worries. MDD+GA assigned higher importance compared to MDD-GA ($t(38) = 2.19, p < .05$) and MDD-GA attributed more importance compared to HC ($t(43) = 2.10, p < .05$, cf. Table 2.1).

2.2.3 Psychiatric Questionnaires

Participants completed a series of self-report psychiatric questionnaires assessing depression using Montgomery-Asberg Depression Rating Scale (MADRS; Montgomery and Asberg, 1979), Beck's Depression Inventory (BDI; Beck et al., 1961), and Hamilton Depression Rating Scale (HAM-D; Hamilton, 1960), anxiety using

Hamilton Anxiety Rating Scale (HAM-A; Hamilton, 1959) and the trait version of State-Trait-Anxiety-Inventory (STAI; Spielberger, 1983), and worry using Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) and the Worry Domains Questionnaire (WDQ; Tallis et al., 1992). These were not used as diagnostic tools but rather to stratify the severity and type of symptoms. The scores are reported in Table 2.1 and their distribution in Figure A.2.

2.2.4 Quantification and Statistical Analysis

Our main dependent variables were the participant's three risk preference parameters, namely λ (loss aversion), ρ (risk aversion), and μ (choice consistency) using a prospect theory utility function with power utility from DOSE (Wang et al. (2010), cf. 1.5.3 for more details on the model). Importantly, the parameters were estimated separately for each participant during the DOSE, as each participant's gambles were optimized for their particular economic preferences. The DOSE method has improved parameter recovery for ρ and related preference parameters, as it uses a Bayesian adaptive approach to select the most informative sequence of questions for each participant, updating beliefs after each response. Simulation studies demonstrate that DOSE can recover ρ with roughly twice the accuracy of traditional elicitation methods, even with fewer questions, and produces more granular and stable individual-level estimates (Wang et al., 2010).

To test whether the effect of the worry induction on the risk preference parameters differs in anxious and depressive patients, we conducted a 2 x 2 repeated-measure ANOVA with time points (baseline, after worry induction) as the within-subject factor and patient groups (MDD+GA, MDD-GA) as between-subjects factors. Means of the risk preference parameters before and after the induction were then compared with post-hoc independent two-sample t-tests between MDD+GA and MDD-GA, and both MDD and HC as supporting tests. As control analyses, we added the most distinctive anxiety (HAM-A) and depression (HAM-D) scores first separately and then jointly as a within-subjects factor in addition to condition and timepoint in a linear mixed effect model. The Log10 Bayes Factors (LBF) were calculated for each test. For the ANOVA and mixed effects models, we compared models including the effect with equivalent models excluding the effect. For the post-hoc contrasts, we used Bayesian analysis of one- and two-sample designs.

As secondary outcome variables, the propensity to choose the gamble over the certain reward option was calculated for each participant. While this variable cannot be used as an index for risk-taking, as each participant made decisions about differently valued gambles, it can be computed separately for mixed and gain-only gambles. The same procedure as mentioned above (repeated-measure ANOVA and post-hoc t-tests) was employed to compare the propensities between groups. Results are reported in A.2.

Additional exploratory analyses were conducted to compare other collected variables such as questionnaire scores and worry-related measures (e.g., time spent worrying, importance) using the same contrasts as mentioned earlier. We also conducted bi-variate correlations within each group and across all participants.

Before running the statistical tests, the distributions of all outcome variables were checked. The distribution of λ, ρ, and μ were slightly positively skewed. Previous studies used log transformation to reduce skewness (Charpentier et al., 2017), but in our data, this resulted in even more skewness. Therefore, we applied the square-root transformation, which is weaker and reduced the skewness of ρ in our data to an acceptable level (between -.5 and .5). We did not transform any other variables as this did not reduce their skewness. Additionally, in line with previous work, the negation of ρ was taken as the final index of risk aversion such that higher values of ρ indicated higher levels of risk aversion.

Finally, we conducted supplementary analyses to more closely replicate the analyses of Charpentier et al. (2017). In their study, anxious patients and healthy controls made economic choices that were embedded in an emotional memory task involving both emotional and neutral faces as well as objects. They estimated risk and loss aversion across all trials rather than separately for each emotion condition, as this was the winning model (according to lowest Bayesian information criterion) which outperformed all other models. Consequently, they did not differentiate between trials that may have involved emotional priming and those that did not. To better match their methodology in order to replicate the results, we combined our participant variables without distinguishing between baseline and post-induction time points; thereby collapsing time points.

2.3 Results

2.3.1 Task Characteristics

As expected, healthy people were loss aversive and risk-seeking before and after the worry induction (cf. Figure 2.1). Loss aversion (baseline: $t(22) = 6.26, p < .001, LBF = 3.86$; post-induction: $t(22) = 6.31, p < .001, LBF = 3.91$) and risk aversion (baseline: $t(22) = -28.2, p < .001, LBF = 15.6$; post-induction: $t(22) = -30.69, p < .001, LBF = 16.35$) were both different from zero.

2.3.2 Risk Preferences

In accordance with our hypothesis, our primary analysis uncovered a significant interaction between time (baseline vs. post-induction) and condition (MDD+GA vs. MDD-GA) for risk aversion ($\eta_p^2 = .10, F(1, 38) = 4.18, p < .05, LBF = .77$; cf. Figure 2.1.A). Control analyses indicated that this was not the case for loss aversion ($F(1, 38) = .06, p = .81, LBF = -1.10$; cf. Figure 2.1.B) or choice consistency ($F(1, 38) = 1.51, p = .22, LBF = -.65$; cf. Figure 2.1.C). Post-hoc t-tests revealed that risk aversion was enhanced after worry induction in MDD+GA compared to MDD-GA patients ($t(38) = 2.08, p < .05, LBF = .22$) and compared to healthy controls ($t(38) = 2.21, p < .05, LBF = .79$). In an exploratory control analysis, there was no significant difference between MDD patients and HC ($t(38) = -0.60, p = .55, LBF = -.52$). Before the induction, risk aversion was descriptively similar between all three groups. Across both patient groups, we found a significant effect of time on choice consistency ($F(1, 38) = 8.36, p < .01, LBF = 1.99$), which was not present for risk aversion ($F(1,38) = 0.29, p = .59, LBF = -1.40$) or loss aversion ($F(1,38) = 0.93, p = .34, LBF = -1.10$).

Next, we conducted control analyses to address how our primary results relate to psychiatric symptom scores. We added the most distinctive anxiety questionnaire (HAM-A) as a covariate. This revealed an interaction between timepoint and anxiety in risk aversion ($t(32) = 2.50, p < .05, LBF = .77$), while the interaction between timepoint and patient group was no longer significant ($t(32) = 0.99, p = .33, LBF = -0.79$), suggesting that the group differences we found are indeed explained by anxiety. However, we also found a 3-way interaction between timepoint, anxiety, and group ($t(32) = -2.29, p < .05, LBF = 1.32$; cf. Figure 2.2.A). Taken together, these findings indicate that the effect of worry induction scaled with HAM-A scores only in the MDD+GA group, not in the MDD-GA group, pointing to qualitative differences between these two groups beyond questionnaire scores.

When adding the most distinctive depressive symptom score (HAM-D) as a covariate in addition to anxiety, the interactions between timepoint and anxiety ($t(28) = 2.88, p < .01, LBF = 21.99$) and between timepoint, anxiety, and group remained significant ($t(28) = -2.14, p < .05, LBF = 20.09$), suggesting that the relation to patient group and anxiety scores is not explained by depression. There was no significant effect when only HAM-D was added as a covariate.

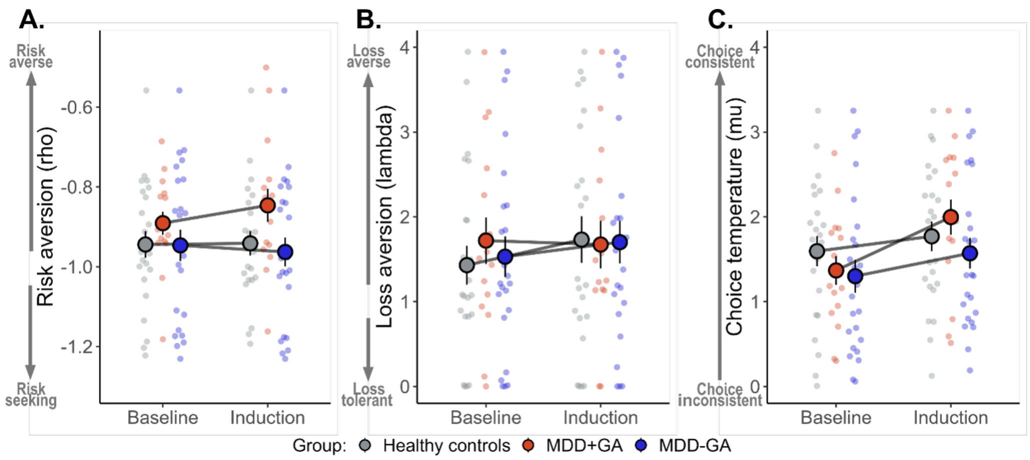

Figure 2.1: **Enhanced risk aversion but not loss aversion in MDD+GA patients after worry induction compared to MDD-GA.** Mean estimates of risk aversion **(A)**, loss aversion **(B)**, and choice temperature **(C)** plotted separately for MDD+GA, MDD-GA, and HC. There were no significant group differences in baseline levels (i.e., before worry induction) of risk aversion (MDD+GA: -.89 vs. MDD-GA: .-95, t = -1.14, p = .26; MDD: -.92 vs. HC: -.94, t = -.47, p = .64), loss aversion (MDD+GA: 1.72 vs. MDD-GA: 1.53, t = .52, p = .61; MDD: 1.61 vs. HC: 1.43, t = .61, p = .55) or choice consistency (MDD+GA: 1.37 vs. MDD-GA: 1.3, t = .26, p = .79; MDD: 1.33 vs. HC: 1.6, t = -1.21, p = .23).
Error bars represent SEM.

Table 2.1: **Demographics, Questionnaire Scores, and Participants' Characteristics.**
*T-test results: * (<.05), ** (<.01), *** (<.001), tr: trend (<.10), n.s.: non-significant (>.10). Mean ± SD.*

	MDD+GA	MDD-GA	MDD+GA vs. MDD-GA	HC	MDD vs. HC
Women:Men	7:9	12:12	n.s.	10:13	n.s.
Age (years)	41.20 ± 9.94	44.00 ± 11.50	n.s.	41.00 ± 13.40	n.s.
STAI-S Score	53.53 ± 9.46	49 ± 9.02	n.s.	26.7 ± 5.68	****
STAI-T Score	47.67 ± 9.58	45.9 ± 11.97	n.s.	22.61 ± 7.00	****
HAM-A Score	22.67 ± 6.63	15.21 ± 7.37	***	2.52 ± 2.43	****
HAM-D Score	22.64 ± 6.40	15.13 ± 7.75	***	.82 ± .96	****
MADRS Score	24.31 ± 8.25	17.92 ± 7.81	*	1.09 ± 1.12	****
BDI Score	29.06 ± 8.63	22.32 ± 10.08	*	3.45 ± 3.76	****
PSWQ Score	61.75 ± 13.43	54.17 ± 10.21	tr.	37.65 ± 8.57	****
WDQ Score	63.31 ± 19.11	44 ± 21.28	**	13.87 ± 9.51	****
Number of episodes	2.25 ± 1.65	2.00 ± 2.02	n.s.	n/a	n/a
Length of current episode (months)	29.38 ± 24.56	78.27 ± 279.90	n.s.	n/a	n/a
Number of comorbidities	1.56 ± 1.09	1.21 ± .41	n.s.	n/a	n/a
Discomfort levels (/100)	45.89	46	n.s.	15.57	**
Time spent worrying (%)	77.75	62.83	n.s.	43.05	****
Probability of happening (%)	71.67	64.76	n.s.	52.03	n.s.
Number of steps	7.88	8.54	n.s.	6.52	tr.
Importance (/7)	6.88	6.25	*	5.43	***

2.3.3 Effect of Anxiety and Depression on Risk Preferences

To further investigate whether the differences in MDD+GA are explained by anxiety or depression, we conducted additional exploratory analyses using the questionnaire scores. To avoid issues with interpreting p-values in multiple exploratory tests, we report only R^2 values. Across all included groups, the change in risk aversion from baseline to post-induction was not correlated with any psychiatric scores (all $R^2 < .05$). Consistent with the previous covariate analysis, it was negatively correlated with HAM-A scores within the MDD+GA group ($R^2 = .31$), but not in the other groups (both $R^2 < .05$; cf. Figure 2.2.B). Given HAM-A's strong positive correlations with depressive scores (e.g., HAM-A and HAM-D: $R^2 = .44$; cf. Appendix Figure A.3), we controlled for HAM-D using partial correlations. The correlation between HAM-A and risk aversion differences in MDD+GA increased when controlling for HAM-D ($R^2 = .36$), MADRS ($R^2 = .33$), and BDI scores ($R^2 = .46$). A similar effect was observed in the change in propensity to gamble in gain-only trials from baseline to post-induction (cf. Appendix Figure A.1).

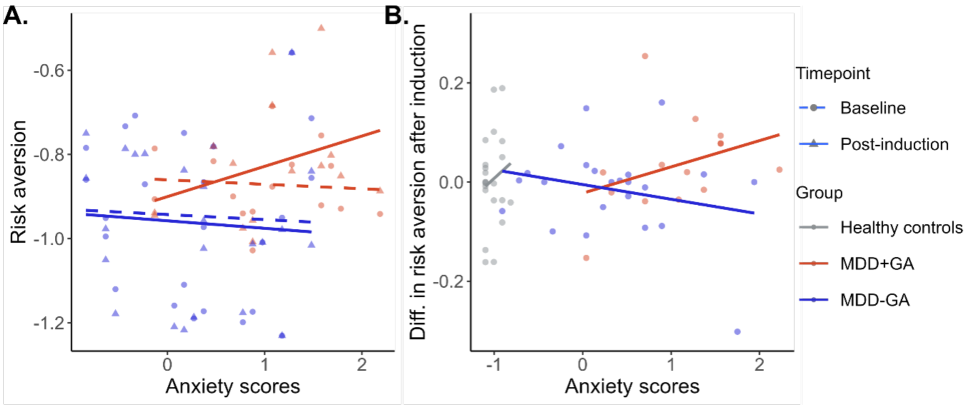

Figure 2.2: **Effect of anxiety on risk aversion in MDD+GA but not MDD-GA patients after worry induction.** **(A)** Regressions between risk aversion and scaled anxiety (HAM-A) scores. Dotted line corresponds to baseline and full line to post-induction. **(B)** Correlation between the difference in risk aversion after worry induction and scaled anxiety (HAM-A) scores.

2.3.4 Replication of Analyses from Previous Work

We successfully replicated previous findings, demonstrating an enhanced risk aversion in MDD+GA compared to both MDD-GA ($t(78) = 2.36, p < .05, LBF = .30$) and HC ($t(78) = 2.21, p < .05, LBF = .25$) when collapsing time points according to the original experimental design. No significant differences were observed between groups in loss aversion or choice consistency (cf. Appendix Table A.2). For findings related to the propensity to gamble, refer to Appendix A.2.

2.4 Discussion

The present study sought to clarify whether economic risk avoidance in GA is a trait-like causal factor or a state-like feature triggered by worry. We found that patients with MDD+GA are characterized by increased risk aversion after a worry induction but not at baseline when compared to MDD without GA. As anticipated, GA had no relation with loss aversion. This is in accordance with our hypotheses and previous studies showing that anxious patients do not overweigh losses (Charpentier et al., 2017; Ernst et al., 2014) or punishments per se (Aylward et al., 2019).

By incorporating individuals diagnosed with MDD (without GA symptoms) and healthy participants, we were able to compare the decision-making behaviors of anxious patients to the two control groups to discern specific differences associated with anxiety and depression. In particular, by comparing MDD patients with and without GA, we can rule out that in-patient status or unspecific patient characteristics shared by both MDD groups contribute to our results. Little is known about how comorbid depressive disorders impact risk preferences, but evidence seems to suggest that non-suicidal depressed patients exhibit similar risk and loss aversion as healthy controls (Baek et al., 2017; Hadlaczky et al., 2018). Our results are in line with this and suggest that heightened risk aversion is associated with anxiety symptoms in depression, rather than depression as such.

When considering psychiatric questionnaire scores, several facets of anxiety could be separated. MDD+GA patients reported higher worry and anxiety than MDD-GA, as measured by WDQ, PSWQ, and HAM-A, but not when measured by STAI. It has been suggested that STAI lacks divergent validity as it does not distinguish well between clinically anxious or depressed samples (Knowles and Olatunji, 2020). In addition to higher anxiety levels, MDD+GA also reported higher levels of depression than MDD-GA, as measured by HAM-D, MADRS, and BDI. Previous studies found that comorbid depression–anxiety groups have more severe symptoms, with higher levels reported of depression and anxiety than pure depression or anxiety disorder (Penninx et al., 2011; Schoevers et al., 2003, 2005; Starcevic et al., 2007; ter Meulen et al., 2021). The mixed presentation is also associated with a worse prognosis even when controlling for differences in severity and duration of symptoms (Penninx et al., 2011). Nonetheless, we also did not find any continuous relation between the risk preference estimates and depression scores. Instead, we observed an effect of anxiety on risk aversion in MDD+GA, which became more pronounced after correcting for depression scores (but not the other way around). This pattern suggests that it is not merely the overlapping characteristics of anxiety and depression that drive this effect, but rather the distinct elements of anxiety. This underscores a unique link between anxiety in MDD+GA and risk aversion that goes beyond its strong association with depression. Additional analysis showed that this effect of anxiety arose after the worry induction, suggesting that in a state of increased worry, higher anxiety is linked to higher risk aversion. However, it is still possible that comorbid GA symptoms might rely on different mechanisms than GAD in non-depressed people. For instance, comorbid anxiety disorder was associated with higher worry across primary diagnoses including MDD, bipolar disorder, and psychosis, even after controlling for GAD (Kertz et al., 2012). Charpentier et al. (2017)'s exploratory analysis found no difference in risk aversion when comparing pure anxious patients (N=12) to anxious-depressed patients (N=13). It is important to consider that their study might have been underpowered to detect subtle distinctions, highlighting the need for further investigation in this area. Finally, it is worth noting that different mechanisms might be at play in participants with varying anxiety levels. While our results and those of Charpentier et al. (2017) suggest enhanced risk aversion but not loss aversion in GA, anxiety questionnaire scores in healthy individuals were linked to enhanced loss aversion but not risk aversion in Xu et al. (2020). This suggests that GA may be categorically different from enhanced anxiety in healthy people.

By examining risk preferences both before and after an incidental worry induction, we could determine whether the previously observed enhanced risk aversion in GA is a potential cause of worry chains, reflecting a state-like characteristic rather than an enduring trait of the condition. We found no differences between groups before the induction. Instead, we uncovered a significant interaction between time and patient groups, which implies an emotion-induced enhancement in risk aversion. While Charpentier et al. (2017) found that risk parameters were not influenced by emotional priming on the same trial, their presentation of a visual prime (e.g., fearful faces, objects) before each gambling decision might have a longer-lasting spillover effect on subsequent non-priming. We note that our experimental design did not aim to specifically investigate the effect of worry induction over any other forms of emotional priming. It could be interesting to explore whether the induction of rumination, which has traditionally been linked to MDD (Nolen-Hoeksema, 1987; Watkins and Roberts, 2020), yields different outcomes. Although worry and rumination share similarities as repetitive perseverative negative thinking processes, they differ in their content and temporal orientation. Worry focuses on potential future threats, whereas rumination centers on past negative events (Kircanski et al., 2015; Nolen-Hoeksema et al., 2008).

One of the primary factors proposed to underlie heightened risk aversion in GA is an inclination to overestimate the likelihood of negative outcomes. Our current investigation did not reveal any intergroup differences concerning participants' estimated probability of the realization of their own worries, thereby contributing to the conflicting body of evidence. This lack of effect could be explained by two non-mutually exclusive possibilities. Firstly, a distinction might exist between estimating the probability of a negative event in a general context versus for one's own personal circumstances. While prior studies focusing on general estimations found mixed evidence (Butler and Mathews, 1983; Mitte, 2007), those that delved into self-contrasted scenarios did reveal more consistently such an impact (Butler and Mathews, 1983; Miranda and Mennin, 2007; Mitte, 2007). In our study, participants were asked to gauge the overall likelihood of their most concerning event occurring, rather than evaluating whether it was more probable for the event to happen to them as opposed to others. Secondly, GAD may rather exhibit a tendency to underestimate the probability of positive events (Blair et al., 2017; Butler and Mathews, 1983; Stöber, 1997), especially in self-contrasted scenarios. This could potentially explain why, similar to Charpentier et al. (2017), we found that anxious individuals gambled significantly less on gain-only trials (cf. Appendix A.2) without displaying any noticeable differences from the other groups in mixed gambles. This idea still aligns with the concept of interpretation bias, which suggests that individuals with anxiety tend to inaccurately interpret ambiguous stimuli as negative. Rather than solely attributing it to the overestimation of negative events, another possible explanation lies in the underestimation of positive events. Future studies should add loss-only trials in addition to the existing gain-only trials to assess whether risk aversion in the loss and gain domains differ between groups.

Conceptually, we replicated Charpentier et al. (2017) using a clinically different sample from a different country and employing a similar experimental task (cf. Appendix Table A.2). It is noteworthy that the log transformation applied in their study resulted in higher skewness within our dataset, but it also yielded stronger effects. In accordance with our data, we decided to utilize a milder transformation (e.g., square root) for parameters where it improved skewness. Moreover, Charpentier et al. (2017)'s results on risk aversion were significant regardless of the transformation (tested using their data available online). This process emphasizes the significance of verifying the adequacy of replication methods for the specific dataset and if necessary adapting the analysis accordingly.

In interpreting the results of this study, it is important to consider the possibility that observed changes in risk aversion may not be solely attributable to the worry induction intervention itself. Instead, these effects could partially reflect non-specific factors such as the passage of time or repeated exposure to the DOSE task. Repeated task performance can lead to learning effects, habituation, or changes in participant engagement due to fatigue or other. Although the study design included a within-subjects baseline and post-induction comparison, without a control group undergoing repeated testing without the worry induction, it is not possible to fully rule out the influence of time or repetition. Nonetheless, it seems unlikely that repetition or time effects would selectively impact only the anxious groups, especially since similar findings have previously been observed within a single session. Additionally, Prospect Theory model parameters show good-to-excellent two-week test-retest reliability (Mkrtchian et al., 2023), underscoring their stability over time without intervention. Thus, it seems unlikely that the observed effects are due to temporal factors.

This study has several limitations that warrant consideration. Firstly, our relatively small sample size might obscure small differences between groups at baseline. Despite the potential for some contrasts to be underpowered, the robustness of our main analysis—examining the interaction between time and patients—remains intact and was not underpowered (cf. Appendix A.1 for more details on power analyses). Of note, MDD patients without GA had relatively high worry scores in the PSWQ, which might have influenced their baseline risk preferences. However, this alone is unlikely to explain the lack of group differences at baseline, since both groups were not appreciably different from healthy participants. Additionally, our worry induction procedure was established as effective in previous research Ruscio and Borkovec (2004). However, we did not ask participants to rate their worry and discomfort before and after the worry induction procedure, we could not verify how effective the procedure was in individual participants. Another limitation pertains to the generalizability of our findings. Our sample, composed exclusively of psychiatric in-patients with severe depression, might not reflect the characteristics of broader populations, such as those with milder forms of depression or individuals in outpatient settings. Finally, our interpretation partially relies on the DOSE task's capacity to measure both trait-level and state-level risk aversion. However, we lack definitive data to support this capability. Future research should aim to clarify this aspect to strengthen the validity of interpretations based on this measure.

To conclude, better understanding differences between patient groups with comorbidities is crucial to allow better assessment of potential treatments (Harmer et al., 2011) or the development or refinement of relevant cognitive interventions (Robinson et al., 2013). The current study suggests that anxious individuals exhibit distinct patterns of decision-making compared to depressed and healthy controls, characterized by heightened risk aversion but unchanged loss aversion after a worry induction. These findings could directly inform, improve, or lead to the development of psychological interventions aimed at addressing both mental health conditions. For example, therapeutic strategies for GA could benefit from focusing on reducing sensitivity to uncertainty rather than negative outcomes per se.

3

Behavioral Cautiousness

The content of this chapter is based on the results published as a preprint in Sporrer, J. K., Melinscak, F., & Bach, D. R. (2024). Transdiagnostic psychiatric symptom dimensions predict behavioral cautiousness. *PsyArxiv*.
doi.org/10.31234/osf.io/3u47k.

3.1 Introduction

Avoiding threats to one's integrity is a common and adaptive behavior (LeDoux et al., 2017). Yet, real-life situations often mandate pursuing rewards in circumstances that entail danger, exemplified by risky foraging. Such situations give rise to conflicting motivations to approach the reward, and avoid the threat (McNaughton et al., 2016). Solving this dilemma requires a careful assessment of available options, and the probabilities and magnitudes of the outcomes associated with each of them (Aupperle and Paulus, 2010). Approach-avoidance conflict (AAC) tests encapsulate this situation in a well-defined laboratory setting (Bach, 2021; Rodgers et al., 1997). Anxiolytic substances—i.e., drugs that reduce subjective feelings of anxiety in clinical conditions—crucially alter animals' cautiousness in such tasks (cf. Cryan and Sweeney 2011 for a review). This has led to a long history of employing them as a primary preclinical model in anxiety disorder research, and for the development of anxiolytic drugs. More recently, AAC tasks have been translated to humans, and validated by cross-species similarity of underlying neural substrates including their sensitivity to benzodiazepines and other anxiolytics (cf. Bach 2021 for a review). An initial assumption in this translational effort was that AAC task readouts should relate to trait anxiety (Aupperle et al., 2011) and might even be used as diagnostic tests for anxiety disorders (cf. Bach 2021; Stephan et al. 2016 for a review). While this appears intuitively plausible, evidence that cautiousness metrics such as approach rate in AAC tasks relates to trait anxiety in common questionnaires is inconsistent (cf. Fricke and Vogel 2020 for a review). Specific relations reported in individual studies (Bach, 2015; Fung et al., 2019; Walz et al., 2016) were largely not replicated (Biedermann et al., 2017; Gromer et al., 2021; Kallai et al., 2007; Struijs et al., 2018, 2017; Walz et al., 2016), including in a large sample of healthy young people (N=781; Bach et al., 2020). Similarly, a large-scale clinical study found no relation between approach-avoidance behavior with anxiety disorder, depression, or substance abuse diagnosis (Smith et al., 2021). At the same time, cautiousness in AAC tasks appears to be a relatively stable trait as shown by 2-year test-retest reliability (Bach et al., 2020). This raises an important question: if anxiolytic-sensitive cautiousness, and self-reported trait anxiety, are both stable traits but not specifically related to each other, then what other psychiatric symptom dimensions predict cautiousness?

This is the question we sought to address here in two large online samples and with a comprehensive psychiatric symptom battery. To this end, we capitalized on a human AAC task (Bach, 2015) with a simple and abstract visual design that could be presented online. In this task, a participant can decide whether, and how rapidly, to approach a reward, under risk of being virtually attacked by a predator and incurring a variable loss. As with many AAC tasks, its core behavioral indices are passive avoidance, i.e., rate of avoidance decisions, and behavioral inhibition, i.e., latency to initiate approach. Previous work has consistently demonstrated that both of these indices linearly increase with increasing threat probability and threat magnitude (i.e., potential loss; Abivardi et al. 2020; Bach 2015, 2017; Castegnetti et al. 2020; Khemka et al. 2017), a group effect that we replicated in the present sample. Notably, behavioral inhibition in this task is not reward-maximizing but might be optimal if the agent has assumptions about the temporal coupling of reward and threat (Bach, 2015). Thus, in a secondary task, we assessed people's implicit beliefs about such reward correlations (Bach, 2017). We then asked whether passive avoidance and behavioral inhibition, and their relation to threat probability and magnitude, are linked to psychiatric symptom dimensions. For comparability with previous work on other behavioral dimensions, we utilized an exhaustive clinical questionnaire battery with a known three-factor structure (Gillan et al., 2016; Hopkins et al., 2022; Rouault et al., 2018): 'Compulsive Behavior and Intrusive Thought (CIT)', 'Anxious-Depression (AD)' and 'Social Withdrawal (SW)', which we also replicated in our samples. We added further anxiety questionnaires, and questionnaires found relevant in previous work on AAC tasks. As we had no prior hypothesis on which symptom dimensions or questionnaires would relate to cautiousness, we opted for a rigorous exploration-confirmation approach. Hypotheses were first generated based on analysis of a discovery sample (N = 315), and then confirmed after pre-registration in a second sample (N = 690).

3.2 Methods

3.2.1 Participant Details

Adult participants (> 18 years old) were recruited online using Amazon Mechanical Turk (MTurk, https://www.mturk.com/). We included 315 participants in the discovery sample (149 females, mean age $\pm$ SD: 36.40 $\pm$ 11.01; October-November 2021) and 690 in the confirmation sample (338 females, mean age $\pm$ SD: 33.41 $\pm$ 9.89; February-March 2023; cf. Table B.1). In total, the combined sample comprised 1005 participants (487 females, mean age $\pm$ SD: 34.35 $\pm$ 9.89).

Participants who completed the task per protocol were included if they did not meet any of the following exclusion criteria:(1) Pressed (or did not press) the same button on the keyboard in more than 95% of the trials, (2) Responded incorrectly to all three attention checks in the questionnaires (cf. Appendix B.1.1), (3) Returned to the safe place in fewer than 50% of the trials, indicating lack of understanding of the task, (4) Performance at chance level. In the discovery sample, we excluded 189 (37%) participants out of 504, leaving 315 participants for analysis (cf. Figure B.1 for an exclusion flowchart). In the confirmation experiment, we excluded 478 (41%) participants out of 1168, leaving 690 participants for analysis. Although there are growing concerns about online data quality (Burnette et al., 2022a,b; Chandler et al., 2014; Zorowitz et al., 2023), further validation tests suggest our findings are unlikely to come from spurious results (Zorowitz et al. 2023; cf. Appendix B.2.2).

Additionally, to ensure that participants understand the task, they had to correctly answer five questions on a task comprehension test, after reading the instructions and before starting each task. They could reread the instructions after each try, but if they failed the test more than five times, they could not take part in the experiment.

After reading an information sheet about the experiment, participants confirmed their consent online. All procedures were in accordance with the Declaration of Helsinki and local regulations. The study was approved by the Governmental

Ethics Committee (Kantonale Ethikkommission Zurich, BASEC 2016-00068). Participants were paid a base sum of 10$ plus a bonus ranging from 2$ to 14.5$ (with a mean of 12$) conditional on task performance and on passing several attention checks.

3.2.2 Experimental Design

Participants performed an AAC task embedded in an online computer game (cf. Figure 3.1), based on previous studies in a lab setting (Abivardi et al., 2020; Bach, 2015, 2017; Castegnetti et al., 2020; Khemka et al., 2017). On each trial of this "scoop-and-run" task, participants could collect one monetary token (approach motivation) under threat of getting caught by a predator and consequently losing an explicitly signaled number of tokens (avoidance motivation). 144 trials were presented in randomized order, evenly distributed across 6 different levels of potential token loss, and three different threat probabilities[1]. At the start of each trial, the participant was in a “safe place”, the bottom grid block, and was tasked to decide whether to collect a token that would appear on either side after an interval that was the sum of a fixed delay (500 ms) and a random sample from a truncated gamma distribution ($\kappa = 2, \theta = 1, \mu = 2\,s$), truncated at $6\,s$. In case the participant did not collect the token, it disappeared after a variable time drawn from the same gamma distribution. Below the grid, the potential loss of the current trial (0-5 tokens) was indicated by red tokens. A "sleeping" predator was waiting opposite the safe place in the top grid block and could catch the participant if they were outside the safe place. Wake-up of the predator followed a Bernoulli process, independently determined in successive time bins of 20 ms. Three threat levels, corresponding to different wake-up rates, were represented by three different frame colors (cyan, yellow, and purple) with a random relation of color and threat level. If the player was outside of the safe place for 100 ms, a value established in previous work (Bach, 2015), then this resulted in catch probabilities of p1 = .1, p2 = .2, or p3 = .3. The

[1]While the term "threat" may be open to interpretation, its use here aligns with prior literature employing this task (Bach, 2015, 2017) and similar risky-foraging paradigms (e.g. Mobbs et al., 2007), and also reflects how the experimental design was framed to participants.

actual catch rates depend on the participants' individual return latencies. Crucially, threat probabilities were not explicitly instructed but it was emphasized that the predators were differently dangerous and that the participants needed to learn this difference. If the participant got caught, then the token disappeared, the predator turned red, and the indicated potential loss was realized. In all cases, the trial ended 1 s after the pre-determined token disappearance time.

The second block comprised 57 trials from a different task, the predator exposure task, randomly interspersed with 15 AAC task refresher trials. As the graphical set-up is similar, the type of task was signaled with either a grey token on AAC task trials or a grey circle for predator exposure task trials. In the predator exposure task, participants could not move to the sides of the grid and were instructed to expose the awake predator by pressing the up-arrow key (i.e. a motor action unavailable on AAC trials). If the predator was awake during the attempt, it turned red and the trial would end early. Otherwise, it would turn black and the trial would continue to the pre-determined end. This feedback allowed participants to update their knowledge of the experimental statistics (which were maintained throughout all blocks), according to which the probability of being awake was independent of time or token appearance and was randomly determined at each capture attempt. Participants were explicitly informed that the tokens were irrelevant to the task and could not be collected. If participants correctly believed that catch probabilities were constant, approach time would not depend on token appearance and the optimal strategy would be to approach the predator as soon as the trial starts to shorten the experiment. On the other hand, if participants incorrectly assumed a temporal threat-reward correlation, as suggested in previous work (Bach, 2015, 2017), then the reward-maximizing strategy would be to approach whenever a token appeared, at the maximum of their subjective threat wake-up function.

After the two tasks, participants were asked to estimate the probability of getting caught if they left the safe place for each of the three threat levels on a continuous scale ranging from 0 to 100% (with a 1% step increment). The association between the color of the threat and threat level depended on behavior and had to be implicitly learned during the experiment.

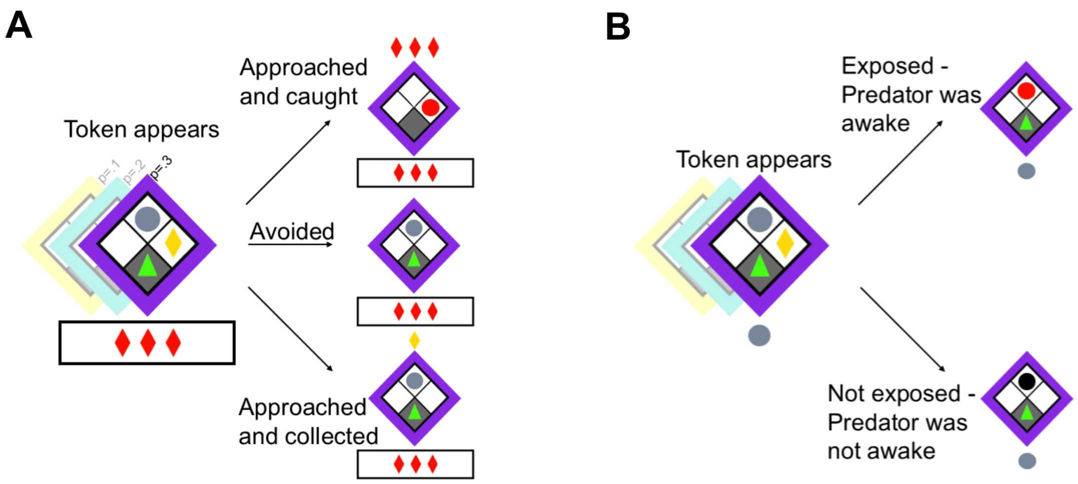

Figure 3.1: **Experimental set-up for the approach-avoidance conflict (AAC) task. (A)** where participants could approach a reward at the risk of getting caught by a predator with varying levels of threat probability (shown by the color of the grid, learned beforehand by direct experience) and magnitude (signaled by the number of red tokens). In the predator exposure task **(B)**, participants had to guess when the predator awoke, following the same threat probabilities as in the AAC task. As such, the probability of awakening was independent of time or token appearance and was randomly determined at each capture attempt.

3.2.3 Psychiatric Questionnaires

After completion of the behavioral tasks, participants were asked to answer a battery of self-report questionnaires based on previous work (Gillan et al., 2016; Rouault et al., 2018), assessing a range of psychiatric symptoms including depression (Zung Self-Rating Depression Scale, ZDS, Zung 1965), generalized anxiety (Generalized Anxiety Disorder 7-item scale, GAD-7, Spitzer et al. 2006), schizotypy (Short Scales for Measuring Schizotypy, SSMS, Mason et al. 2005), impulsivity (Barratt Impulsiveness Scale-11, BIS-11, Patton et al. 1995), Obsessive-Compulsive Disorder (Obsessive-Compulsive Inventory-Revised, OCI-R, Foa et al. 2002), social anxiety (Liebowitz Social Anxiety Scale, LSAS, Liebowitz 1987), eating disorders (Eating Attitudes Test, EAT, Garner et al. 1982), apathy (Apathy Evaluation Scale, AES, Marin et al. 1991), alcoholism (Alcohol Use Disorders Identification Test, AUDIT, Saunders et al. 1993), and a short IQ evaluation (International Cognitive Ability Resource, Condon and Revelle 2014).

In addition to this battery, they also answered questionnaires assessing trait anxiety (State-Trait Inventory for Cognitive and Somatic Anxiety, STICSA-T, Ree et al. 2008), sensation seeking (Brief Sensation Seeking Scale, BSSS, Hoyle et al. 2002), and the daringness subscale of the Child and Adolescent Disposition Scale (CADS, Lahey et al. 2010) to replicate and extend previous findings with other AAC tasks (Bach, 2015; Bach et al., 2020). We did not include questionnaires based on analysis of rodent AAC tasks (e.g., reinforcement sensitivity inventories) as there is no empirical support for a relation with behavioral readouts in humans. See Appendix for questionnaire scores (cf. Table B.1), distributions (cf. Figure B.2), and correlations (cf. Figure B.3).

As a measure of data quality, attention checks were added to three questionnaires to ensure participants read the questions carefully (cf. Appendix B.1.3).

3.2.4 Quantification and Statistical Analysis

The exclusion criteria, pre-processing steps (cf. Appendix B.1.2), and analysis plan were based on a discovery sample and pre-registered before the confirmation sample was recruited (osf.io/5hmgk/; registered on 05 January 2023). The analysis plan encompassed 3 steps. Step 1 was a manipulation check and consisted of replicating the 4 behavioral group-level effects of the task previously shown in lab settings (Abivardi et al., 2020; Bach, 2015, 2017; Castegnetti et al., 2020; Khemka et al., 2017). Step 2 was a check of the questionnaire battery and consisted of replicating a 3-factor solution of the transdiagnostic symptom dimensions (Gillan et al., 2016; Hopkins et al., 2022; Rouault et al., 2018). For steps 1-2, every single precondition needed to be confirmed at $p < .05$ to progress to the next step; hence there was no multiple comparison correction. The goal of step 3 was to assess our main research questions and investigate associations between behavioral variables of interest (indexing passive avoidance and behavioral inhibition), and psychiatric symptom dimensions extracted in step 2 and individual psychiatric questionnaires. Holm-Bonferroni method was applied to correct for multiple comparisons across 20 pre-registered hypotheses in the confirmation sample.

Step 1 – Replication of behavioral inhibition and passive avoidance

To test whether any of the behavioral variables were influenced by the independent variables, we conducted (Generalized) Linear Mixed-Effects (LME) Models, using `glmer()` and `lmer()` from the `lme4` package in R. These models can deal with unbalanced data. Each model embodied a 3 x 6 factorial design with threat level (low/medium/high) and potential loss (0–5 tokens), specified in the syntax of the lme4 R package as: *Dependent variable   1 + threat level * potential loss + (1 | participant).*

Step 2 – Replication of the 3-factor psychiatric symptom dimensions

To replicate a previously established latent transdiagnostic structure (Gillan et al., 2016; Hopkins et al., 2022; Rouault et al., 2018), we applied a factor analysis with Maximum Likelihood Estimation using the `fa()` function from the `Psych` package in R with an oblique rotation (`oblimin`). We selected the number of factors based on Cattell's criterion (Cattell, 1966) using the Cattell-Nelson-Gorsuch (CNG) test from the `nFactors` package. The participants' factor scores were estimated using the Thurstone method. Finally, to ensure that our extracted latent structure replicated previous findings, we compared the item loadings and participants' z-scored scores using the item weights between our current study and that of Hopkins et al. (2022) who had access to a substantially higher subject-to-variable ratio (N = 4782).

The CNG test revealed a 3-factor latent structure that was concordant with previous studies (Gillan et al., 2016; Hopkins et al., 2022; Rouault et al., 2018). As the loadings across items had large positive correlations between studies (cf. Figure B.4), we adopted the same labels as in the previous studies. For Factor 1 'Compulsive Behavior and Intrusive Thought (CIT)', the highest loadings came from the Alcoholism, OCD, Eating Disorders, Impulsivity, and Schizotypy questionnaires. Factor 2 'Anxious-Depression (AD)' was dominated by items from the Generalised Anxiety, Depression, and Apathy questionnaires. Lastly, Factor 3 'Social Withdrawal (SW)' had the highest average loadings from the Social Anxiety questionnaire, with some significant contributions from Generalised Anxiety and Eating Disorder questionnaires (cf. Figures 3.2,B.3, and B.5).

Exploratory inclusion of the three additional questionnaires assessing daringness, sensation seeking, and trait anxiety did not alter the 3-factor latent structure (cf. Appendix B.1.3 and Figure B.6).

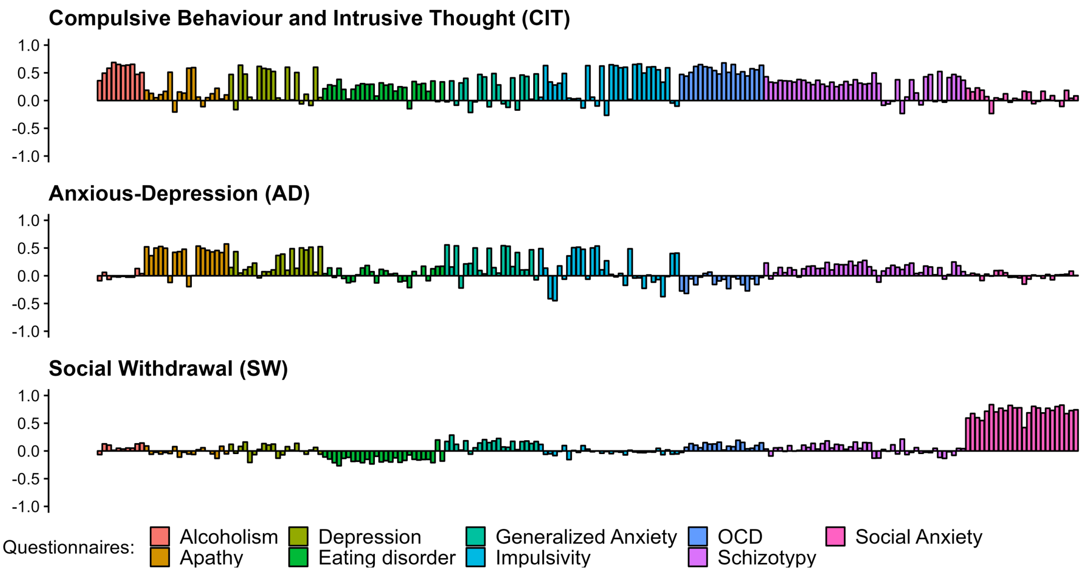

Figure 3.2: **Three latent transdiagnostic symptom dimensions explained the shared variance between all questionnaire items.** Loadings of all individual questionnaire items (color-coded by questionnaire) onto each factor. This plot is based on the combined sample, Figure B.5 shows the results separately for the discovery and confirmation samples. Cf. Figure B.3 for the eigenvalues and correlation matrices between questionnaires in both samples.

Step 3 – Effect of the psychiatric symptom dimensions and demographics on task readouts

To test the extent to which the interindividual differences predict behavior, we included each symptom dimension score or demographics as a z-scored fixed effect predictor into a simplified model from step 1 (i.e. without interaction of the within-subject factors). In the syntax of the `lme4` package, the specification was: *Dependent variable 1 + (threat level + potential loss) * IV + (1 | participant).* For all hypotheses relating to symptom dimensions, we ensured that the tests in the discovery sample remained significant in a model in which all demographics were added as covariates, and vice-versa for the hypotheses relating to demographics.

To test the extent to which the symptom dimensions predict recollection of threat memory (i.e. learned association between color and threat level), we included each score as a z-scored fixed effect predictor of estimated catch rates while accounting for true catch rates. In the syntax of the `lme4` package, the specification is *Estimated catch rates   1 + symptom dimension * actual catch rate + (1 | participant)*.

Estimating explained variance in symptom dimensions is complicated by the fact that mixed-effects models include within-subjects effects and trial-by-trial data. Thus, we estimated the proportion of explained variance in a regression model with symptom dimension as the dependent variable and six per-participant predictor variables as predictors: each participant's average approach rate and latency, and their respective linear relation to threat probability and magnitude.

To investigate a subjective prior assumption that the presence of tokens alerts the predator, we estimated the influence of each symptom dimension on the percentage of predator exposure attempts before the token appearance in a linear regression.

3.3 Results

We investigated the relationship between AAC behavior and three transdiagnostic symptom dimensions, namely Compulsive Behavior and Intrusive Thoughts (CIT), Anxious-Depression (AD), and Social Withdrawal (SW). Twenty hypotheses (denoted as H) were first generated based on a large discovery sample ($N_1 = 315$) and then confirmed after pre-registration in a second sample ($N_2 = 690$) with correction for multiple comparisons (cf. Table 3.1). To put our findings into context, we report effect sizes and inference statistics from post hoc exploratory analyses conducted on the combined sample.

We confirmed that in line with previous lab-based results (Abivardi et al., 2020; Bach, 2015, 2017; Bach et al., 2019; Castegnetti et al., 2020; Khemka et al., 2017), passive avoidance (i.e., proportion of avoidance decisions) and behavioral inhibition (i.e., approach latency) increased with threat probability and magnitude (cf. Figures 3.3.A and B.7, Table B.2). We also confirmed that the factorial structure of our questionnaire battery matched the one found in previous work (cf. Appendix B.1.3).

3.3.1 Approach-Avoidance Conflict Decisions

We found a strong and robust relation between AAC behavior and CIT (cf. Figures 3.3.B and B.8; Table 3.1). Individuals with high CIT scores exhibited a greater inclination to approach tokens (i.e., reduced passive avoidance, H1; $p < .0001$). Quantitatively, one standard deviation increase in CIT was associated with a 59.2% increase in approach rate, all other predictors being held equal. Secondly, individuals with high CIT scores were delayed in approaching tokens (i.e., higher behavioral inhibition, H8; $p < .0001$), such that one standard deviation increase in CIT corresponded to a 128 ms increase in approach latency. Finally, people with high CIT had a higher tendency to approach in the incorrect direction, i.e., the side opposite to the token (H20; $p < .0001$). Post-hoc analyses further indicated that while individuals with higher CIT levels were significantly more likely to be caught ($\beta = 3.2, t(997) = 6.71, p < .0001$), there was no significant association with reduced task performance as measured by total tokens won ($\beta = -0.72, t(997) = -0.99, p = .32$).

Next, AAC behavior in individuals with high CIT was less dependent on parametric threat features. The decrease in approach rate towards the token with higher threat probability (H2; $p < .0001$) and magnitude (H3; $p < .0001$) was less pronounced in people with high CIT. Similarly, the increase in approach latency with higher threat probability (H9; $p < .01$) and magnitude (H10; $p < .0001$) was less pronounced in people with high CIT. All in all, behavior explained 37.4% variance in CIT.

The next best predictor was IQ, as measured with the International Cognitive Ability Resource, with behavior explaining 8.0% variance (cf. Figure B.8). AAC decisions of higher IQ people depended more strongly on parametric threat features. The decrease in approaching tokens under higher threat probability (H5; $p < .0001$) and magnitude (H6; $p < .0001$) was more pronounced in people with higher IQ. Similarly, the increase in approach latency with higher threat magnitude was more pronounced in people with high IQ (H15; $p < .0001$).

Crucially, no strong relation between behavior and AD or sex was found, and those that were initially hypothesized were rejected in the confirmation sample (cf. Table 3.1). Indeed, no relation of any behavioral index with AD exceeded $R^2 = .01$ in the combined sample. Thus, in line with previous work, we found no evidence that AAC decisions related to subjective reports of transdiagnostic anxiety.

Notably, when analyzing individual questionnaire scores, we found that some of the aforementioned relationships were profoundly non-specific, i.e., extended across multiple questionnaires. In a post-hoc analysis across the combined sample, 11/12 questionnaires had a positive relation at $p < .05$ with approach choices (cf. Table B.3 and Figure 3.4). Among those, sensation-seeking and daringness had the strongest effect. In contrast, the remaining 1/12, social anxiety, had no effect. Similarly, 12/12 questionnaires had a significant positive relation at $p < .05$ with approach latency. Among these, OCD and daringness had the strongest effect. Nonetheless, the effect of CIT was larger than any of these individual questionnaires. CIT also accounted for the greatest proportion of behavioral variance at 37.4%, notably exceeding the variance explained by daringness and OCD, the two highest among the questionnaires, at 21.2% and 21.1% respectively (cf. Table B.3). Thus, while altered approach-avoidance behavior showed broad and non-specific relationships to most questionnaire scores, the patterns were better explained by a single transdiagnostic dimension than by any individual questionnaire.

3.3.2 Subjective Prior Assumptions

To investigate the subjective prior assumption that the presence of tokens alerts the predator, participants completed a predator exposure task. Here, optimal behavior according to the task statistics was to make an exposure attempt early in the trial, regardless of token appearance. Across participants in the combined sample, the majority of the exposure attempts (58.8%) were made after the token had appeared, confirming previous work in lab-based settings (cf. Figure 3.5.A). People with higher CIT made more exposure attempts before the token appeared than those with low CIT (H18; $p < .0001$). A one standard deviation increase in CIT score corresponded to a 161.33 ms reduction in exposure latency. This suggests that people with high CIT might have a less strong prior assumption that the presence of rewards alerts the predator.

3.3.3 Biased Threat Memory

After the two behavioral tasks, participants estimated the probability of getting caught by each predator. As expected, the estimated catch rates depended on threat level (ANOVA: $F(2, 1004) = 29.98, p < .0001$) and true catch rates (LME: $t(1, 2009) = 131.59, p < .0001$) across all participants. However, the relation between true and estimated catch rate was far from perfect (regression coefficient $\beta = .21$), and there was a significant intercept ($t(1, 2009) = 4417.79, p < .0001$): participants estimated catch rate on average 36.6% higher than the true catch rate (cf. Figure 3.5.B), which is in line with previous lab-based results (Bach et al., 2019).

CIT was linked to a biased learned association between color and threat level: people with high CIT globally overestimated catch rates more, even while accounting for true catch rates (H16; $p < .0001$). Additionally, the estimated catch rates of people with high CIT depended less on actual catch rates (H17; $p < .0001$). Post-hoc exploratory analysis suggested that the effect of the CIT scores was driven both by a higher memory bias ($\beta = .21, t(997) = 6.70, p < .0001$) and lower memory precision ($\beta = -.08, t(997) = -2.54, p < .01$) as indexed by the intercept and the slope relative to variations between estimated and actual catch rates. This analysis indicates that high CIT scorers were significantly impaired in their ability to differentiate between threat levels. To illustrate this, the top 25% of CIT scorers demonstrated a marked inability to accurately gauge threat based on catch rates (cf. Figure 3.5.C; cf. Appendix B.2.4).

Table 3.1: **Pre-registered hypotheses and the results of the confirmation analysis.** P-values from the discovery sample are not corrected for multiple comparisons and are presented as a heuristic guide only. For the confirmation sample, uncorrected P-values are presented; all of these were significant after comparing to adjusted alpha-rates according to the Holm-Bonferroni method.

Hypotheses	**Discovery Sample**	**Confirmation Sample**
H1. People with high CIT approach more often	$\beta = .693$ (.09), $t(1, 44823) = 58.92$, $p < .0001$	$\beta = .367$ (.07), $t(1, 98245) = 31.2$, $p < .0001$
H2. The decrease in approach with higher threat level is less pronounced in people with high CIT	$\beta = .267$ (.03), $t(1, 44823) = 102.23$, $p < .0001$	$\beta = .132$ (.02), $t(1, 98245) = 62.59$, $p < .0001$
H3. The decrease in approach with higher potential loss is less pronounced in people with high CIT	$\beta = .955$ (.04), $t(1, 44823) = 546.94$, $p < .0001$	$\beta = .528$ (.02), $t(1, 98245) = 463.76$, $p < .0001$
H4. The decrease in approach with higher potential loss is more pronounced in people with high AD	$\beta = -.111$ (.04), $t(1, 44823) = 9.64$, $p < .01$	Not confirmed
H5. The decrease in approach with higher threat level is more pronounced in people with high IQ	$\beta = -.247$ (.02), $t(1, 44966) = 112.33$, $p < .0001$	$\beta = -.081$ (.02), $t(1, 98819) = 21.21$, $p < .0001$
H6. The decrease in approach with higher potential loss is more pronounced in people with high IQ	$\beta = -.638$ (.04), $t(1, 44966) = 278.66$, $p < .0001$	$\beta = -.182$ (.03), $t(1, 98819) = 50.97$, $p < .0001$
H7. The decrease in approach with higher threat level is less pronounced in males	$\beta = .2$ (.05), $t(1, 44680) = 17.19$, $p < .0001$	Not confirmed
H8. People with high CIT approach later	$\beta = .087$ (.01), $t(1, 32215) = 35.54$, $p < .0001$	$\beta = .098$ (.01), $t(1, 79978) = 114.76$, $p < .0001$
H9. The increase in approach latency with higher threat level is less pronounced in people with high CIT	$\beta = -.014$ (.00), $t(1, 32215) = 26.18$, $p < .0001$	$\beta = -.005$ (.00), $t(1, 79978) = 10.13$, $p < .01$
H10. The increase in approach latency with higher potential loss is less pronounced in people with high CIT	$\beta = -.036$ (.00), $t(1, 32215) = 72.38$, $p < .0001$	$\beta = -.013$ (.00), $t(1, 79978) = 36.09$, $p < .0001$
H11. People with high AD approach earlier	$\beta = -.053$ (.02), $t(1, 32215) = 11.97$, $p < .001$	Not confirmed
H12. The increase in approach latency with higher potential loss is more pronounced in people with high AD	$\beta = .012$ (.00), $t(1, 32215) = 7.26$, $p < .01$	Not confirmed
H13. People with high IQ approach earlier	$\beta = -.065$ (.02), $t(1, 32335) = 18.73$, $p < .0001$	Not confirmed
H14. The increase in approach latency with higher threat level is more pronounced in people with high IQ	$\beta = .015$ (.00), $t(1, 32335) = 25.24$, $p < .0001$	Not confirmed

Table 3.2: Continuation of Table 3.1

Hypotheses	Discovery Sample	Confirmation Sample
H15. The increase in approach latency with higher potential loss is more pronounced in people with high IQ	$\beta = .023$ (.00), $t(1, 32335) = 25.35$, $p < .0001$	$\beta = .008$ (.00), $t(1, 80372) = 16.96$, $p < .0001$
H16. People with high CIT overestimate catch rates more	$\beta = 19.36$ (1.49), $t(1, 625) = 168.33$, $p < .0001$	$\beta = 11.11$ (.9), $t(1, 1367) = 150.76$, $p < .0001$
H17. For people with high CIT, estimated catch rates depend less on actual catch rates	$\beta = -.29$ (.04), $t(1, 625) = 61.9$, $p < .0001$	$\beta = -.15$ (.02), $t(1, 1367) = 51.43$, $p < .0001$
H18. People with high CIT make more predator exposure attempts before token appearance	$\beta = 5.66$, $t(1, 309) = 3.19$, $p < .01$	$\beta = 6.31$, $t(1, 645) = 4.9$, $p < .0001$
H19. People with high AD make fewer predator exposure attempts before token appearance	$\beta = -4.88$, $t(1, 309) = -2.73$, $p < .01$	Not confirmed
H20. People with high CIT more often approach in the incorrect direction	$\beta = -.426$ (.07), $t(1, 33892) = 32.5$, $p < .0001$	$\beta = -.231$ (.05), $t(1, 84335) = 25.09$, $p < .0001$

3.4 Discussion

Cautiousness in AAC tests is a stable behavioral trait (Bach et al., 2020) and is sensitive to anxiolytic drugs (Bach, 2021; Cryan and Sweeney, 2011) across species including humans (Bach et al., 2018; Korn et al., 2017) but with no strong evidence of a relation to self-reported anxiety (Bach, 2021; Bach et al., 2020; Fricke and Vogel, 2020). Here, we asked what are the clinically relevant personality traits that predict this behavioral trait. We included a large questionnaire battery, including a brief IQ test, as well as demographic information, in our analysis. In a discovery sample ($N_1 = 315$), we formed and pre-registered 20 hypotheses which were subsequently tested in a large independent sample ($N_2 = 690$).

Our primary finding highlights transdiagnostic compulsivity (i.e. CIT) as the main predictor of all behavioral readouts. High transdiagnostic compulsivity was related to decreased passive avoidance (i.e. lower avoidance rate) but also heightened behavioral inhibition (i.e. later approach). Interestingly, while participants with low levels of transdiagnostic compulsivity did not lack behavioral inhibition, those with high transdiagnostic compulsivity exhibited a heightened form. In addition, for those with high transdiagnostic compulsivity, both passive avoidance and enhanced behavioral inhibition were less dependent on trial-by-trial threat characteristics.

In a secondary task, we assessed implicit beliefs on threat-reward correlation as a potential mechanistic explanation for behavioral inhibition. We found that high transdiagnostic compulsivity was linked to a diminished belief that the presence of tokens alerts the predator, which indicates an altered or simplified representation of threat and reward relations. This links with previous studies suggesting that compulsive behavior may be explained by the difficulty in building an accurate explicit model of the world (Seow et al., 2020), which suggests they may rely on simplifications or shortcuts. While additional experiments are needed to determine exactly which mechanisms are at play behind the amplified behavioral inhibition, it seems unlikely that these processes operate under the same model-based control commonly observed in the rest of the population (Bach, 2015, 2017). Additionally, numerous studies have provided compelling evidence that individuals with compulsive disorders (Gillan et al., 2016, 2011; Morris et al., 2015; Voon et al., 2015), as well as healthy people with high CIT (Gillan et al., 2016), show reduced goal-directed control. Studies hint towards the idea that model-based learning deficits in compulsive individuals predict the presence of habits (Gillan et al., 2014). In the context of behavioral inhibition, a similar mechanism might be at play at least to the extent that it seems to rely on a basic association between cues (i.e. rewards) and response (i.e. avoidance) that does not adapt to environmental characteristics (i.e. threat probability and magnitude). Furthermore, our pre-registered findings revealed that higher IQ, associated with enhanced goal-directed control (Schad et al., 2014), was linked to higher integration of threat features into behavior, and post-hoc analysis revealed an increased belief in threat-reward correlations with high IQ (cf. Appendix B.2.3). These results, concordant with previous findings (Bach et al., 2020), suggest a form of attenuated behavioral inhibition, further endorsing the premise that variations in behavioral inhibition could stem from differences in goal-directed control abilities.

On the other hand, the decreased dependency on threat probability in people with high transdiagnostic compulsivity could be the result, at least partially, of an inability to discern between threats. The association between the color of the threat and threat level depended on behavior and had to be implicitly learned during the experiment. Given people with high transdiagnostic compulsivity approached more, they had more opportunity to learn, but in fact their estimated catch rates were less accurate. Interestingly, it has been posited that OCD and general compulsivity show deficits in the subjective reporting but not the learning of uncertain stimulus statistics (Lee et al., 2023; Vaghi et al., 2017) such that OCD patients can construct an accurate internal representation of the environment characteristics but fail to use it to guide behavior. In contrast, in the present study, people with high transdiagnostic compulsivity had deficits both in behavior and in explicit reports of threat statistics. Overestimation of threat has been implicated in the pathogenesis of OCD, mapping high on the CIT factor, due to biased processing of threat-related information (Moritz and Jelinek, 2009; Moritz and Pohl, 2009; Sookman and Pinard, 2002; Tolin et al., 2003).

Sex has been previously found to be the best predictor of cautiousness in a somewhat different risky foraging task in young people (Bach et al., 2020), in line with extensive self-report literature (Byrnes et al., 1999). In the current study, we did not observe any relation between cautiousness and sex but it was also not designed to investigate sex differences. In Bach et al. (2020), males were less cautious when their potential loss was smaller but adapted their behavior to the same level as females when the potential loss was higher (Bach et al., 2020). This may mean they are not more reckless overall, but adjust to risk more (Lewis et al., 2022).

Nonetheless, we replicated an intriguing finding from Bach et al. (2020), who identified daringness as the best self-reported predictor of behavioral cautiousness. Similarly, in our study, daringness strongly predicted cautious behavior. Although the associations were not unique, daringness consistently emerged as the best individual questionnaire predictor. However, its predictive strength was still outpaced by transdiagnostic compulsivity, which it notably maps onto.

Finally, we add to a large body of literature suggesting that cautiousness in

AAC tests, although these are often referred to as "anxiety tests," does not specifically relate to self-reported anxiety (Bach, 2021; Fricke and Vogel, 2020). Our study found no connections with transdiagnostic anxiety (AD). However, it is worth noting that the AD factor might relate more closely to apathy and depression rather than anxiety, as general anxiety scores demonstrated closer ties with apathy and depression than with another trait anxiety questionnaire (cf. Appendix B.1.3). Additionally, some connections between some (but not all) anxiety-related questionnaires and approach-avoidance biases were found. However, these links were not unique and exhibited the same directional patterns as other psychiatric questionnaires, some of which consistently explained more variance than anxiety. These findings are unlikely to be due to the specificity of the online sample, as they were equally, if not more, anxious than in-person studies (cf. Appendix B.2.1). In sum, the validity of using AAC tests for etiology research related to anxiety disorder in healthy humans may be questioned (Bach, 2021; Bach et al., 2020).

Beyond the specific findings of our study concerning behavioral inhibition and avoidance, our data underscore the advantages of employing transdiagnostic dimensions in contrast to traditional methods of phenotyping. Similar to Seow et al. (2020), we identified nonspecific patterns of correlation with task-related variables when we scrutinized commonly used yet infrequently compared clinical questionnaires. For instance, all twelve questionnaires were consistently associated with behavioral inhibition, yet only the compulsive factor distinctly aligned with this association and exhibited a larger effect than any individual questionnaire. This approach not only addresses collinearity concerns among questionnaires but also tackles their inherent heterogeneity. Moreover, our findings affirm the generalization of these symptom dimensions to another independent dataset, reinforcing their utility across various experimental designs for a comprehensive understanding of their impact on cognitive functions (Gillan et al., 2016; Katyal et al., 2023; Lee et al., 2023; Rouault et al., 2018; Seow et al., 2020, 2021).

An important consideration is whether CIT might in fact represent a general factor of psychopathology, often referred to as the p-factor (Watts et al., 2023). However, the strong correlation between our factor loadings and those identified in prior studies which delineated distinct effects associated with these factors, makes this association between CIT and the p-factor speculative and unlikely. Additionally, a second-order principal component, which represents a broad-spectrum psychopathology factor (Caspi et al., 2014; Lahey et al., 2021), did not explain behavior better than CIT, and several individual questionnaires (cf. Appendix B.1.4). Yet even if each three symptom dimensions represents distinct facets of psychopathology; higher CIT might still signal a propensity towards more severe psychopathological states, negative outcomes, and generally exert a more adverse impact on cognitive functions. This alternative explanation seems improbable under the premise of a linear relationship between psychopathology levels and approach-avoidance biases, given that the observed effects did not intensify in the confirmation sample, despite its consistently higher psychopathology scores compared to the discovery sample.

Another possible explanation for the results is that CIT is linked to less engagement in the task or that intrusive thoughts, for example, could cause momentary disengagement on some trials for high-CIT individuals. However, the combination of stringent exclusion criteria, comprehension and attention checks, performance-based incentives, and replication in a pre-registered sample suggest that the relationship between CIT and AAC task behavior is not simply due to inattentiveness or disengagement, both overall and on individual trials. Additionally, CIT was not associated with lower performance. Supporting this, recent work by Moutoussis et al. (2021) demonstrates that individual differences in decision-making—captured by a construct called "decision acuity"—are stable, meaningfully related to psychopathology, and underpinned by brain connectivity patterns. Their findings indicate that such cognitive differences are not merely artifacts of disengagement or inattention, but reflect genuine psychological and neurobiological variation. Together, these lines of evidence strengthen the case that the observed association between CIT and AAC task behavior reflects true individual differences rather than reduced task engagement.

While web-based data collection offers many benefits in psychiatry research (Crump et al., 2013; Gillan et al., 2016; Mason and Suri, 2012), concerns regarding the quality of data—particularly on mTurk—have been on the rise (Burnette et al., 2022a,b; Chandler et al., 2014; Zorowitz et al., 2023). Our exclusion rate, while somewhat higher than the typical 3-37 % range found in a meta-analysis, is not unusual for such studies (Shapiro et al., 2013). Interestingly, Zorowitz et al. (2023) demonstrated that inattentive responding can lead to spurious correlations. By addressing each possible source of spurious correlations through further tests and validations, we believe it is unlikely that our findings are the result of false-positive correlations (cf. Appendix B.2.2).

In conclusion, we find that a transdiagnostic symptom dimension assessing Compulsive Behavior and Intrusive Thoughts is the main predictor of cautious behavior in an AAC task. Even if AAC tests have been extensively used to characterize the effects of anxiolytic agents and probe neural circuitry related to anxiety (Bach, 2021), they might not specifically relate to self-reported anxiety.

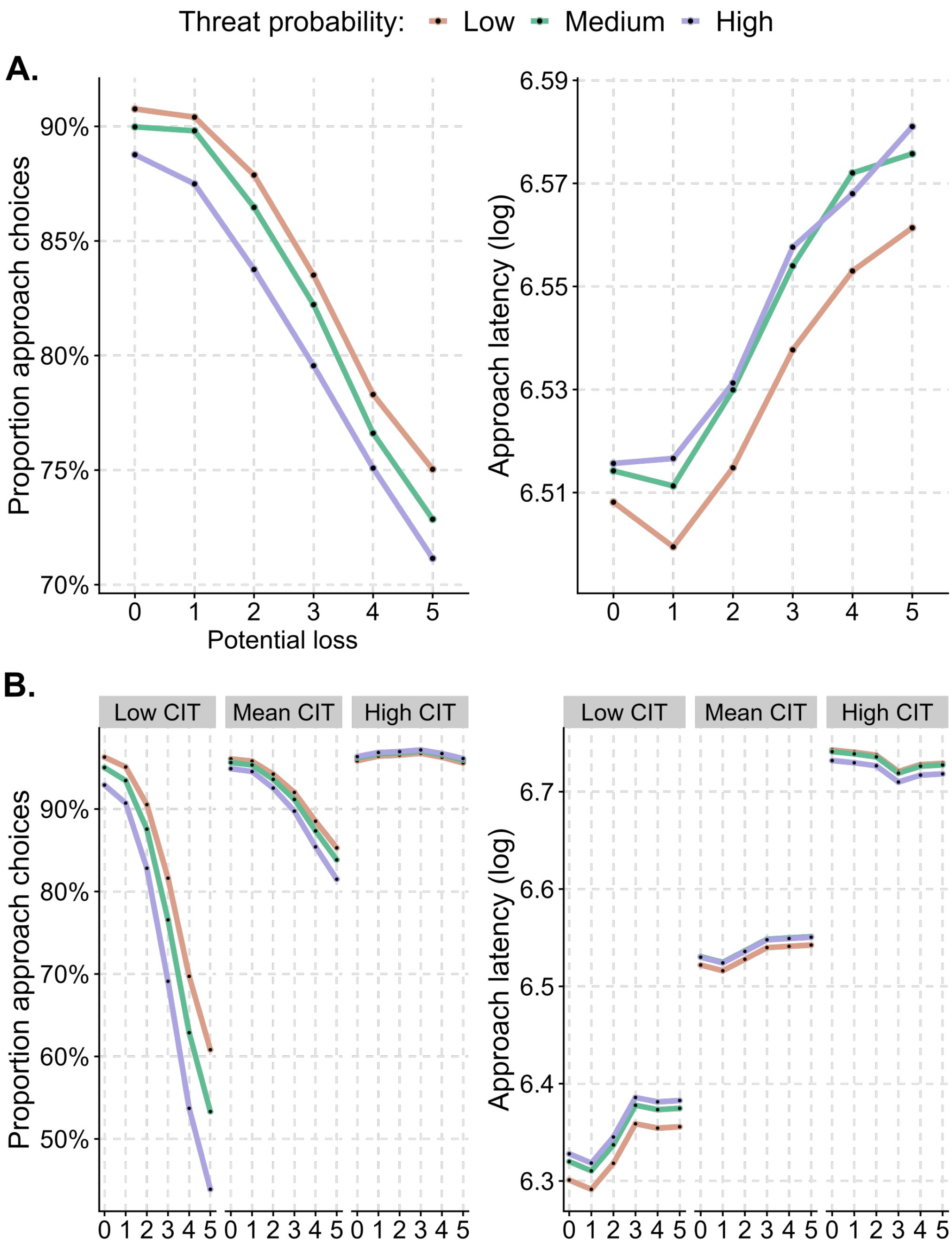

Figure 3.3: **Increasing threat probability and potential loss enhance passive avoidance and behavioral inhibition.** **(A)** Proportion of approach-avoidance decisions, indexing passive avoidance (left) and approach latency, indexing behavioral inhibition (right). **(B)** Estimated marginal means of approach choice (left) and latency (right) depending on CIT (Compulsive Behavior and Intrusive Thought) symptom dimension scores while other predictors are kept fixed. This plot is based on the combined sample, Figures B.7 and B.8 show the results separately for the discovery and confirmation samples. *Low CIT: -1.5, Mean CIT: 0, and High CIT: +1.5.*

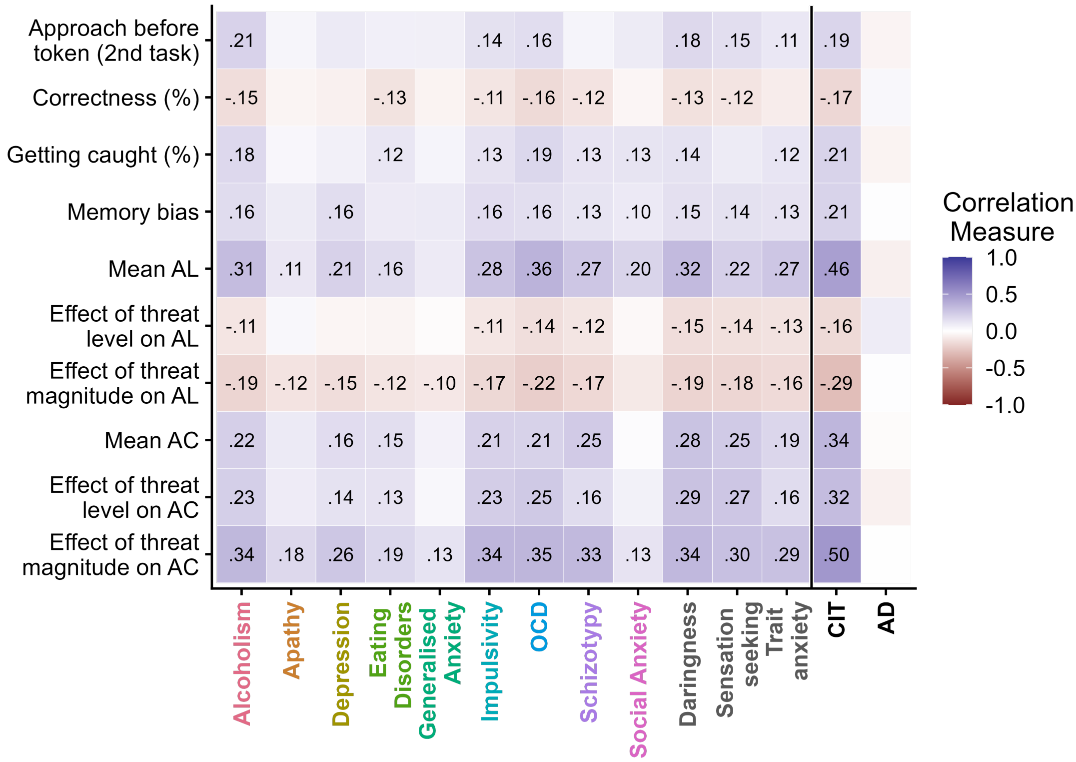

Figure 3.4: **Correlation matrix between behavioral indices and questionnaires scores** (left of black line) **and symptom dimensions scores** (right of black line). The color scale and the numbers indicate the correlation coefficient. The number is only present when the absolute r > .10. This plot is based on the combined sample. *AC: Approach choices, AL: approach latency.*

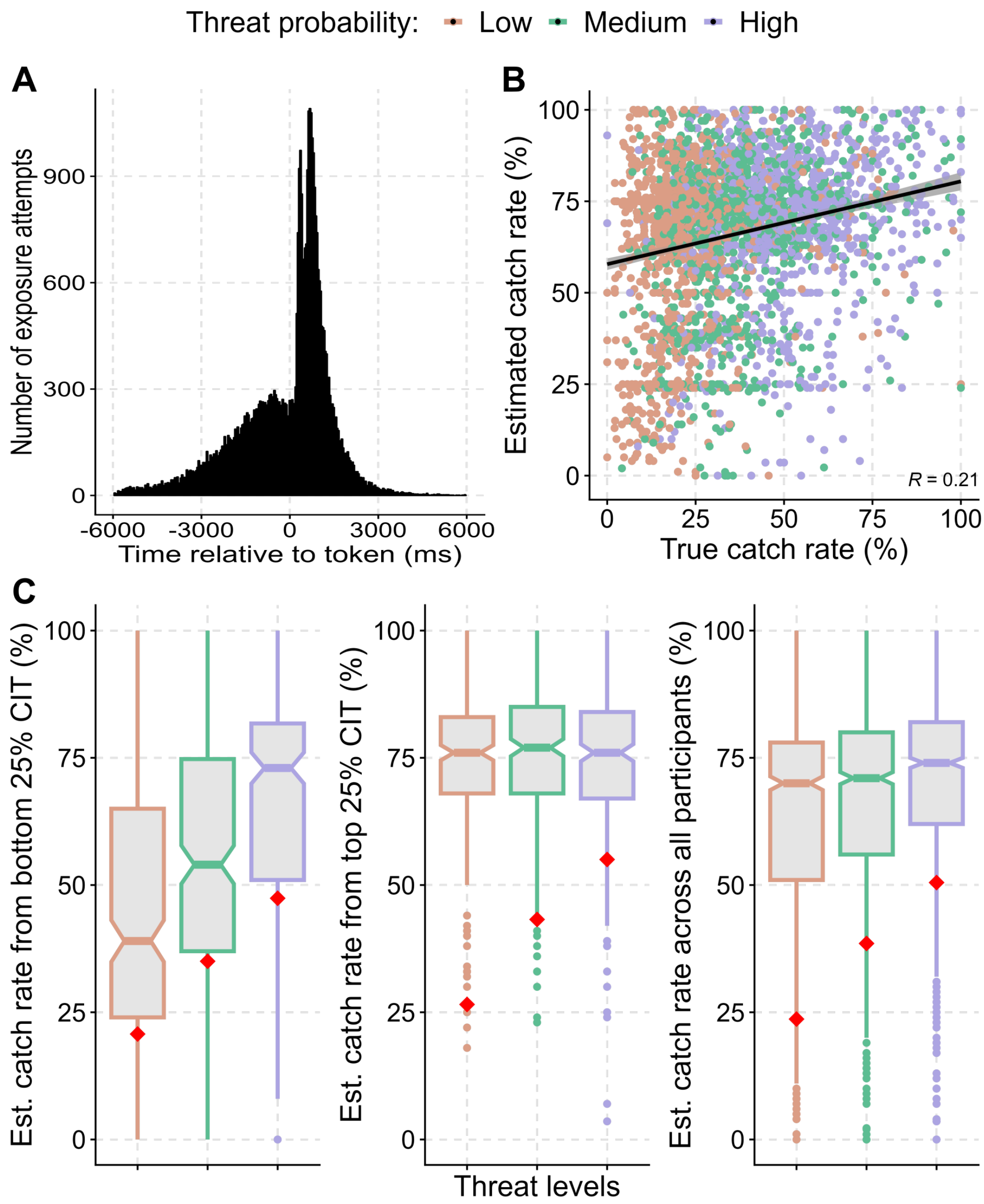

Figure 3.5: **Subjective prior assumptions and threat memory. (A)** Time of threat exposure attempts relative to token appearance. **(B)** Across participants, the estimated catch rates depended on the true catch rate which had to be learned during the experiment. **(C)** CIT (Compulsive Behavior and Intrusive Thought) is linked to biased threat memory such that the top 25% CIT scorers (center) did not distinguish between different threat levels and overestimated their probabilities. While the bottom 25% CIT scores (left) and all participants (right) distinguished the threats better, they still overestimated the threat probabilities. Actual threat rates for each level are denoted by red diamonds. This plot is based on the combined sample, Figures B.9 and B.10 show the results separately for the discovery and confirmation samples. *Est.: estimated.*

4

Escape Decisions

The content of this chapter has been published as a research article in Sporrer, J. K., Brookes, J., Hall, S., Zabbah, S., Hernandez, U. D. S., Bach, D. R. (2023). Functional sophistication in human escape. *Iscience*. doi.org/10.1016/j.isci.2023.108240.

Additional unpublished results have been incorporated into this chapter, offering a unified perspective on prior findings regarding individual differences.

4.1 Introduction

Field observations and laboratory experiments have revealed that many non-human species employ complex and sophisticated defensive behaviors (Evans et al., 2019, 2018), to escape from immediate threat. However, little is known about human escape actions and the computational mechanisms that control them, as this is difficult to study for ethical reasons. Previous work has used imagined threat scenarios (Blanchard et al., 2001), withdrawal from mild aversive stimuli or cues associated with them (Dinsmoor, 1977), or third-person view computer games (Bach, 2021; Fung et al., 2019; Mobbs et al., 2007; Qi et al., 2018; van Meurs et al., 2014) which restrict possible actions to key presses or joystick movements and tends to minimize behavioral dynamics. This mismatches with the rich and large repertoire of defensive actions commonly seen in real life and is likely to underestimate the complexity of the action space, and the ensuing decision problem (Bach, 2017). For example, Homer's Iliad anecdotally describes at least 13 distinct behavioral patterns under conspecific attack, not counting the use of weapons (cf. Appendix Table C.1). However, a systematic empirical assessment of the mechanisms that compute the choice between these behaviors in humans remains elusive (Bach, 2017; Mobbs et al., 2020; Mobbs and LeDoux, 2018).

Here, we investigated human escape behavior in a fully immersive virtual reality (VR) environment in which participants could move freely within a 5 x 10 m physical space (cf. Figure 4.1.A.). With "escape", we refer to behaviors aimed at distancing oneself from an existing threat in the environment, in order to reduce or eliminate harm. This is in contrast to "avoidance", which is often defined as behaviors aimed at preventing threat encounter in the first place (Sege et al., 2018).

In two independent experiments, N1 = 29 (E1) and N2 = 30 (E2) participants were instructed to forage for fruit on a bush, and to stay clear of various threats over 68/36 (E1/E2 part 1) independent trials. We included 13/7 natural and 3/1 artificial (control) threats (cf. Figures 4.1.B., C.1; Appendix Table C.2), which appeared on 60/32 of all trials (cf. Appendix C.1.3 for threat selection). A shelter, which was always 5 m behind the fruit bush, provided protection from all threats. On half of these trials, the threat moved toward the participant ("attack" condition; cf. Figure 4.1.C.) requiring escape, while on the other half, the threat diverted from the attack trajectory after covering 20% of the distance, making any escape unnecessary ("divert" condition; cf. Figure 4.1.D.). All threats were animated to show realistic behavior (URL to replay the experiments: https://osf.io/2b3k7/). Fast animal threats would chase and outrun the participants, forcing them to enter the shelter to survive. Slow animal threats would chase but not outrun the participants, thus requiring escape but not necessarily entry into shelter. The inanimate threat did not chase such that entering the shelter was unnecessary. Analyses in the main text refer to natural threats included across both experiments (cf. Figure 4.1.B., Appendix Tables C.2, C.3). In an exploration-confirmation strategy, hypotheses were generated by inspecting contrasts of estimated marginal means from linear mixed-effects models in E1, which were then tested in E2 with Holm-Bonferroni correction for multiple comparisons (cf. Figure 4.1.E.; Appendix Tables C.4, C.5).

4.2 Methods

4.2.1 Participant Details

To test eligibility, participants first had to complete an anonymous pre-screening questionnaire. To exclude any potential risk, they could not take part if they have a lifetime history of being victim to a life-threatening situation or interpersonal attack, or if they ever had major symptoms of post-traumatic stress-disorder (PTSD). This was assessed using the PTSD Checklist for DSM-5 (PCL-5; Weathers et al., 2013). Exclusion criteria also included the diagnosis of a mental or neurological disorder. To minimize the risk of injury from falling while moving in the virtual environment, they also needed to confirm that their movement abilities, hearing, and vision were unimpaired.

We report data from 29 participants (19 females; mean age $\pm$ SD $=$ 25.9 $\pm$

5.6) in the first experiment (E1) and 30 participants (15 females; mean age $\pm$ SD $= 24.7 \pm 5.3$) in the second experiment (E2). Due to hardware failures, some individual trials were not completed such that participants experienced, on average, 66.5 out of 68 and 35.8 out of 36 trials. One additional participant in E1 and three additional participants in E2 did not complete the experiment per protocol due to VR hardware failure and were not included. Since we were interested in escape behavior (which could only be initiated after the threat appeared), we excluded one additional participant in E2 who moved away from the fruit bush before any threat appeared on 31 out of 36 trials (corresponding to 5 standard deviations above group average).

All participants gave written informed consent before starting the experiment and received a fixed monetary compensation. Experiments complied with all relevant ethical regulations and were approved by the UCL Research Ethics Committee (6649/003).

4.2.2 Experimental Design

Each experiment consisted of a sequence of short encounters (trials) with various threats (cf. Appendix Tables C.2, C.3; Appendix C.1.3 for threat selection), and short breaks in between. Participants were tasked with collecting as many pieces of fruit (resembling kumquats) as possible while staying clear of physical contact with any threat.

Tutorial

An interactive tutorial took place in a white barren environment. First, we instructed participants to walk around the boundaries of the physical environment, run back and forth across the length of the space, and walk backward toward the edge of the physical environment. Secondly, we showed the participants an example of a fruit bush. Participants were taught to hold their hand over any appearing fruit for one second, which made it disappear and play a “string pluck” sound, before the next fruit appeared. The third stage allowed the participant to experience a red display and loud white noise that occurred on threat contact. Finally, participants were shown the shelter and told it would always appear behind their starting position. They entered the shelter to end the tutorial.

Trials

At the start of each trial, the participant was positioned in a low grass clearing surrounded by tall grass, with a single fruit bush 2.5 m in front of them and a shelter 2.5 m behind them. To reduce cue and context conditioning, the color and shape of the fruit bush were randomly varied, and additional bushes, grass patches, animal carcasses, and distant flocks of birds were added randomly around the participant without blocking any escape paths. Once they started fruit collection, a threat could emerge from the grass accompanied by an initial rustling sound after a delay which was uniformly drawn from 1-11 s. A trial could end in one of three ways: 1. Contact of any body part with a threat (approximated by a set of simple volumes), turning the display red, playing an uncomfortable high amplitude white noise sound, and removing any collected fruit in that trial while playing a chime sound. 2. Upon entering the shelter, the door was slammed shut (unless the threat was already in close proximity, thus blocking the door, leading to outcome 1). 3. After a predetermined time, if none of outcomes 1-2 occurred. After outcomes 2-3, a white display appeared, and collected fruit were added to a total count. In all cases, verbal feedback was given on a sign in front of the participant (e.g., "you were killed by the threat", "you escaped safely", or "you survived"). Afterwards, the participant was placed in a transition environment with floor markers, on which they had to stand in order to start the next trial.

4.2.3 Experimental Conditions

In E1 (68 trials) and the first part of E2 (36 trials), we implemented a 2 x 2 factorial design with "threat behavior" (attack/divert) and "time-to-impact" (1.5 s/5

s) as the two independent variables. Two of the 16 threats in E1 (crocodile and time bomb) were not compatible with divert behavior, and so were only used with the "attack" animation. For E1, this resulted in 60 trials, plus 8 no-threat trials visually identical to threat trials. For E2, this resulted in 32 trials, plus 4 no-threat trials.

Threat Behavior

Threats approached from a randomly selected angle (-45, 0, 45 degrees) and ran straight towards the fruit picking position. The attack angle did not affect the total distance traveled, as the threat path consisted of two approximately straight paths. In attack trials, the threat started chasing the participant after 75% of the time it took to reach the fruit picking position (calculated based on the threat's speed). In divert trials, the threat changed target to an invisible object on the far left or right of the participant after 20% of the time. Its target depended on the initial approach angle: left (-45 degrees) led to right diversion (+90 degrees), right (+45 degrees) led to left diversion (-90 degrees), and center (0 degrees) resulted in random left (-110 degrees) or right (110 degrees) diversion with equal probability. All angles were measured around the y-axis (up/down) relative to the positive z-axis (direction the participant would naturally be facing). The rock was set up in a different way. As a non-living object, we did not allow it to divert from its trajectory. Hence, in attack trials, it would roll towards the fruit picking position and beyond, in a ballistic manner. In divert trials, the initial direction of the rock was adjusted by +/-5 degrees. This meant the rock would not hit the participant, but the participant must still watch it for some time to estimate the precise trajectory.

Time-to-impact

We defined time-to-impact as the maximum time available to initiate escape and just about collide with the threat (i.e., if participants were minimally faster, they could escape; if they were minimally slower, they would get killed; cf. Appendix Figure C.1). Threats could be either faster or slower than the participant. For slow animals and the ballistic rock, this was simply the interval between the first occurrence of the threat and their arrival at the fruit-picking position. For chasing fast threats, this was the time that would lead to simultaneous arrival of the threat and the participant at the shelter, given assumed participant speed of 2 m/s based on pilot data. In other words, time-to-impact was the participant's lead time at the shelter if they could escape with zero delay. For all moving threats (excluding time bomb and crocodile in E1), time-to-impact was realized by adjusting initial threat distance and the placement of surrounding covering grass from which the threat appeared (cf. Table 4.1). Large threats appeared from tall grass placed around the participant. Smaller threats appeared from patches of shorter grass placed in the same manner. However, decoy tall grass was placed in small threat scenarios, and decoy short grass in large threat scenarios (at a randomized radius) so that the presence of either type of grass could not be used to predict the type of threat.

We used 1-dimensional constant velocity equations to derive the required initial position of the threat (S_T). In the first case, the threat is faster than the participant ($V_T > V_P$). Assuming constant velocity, the relationship between time, velocity, and position for the threat is given by:

$$S_{\text{esc}} = S_T + V_T \cdot T_{\text{esc}}$$

$$T_{\text{esc}} = \frac{S_{\text{esc}} - S_T}{V_T}$$

And for the participant by:

$$S_{\text{esc}} = S_P + V_P \cdot (T_{\text{esc}} - T_{\text{Plan}})$$

$$T_{\text{esc}} = \frac{S_{\text{esc}} - S_P}{V_P} + T_{\text{Plan}}$$

Combining these, we can solve for the initial distance of the threat (S_T):

$$S_{\text{esc}} = S_T + V_T \cdot \left(\frac{S_{\text{esc}} - S_P}{V_P} + T_{\text{Plan}} \right)$$

$$S_T = S_{\text{esc}} - V_T \cdot \left(\frac{S_{\text{esc}} - S_P}{V_P} + T_{\text{Plan}} \right)$$

Where S_{esc} is the safe house position, S_P is the expected position of the participant (fruit picking position), T_{esc} is the estimated escape time, and T_{Plan} is the time to plan and initiate their escape. S_{esc} and S_P were fixed at -2.5 m and 2.5 m respectively, as they were based on the size of the physical room, which gave a participant a 5 m distance to escape. V_P was assumed to be -2 m/s (running downwards) based on pilot data. The equation was used to a priori programmatically generate the scenarios which would make up each trial, for example placing the threat at its initial position and generating surrounding foliage that would obscure it.

The above method for threat initial placement would not work in the instances where the threat was slower than the participant ($V_T < V_P$) since the threat would not be able to ever catch them during their escape. In these cases, a different equation was used, whereby the participant was expected to get caught when the time-to-impact ended. If $V_T < V_P$, the threat can only catch the participant during the RT period so the movement time of the threat is equal to RT, and will catch the participant at S_P:

$$S_P = S_T + V_T \cdot T_{\text{Plan}}$$

The participant's speed (V_P) is 0 during the RT period, so is not needed. We simply solve the above for S_T. This allows for the calculation of the initial threat position (S_T) based on only the fruit picking position, speed of the threat, and prescribed time-to-impact:

$$S_T = S_P - V_T \cdot T_{\text{Plan}}$$

For the crocodile (in E1), a different environment with murky water was used. The crocodile moved perpendicularly along the bank as it approached to match the participant's position on the bank. Therefore, we positioned the threat such that the time it took to surface from the water and initiate the attack was equal to the time-to-impact for that trial.

The time bomb (in E1) appeared to be “thrown” from the tall grass and landed on the ground in front of the participant 1 second later. The time bomb displayed a timer which ticked down to an explosion that would kill the participant (regardless of distance) if they were not in the safe house. We set the time until detonation to match the sum of the time-to-impact for that trial and the estimated escape time (T_{esc}), starting from the moment it hit the ground.

Non-natural elements in the second part of E2

In the second part of E2 (blocks 2-4), we used non-natural elements in the environment, to address computational characteristics of action-selection mechanisms for one particular threat not used in block 1, namely the panther.

Force shield

In E2 block 2, we tested reinforcer devaluation in a 2 (force shield vs. no force shield) x 2 (panther vs. no threat) x 2 (1.5 vs. 5 s time-to-impact) design. In force shield trials, a visible and audible force shield built up and surrounded the participant at the beginning of the trial and then disappeared visually but continued to protect against the threat. Participants learned about the force shield in a preceding tutorial, during which the panther did not occur. To induce unpredictability and avoid participants from inferring the presence of the threat and the availability of the force shield, we determined the number of trials in this block randomly for each participant (cf. Appendix Table C.7). Thus, we implemented 3 types of trials: 1- certain panther trials that would occur for every participant, 2- uncertain panther trials that would occur with a probability of 75%, and 3- uncertain no-threat trials that would occur with a probability of 75%. Overall, in this block, there are 4 certain trials and 8 uncertain trials with 75% chance of occurring for each participant (cf. Appendix Table C.7).

Hands-up

In E2 block 3, we addressed whether participants could learn to use an instructed non-natural behavior (here, raising both hands above their head) to stop the threat from attacking. Participants were informed at the beginning of each trial whether panthers in this environment would be sensitive to the novel action by a sign in the grass clearing. The factorial design and number of trials were analogous to block 2.

Medusa

In E2 block 4, we sought to investigate whether participants could learn, by trial and error, to identify and avoid a common action (here, looking at the threat) when it led to a negative outcome (here, virtual death by a "magical force"). We used a 2 (panther vs. no threat) x 2 (1.5 vs. 5 s time-to-impact) design. This resulted in the following trials: 2 (time-to-impact) x 3 (repetition) certain panther trials, 2 (time-to-impact) uncertain panther trials, and 2 (time-to-impact) x 2 (repetition) = 4 uncertain no-threat trials (cf. Appendix Table C.7).

Exploratory trials

Block 5 comprised several exploratory trials related to the shelter presence and position. We included 2 trials where the participant started in a cul-de-sac gorge with no shelter, which made escape from the threat impossible.

4.2.4 Equipment

We used an HTC Vive Pro Eye HMD running an experiment built in the Unity Engine with SteamVR & Unity Experiment Framework (Brookes et al., 2019). Vive controllers were held in each hand, and Vive Trackers were attached to the waist and feet to allow for real-time body tracking. The VR headsets included a built-in microphone and eye tracking (cf. Appendix C.1.1 for more details about eye tracking).

4.2.5 Self-reported Questionnaires

We aimed to determine if heterogeneity between participants in threat-related behaviors arose due to individual differences, especially those that have an ethological origin. As such, we wanted to assess predictors of cautiousness such as fear susceptibility, phobia and anxiety levels, and risk preferences (cf. Appendix C.1.2 for more details about questionnaire selection). We implemented all questionnaires using REDCap electronic data capture tools hosted at University College London (Harris et al., 2009).

A few days before the experiment, participants completed self-report questionnaires assessing fear (Fear Survey Schedule-III, FSS; Wolpe and Lang, 1964), trait anxiety (State-Trait Inventory for Cognitive and Somatic Anxiety, STICSA-T; Gros et al., 2007; Ree et al., 2008, 2000), sensation seeking (Brief Sensation Seeking Scale, BSSS; Hoyle et al., 2002), disgust (Disgust Propensity and Sensitivity Scale, DPSS-12; Fergus and Valentiner, 2009), spider phobia (Spider Questionnaire-12, SPQ-12; Zsido et al., 2018), snake phobia (Snake Questionnaire-12, SNAQ-12; Zsido et al., 2018), motion sickness (Motion Sickness Susceptibility Questionnaire, MSSQ; Golding, 1998, 2006), video game usage (Video game usage questionnaire, VGUQ; Tolchinsky, 2013), as well as questions enquiring about their experience and expertise in martial arts.

Immediately before the VR game, participants provided demographic information including sex and gender, age, body weight and height as well as current health and physical status and completed the state portion of the STICSA (Gros et al., 2007; Ree et al., 2008, 2000).

Immediately after the VR game, participants assessed their cybersickness (Simulator Sickness Questionnaire, SSQ; Lane and Kennedy, 1988).

Cybersickness

Cybersickness is part of virtual reality-induced symptoms and effects (VRISE) and is considered to be a subtype of motion sickness induced by immersion into virtual reality (Davis et al., 2014; Saredakis et al., 2020). There is limited consensus about the symptoms provoked by VR as the biological mechanisms are unknown.

To ensure participants did not experience any negative symptoms from the VR, we measured their level of cybersickness (Davis et al., 2014). Compared to the mean cybersickness from similar VR interactive experiments combined in a recent meta-analysis (Saredakis et al., 2020), our sample reported lower cybersickness ($M_{\text{Saredakis et al., 2020}} = 34.3$, $M_{\text{E1}} = 24.2$, $M_{\text{E2}} = 26.6$). Additionally, for ease of comparison between studies, the VRSQ scores were $M_{\text{E1}} = 12.2$, $M_{\text{E2}} = 12.7$, the CSQ Dizziness scores were $M_{\text{E1}} = 0.5$, $M_{\text{E2}} = 0.7$, and the CSQ Difficulty focusing scores were $M_{\text{E1}} = 0.8$, $M_{\text{E2}} = 1$.

4.2.6 Quantification and Statistical Analysis

Task Measures

A priori, we extracted 18 trial-level summary statistics from our data (cf. Appendix Table C.4) for E1; a subset of these were used for statistical analysis of the first part of E2. When averaging over time periods, we use the trapezium rule to ensure the average is accurate over an irregular time sample. (1-3) The three possible outcomes, namely "escape to shelter" by going into the shelter, "survived" by not going into the shelter, or "virtual death" when getting into contact with the threat. We estimated "initiated escape" (4) and "escape initiation time" (5) as the time when participants first moved away from the fruit bush using the head tracker which had the highest data quality. When participants initiated their escape but did not go into the shelter, we considered this an "interrupted escape" (6). We extracted the smallest distance between the participant and the shelter (7) or the threat (8). We extracted peak (9) and mean speed (10) of the participant during escape. The following measures were extracted and averaged over the 1.5 s after the threat appeared (corresponding to the shorter time-to-impact), or during the entire escape (until entering the shelter, or trial end, whichever occurred earlier): (11-12) body orientation (mean cosine of angle between a vector pointing forward from the participant's pelvis, and the line between the participant and the threat), (13-14) head orientation (similar for a vector pointing forward from the participant's forehead), (15-16) fruit picking rate, and (17-18) visual scanning, defined as cumulative angle of head movements. For no-threat trials, the corresponding measures were taken from a random time point defined a priori within Unity. For trials in which participants did not escape, all measures relating to escape were considered missing.

For the second part of E2, we considered the following additional summary statistics: (19) fruit picking rate from threat appearance to the minimum duration of the trial (12.5 s); (20) virtual death by magical force. Furthermore, as we had eye tracker data available for E2, we used gaze orientation (rather than head orientation) to compute (17) for E2 part 2 only; results for head orientation were similar.

Behavioral Analysis

The numerous possible analysis methods combined with the high dimensionality of the dataset posed a formidable multiple comparison problem. Therefore, we opted for a rigorous exploration-confirmation approach. We generated 11 hypotheses by exploring E1 (cf. Appendix Table C.5), which we replicated in E2 while correcting for multiple comparisons (Holm-Bonferroni method). For statistical analysis, we used generalized linear mixed-effects models for binomial variables (virtual death, initiated escape, interrupted escape) and linear mixed-effects models for continuous variables (`glmer()` and `lmer()` from the `lme4` package in R). All models followed a 7 x 2 x 2 factorial design with threats (Elephant, Rock, Human, Bear, Dog, Snake, and Spider), time-to-impact (1.5 s and 5 s), and threat behavior (attack, divert); model syntax DV $\sim$ threat * time-to-impact * threat behavior + (1 | Subject). If models did not converge, we used the `lme4` function `allfit` to iterate through all alternative (g)lmer optimizers. For `glmer`, if the model still did not converge, we removed the integration over random effects (option `nAGQ`). To test pre-defined hypotheses, we estimated marginal means (EMMS) for each cell of the design using `emms()` and generated contrasts using `contrast()` (from the `emmeans` library in R).

For E2 part 2, the a priori primary outcome measures were "escape to shelter" and "minimum distance from shelter" in force shield and hands-up trials, and "virtual death by magical force" in Medusa trials. As secondary outcome measures for force shield trials, we considered "fruit picking rate from threat appearance to the minimum duration of the trial (12.5 s)" and "mean visual scanning over 0-1.5 s after threat appearance". For no-threat trials, the corresponding measures were taken from a random time point defined a priori within Unity. The resulting statistical analysis can be found in Appendix Table C.7.

Questionnaire Analysis

To determine which personal characteristics to include in a multiple regression, we selected questionnaire scores that fulfilled $r^2 > .10$ in bi-variate correlations with a behavioral outcome and retained up to three variables with the highest correlation. Additionally, sex was included in all multiple regressions as it is an important predictor of cautious behavior (Bach et al., 2020). We generated 4 hypotheses by exploring E1 (cf. Appendix Tables C.6 and 4.2), which we replicated in E2 while correcting for multiple comparisons (Holm-Bonferroni method).

Sound Analysis

For E1, a human rater classified, all microphone recordings during trials, whether they contained any voiced sounds and/or voiceless speech, or not. For trials with detected voiced sounds or voiceless speech, a second rater then categorized the time period from threat appearance to trial end, according to the following categories (several possible): (0) No sound in this interval, (1) non-differentiable such as vowels, growls, long hissing, (2) laughing, (3) shrieking, (4) speech apparently directed at the threat, (5) exclamation of surprise, including swearwords, (6) talking to oneself, (7) talking to the experimenter, (8) breathing sounds, (9) ambient sounds. Any occurrence of categories 1-6 was then summarized as "any vocalization".

For E2, we used an automated sound detection algorithm that counts the number of samples above a volume threshold and retains the recording if this number exceeds a time threshold. This algorithm was validated on manual classification as ground truth in E1. We performed a grid search over volume and time thresholds and retained the threshold tuple that resulted in at most 5% misses (within the training sample) and had the lowest number of false alarms. In a 5-fold Monte Carlo cross-validation with 1000 repetitions, this resulted on average in 6.2% misses and 28% false alarms on the test data set. We then used the thresholds optimized on the entire data set in E1, which resulted in a threshold pair of 700 a.u. (volume) and 0.06 s (time) (in-sample performance: 4.1% misses, 34.7% false alarms). Identified sounds were classified in the same way and by the same rater as in E1.

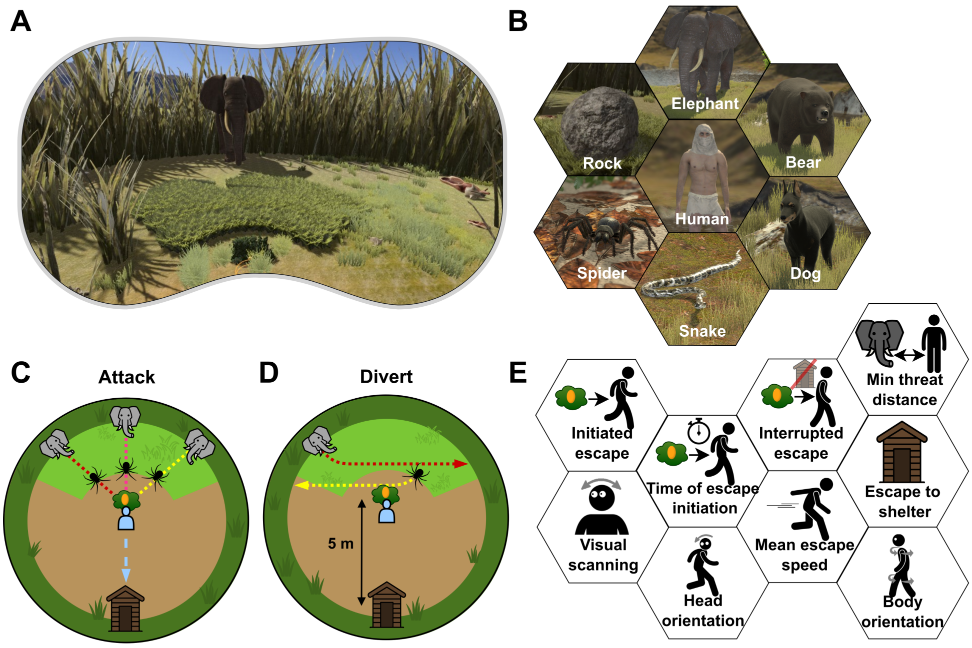

Figure 4.1: **Using fully immersive virtual reality to investigate escape behavior.** **(A)** Participant view (example trial). **(B)** Images of the 7 natural threats used in E1-E2. **(C)** Schematic view of the scenario setup. In the "attack" condition, threats moved toward the participant's position. Fast threats appeared from obscuring tall grass (dark green) and slow threats from short grass (light green) with dashed lines representing their possible trajectories. The escape path toward the shelter is shown in blue. **(D)** In the "divert" condition, the threat deviated after covering 20 % of the distance to the fruit bush. **(E)** Schematic representation of all behavioral outcomes whose hypotheses replicated.

4.3 Results

4.3.1 Escape Decisions

We define "virtual survival" when participants did not have any physical contact with a threat during a trial. When attacked, virtual survival occurred in 81.8%/75.4% of all trials and plateaued after around 3/5 trials (cf. Appendix Figure C.2.B.). Participants collected on average 9.1/9.0 fruits per trial (cf. Appendix Figure C.3.A.).

Beyond complying with explicit instructions to forage and survive, they also engaged in task-irrelevant behaviors that might be adaptive in natural environments. Alarm vocalization is a commonly used defensive behavior in the animal kingdom to deter or distract predators and warn conspecifics (Seyfarth and Cheney, 2003). Participants vocalized toward the threat (e.g., shrieks, squeaks) in 9.1%/18.8% of trials (cf. Appendix Figure C.3.B.) by 83%/73% of the participants. Post hoc analyses of the combined sample revealed that vocalizations occurred significantly more often during threat trials compared to no-threat trials ($t(58) = 5.16, p < .0001$). Additionally, they also showed avoidance behavior, such as seeking shelter in the absence of threat in 16.2%/7.1% of trials.

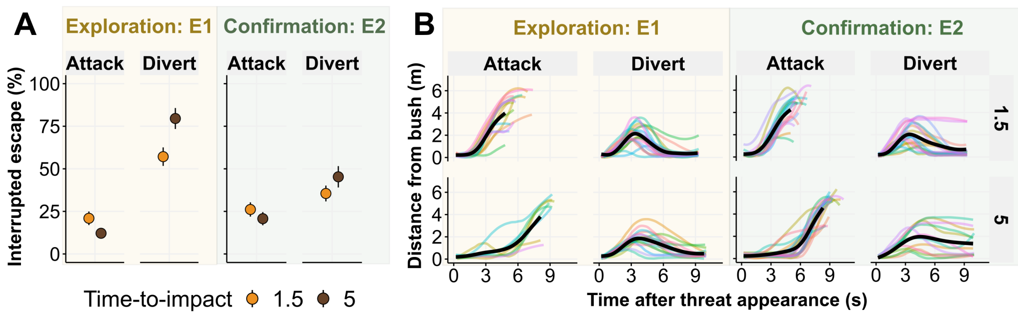

Figure 4.2: **Escape goals are dynamically updated according to environmental changes. (A)** Percentage of interrupted escape. **(B)** Distance of the participant from the fruit bush over time after threat appearance.
Each colored line represents a participant, and the black line is the overall mean. Orange: 1.5 s time-to-impact; brown: 5 s time-to-impact.

Table 4.1: **Statistical results of the eleven hypotheses generated by exploration of E1, and statistically tested in E2 part 1.** Parameter estimates and statistics are derived from linear mixed-effects models. P-values from E1 are not corrected for multiple comparisons and are presented as a heuristic guide only. For E2, uncorrected P-values are presented; all of these were significant after comparing to adjusted alpha-rates according to the Holm-Bonferroni method.

Name	Dependent variable	Contrast	E1 ($\beta \pm SE, t(df), p$)	E2 ($\beta \pm SE, t(df), p$)
H1	Interrupted escape	Divert vs. attack, fast threats	$\beta = -25.3 \pm 2.98, z = -8.49, p < .0001$	$\beta = -7.71 \pm 2.26, z = -3.41, p < .001$
H2	Escape initiation time	1.5 s vs. 5 s, attack	$\beta = -14.4 \pm 0.80, t(528) = -18.15, p < .0001$	$\beta = -12.7 \pm 0.81, t(617) = -15.59, p < .0001$
H6	Escape initiation time	Rock vs. all other threats, 5 s attack	$\beta = 1.60 \pm 0.23, t(528) = 7.01, p < .0001$	$\beta = 1.3 \pm 0.22, t(616) = 5.85, p < .0001$
H3	Initiated escape	1.5 s vs. 5 s, divert	$\beta = 10 \pm 1.71, z = 5.86, p < .0001$	$\beta = 10.7 \pm 1.96, z = 5.44, p < .0001$
H4	Escape to shelter	Feral (elephant, bear) vs. Familiar (human, dog), 1.5 s attack	$\beta = 2.65 \pm 1.09, z = 2.44, p < .05$	$\beta = 2.09 \pm 0.93, z = 2.24, p < .05$
H5	Escape to shelter	Rock vs. all threats, attack	$\beta = -4.26 \pm 0.71, z = -6.1, p < .0001$	$\beta = -2.66 \pm 0.65, z = -4.09, p < .0001$
H7	Mean escape speed	Fast vs. slow threats, attack	$\beta = 1.14 \pm 0.13, t(522) = 8.51, p < .0001$	$\beta = 1.17 \pm 0.14, t(609) = 8.63, p < .0001$
H10	Mean escape speed	1.5 s vs. 5 s, attack	$\beta = 2.48 \pm 0.40, t(522) = 6.25, p < .0001$	$\beta = 2.02 \pm 0.40, t(609) = 5.07, p < .0001$
H8	Body orientation during escape	Fast vs. slow threats, attack	$\beta = -0.88 \pm 0.12, t(522) = -7.60, p < .0001$	$\beta = -0.71 \pm 0.12, t(610) = -6.01, p < .0001$
H9	Head orientation during escape	1.5 s vs. 5 s, attack	$\beta = -1.28 \pm 0.39, t(522) = -3.27, p < .01$	$\beta = -0.85 \pm 0.38, t(609) = -2.25, p < .05$
H11	Visual scanning during 0-1.5 s after threat appears	Human vs. other fast threats	$\beta = -66.5 \pm 13.7, t(736) = -4.86, p < .0001$	$\beta = -69.0 \pm 15.2, t(768) = -4.54, p < .0001$

Table 4.2: **Statistical results of the four hypotheses relating to questionnaires, generated by exploration of E1, and statistically tested in E2 part 1.** Parameter estimates and statistics are derived from GLM models. P-values from E1 are not corrected for multiple comparisons and are presented as a heuristic guide only. For E2, uncorrected P-values are presented; all of these were significant after comparing to adjusted alpha-rates according to the Holm-Bonferroni method.

Name	Dependent variable	Predictors in model	E1 ($r^2, F(df1, df2), p$)	E2 ($r^2, F(df1, df2), p$)
Q-H1	Escape initiation time	Spider phobia (SPQ), fear (FSS), sex	$r^2 = .27$, $F(3, 28) = 3.13$, $p < .05$	$r^2 = .40$, $F(3, 28) = 5.66$, $p < .005$
Q-H2	Minimum distance from threat during escape	Spider phobia (SPQ), fear (FSS), sex	$r^2 = .27$, $F(3, 28) = 3.16$, $p < .05$	$r^2 = .46$, $F(3, 28) = 7.20$, $p < .005$
Q-H3	Fruit picking during 0-1.5 s after threat appears	Spider phobia (SPQ), fear (FSS), sex	$r^2 = .27$, $F(3, 28) = 3.15$, $p < .05$	N.s.
Q-H4	Head orientation during 0-1.5 s after threat appears	Sensation seeking (BSSS), spider phobia (SPQ), fear (FSS), sex	$r^2 = .36$, $F(4, 28) = 3.41$, $p < .05$	$r^2 = .34$, $F(4, 28) = 3.16$, $p < .05$

Escape is often conceptualized as "instinctive": triggered by specific features, with flexible motor implementation, but a predictable end goal (Evans et al., 2019; Gray and McNaughton, 2000). In our experiments, escape was initiated on 93.8%/91.7% of attack trials, and on 54.4%/70.8% of divert trials (cf. Appendix Figure C.4). Once escape was initiated, its ultimate target depended on the threat trajectory. For fast threats, participants did not reach the shelter on 14.2%/23.2% of initiated escapes when attacked. In contrast, when threat diverted, shelter was not reached in 63.9%/37.2% of initiated escapes (H1, $p < .001$; cf. Figures 4.2.A., C.5). Notably in attack trials, escape interruption appeared non-intentional, as it occurred later than during divert trials and invariably led to virtual death (cf. Figures 4.2.B., C.5). Importantly in divert trials, 27.4%/37.5% of interrupted escapes had already been initiated before the threat diverted and were thus presumably started with an intention to go to shelter. These results demonstrate that initiated escapes with an intention to go to shelter can be interrupted when the environment changes. This suggests that escape goals are dynamically updated during escape, rather than being predictable from the outset.

A commonly suggested escape trigger is "defensive distance" (Gray and McNaughton, 2000) or "predatory imminence" (Fanselow and Lester, 1988) of the threat, which is thought to heuristically integrate physical distance and type of threat. Here, we formalized this concept as time-to-impact of the threat, defined as the maximum time available to initiate an escape and just about collide with the threat during escape (i.e., if participants were minimally faster, they would survive). By adjusting physical threat distance and the surrounding covering grass, we implemented two time-to-impact conditions, 1.5 s and 5 s (cf. Appendix Figure C.2; Table 4.1). When threat attacked, participants initiated escape earlier during short time-to-impact (1.3 s/1.2 s) compared to long time-to-impact (2.9 s/2.5 s; H2, $p < .0001$; cf. Figure 4.3.A.). When threat diverted, participants initiated escape more often when time-to-impact was short vs. long (69.2%/82.6% vs. 39.8%/59.0%; H3, $p < .0001$). These results suggest that time-to-impact is an important factor for determining whether and when to initiate escape.

Beyond time-to-impact, biological and physical threat characteristics contributed to participants' escape decisions. For short time-to-impact, participants were more likely to enter the shelter when attacked by fast feral (i.e., elephant, bear; 61.6%/68.3%) than fast familiar threats (i.e., dog, human; 51.9%/56.9%; H4, $p < .05$; cf. Figures 4.3.B., C.6.A.). Participants entered the shelter less often when attacked by the (ballistic) rock than by any other threat (38.6%/40.7% vs. 76.0%/64.9%; H5, $p < .0001$) and initiated escape later than for any other threat when time-to-impact was long (4.8 s/4.2 s vs. 3.2 s/2.9 s; H6, $p < .0001$; cf. Figures 4.3.A., C.7.A.). These results suggest that the decision of whether and when to escape integrates the expected trajectory and behavior of the threat with its time-to-impact.

4.3.2 Escape Implementation

Next, we addressed the implementation of escape once initiated. When attacked, mean escape speed—and thus, energy expenditure (van der Walt and Wyndham, 1973)—depended on the threat's speed (slow threat: 1.4 $\mathrm{m\,s^{-1}}$/1.3 $\mathrm{m\,s^{-1}}$, fast threat: 1.9 $\mathrm{m\,s^{-1}}$/1.9 $\mathrm{m\,s^{-1}}$; H7, $p < .0001$; cf. Figures 4.3.C., C.8). Similarly, participants oriented their body more toward the threat during escape for slow vs. fast threats when attacked (averaged cosine of orientation angle away from threat, ranging from -1 (away from threat) to 1 (toward threat): 0.2/0.1 vs. -0.3/-0.3; H8, $p < .0001$; cf. Figures 4.3.D., C.9.A.). Body orientation during movement affects energy expenditure (Chaloupka et al., 1997) as well as how easy it is to observe the threat, and to return to foraging. Furthermore, head orientation during escape, and mean escape speed, depended on time-to-impact: participants oriented their head more toward the threat (0.0/0.0 vs. -0.1/-0.2; H9, $p < .05$; cf. Appendix Figure C.10.A.) and moved more slowly (1.5 $\mathrm{m\,s^{-1}}$/1.5 $\mathrm{m\,s^{-1}}$ vs. 1.9 $\mathrm{m\,s^{-1}}$/1.9 $\mathrm{m\,s^{-1}}$; H10, $p < .0001$) when they had more time. Finally, visual scanning during the first 1.5 s after threat appeared was less pronounced for a human threat, compared to any other fast threat (cumulative angle of frame-by-frame movements of the participant's forehead: 56.9 $°\mathrm{s}^{-1}$/71.3 $°\mathrm{s}^{-1}$ vs. 66.2 $°\mathrm{s}^{-1}$/79.9 $°\mathrm{s}^{-1}$; H11, $p < .0001$; cf. Appendix Figure C.11). Taken together, this suggests that escape implementation accounts for energy optimization and behavioral affordances (e.g., monitoring threat).

4.3.3 Effect of Individual Differences

Escape decisions were affected by stable personal characteristics (cf. Appendix Tables C.6, 4.2). Across all threats and conditions, fearfulness (Wolpe and Lang, 1964), fear of spiders (Zsido et al., 2018), and sex jointly explained 27%/40% of the between-person variance in escape initiation time (Q-H1, $p < .005$), and 27%/46% of the variance in minimum distance from threat during escape (Q-H2, $p < .005$): generally fearful females with high fear of spiders initiated escape earlier, and left more space between themselves and the threat.

Additionally, escape implementation also depended on stable personal characteristics (cf. Appendix Tables C.6, 4.2). Sensation seeking (Hoyle et al., 2002), fearfulness (Wolpe and Lang, 1964), fear of spiders (Zsido et al., 2018), and sex jointly explained 36%/34% of the between-person variance in head orientation when the threat appeared (Q-H4, $p. < 05$). High sensation seekers oriented more, and people with high fearfulness, fear of spiders, and with female sex, oriented less toward the threat.

Interestingly, we found spider phobia to be the strongest predictor of cautious behavior. In an exploratory post hoc analysis across the combined sample, spider phobia was significantly associated at $p < .05$ with more than 8 behavioral variables (cf. Table C.8). Specifically, individuals with high levels of spider phobia showed a distinct pattern of behavior: for example, they escaped more frequently ($r^2 = .17, p < .001$), were less likely to interrupt their escape ($r^2 = .08, p < .05$), stayed closer to the safe house ($r^2 = .16, p < .005$), maintained a greater distance from the threat ($r^2 = .15, p < .005$), and oriented their head ($r^2 = .18, p < .001$) and body away ($r^2 = .16, p < .005$) from the threat upon its appearance. This effect appears to be specific to spider phobia, as snake phobia was only significantly related to 3 behavioral variables, indicating a more limited influence on cautious behavior (cf. Table C.8).

See Appendix C.2.1 and C.2.2 for more details about replicating previous findings related to sex differences and anxiety, respectively (cf. Table C.8).

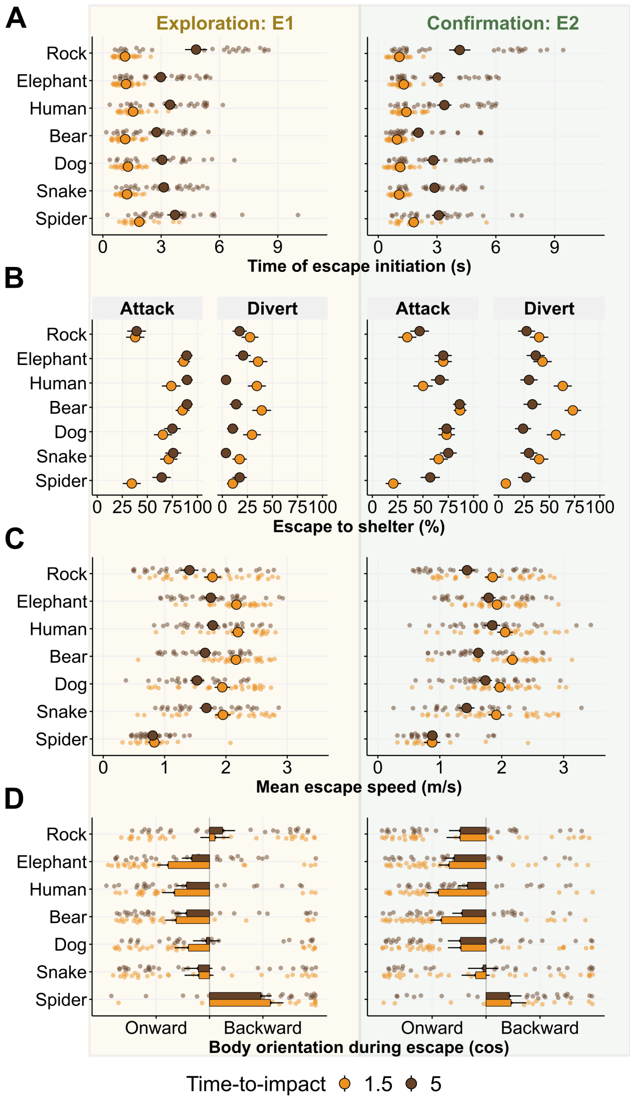

Figure 4.3: **The decisions whether, when, and how to escape depends on threat characteristics. (A)** Time of escape initiation. **(B)** Percentage of shelter entries. **(C)** Mean speed during escape. **(D)** Body orientation (mean cosine of orientation angle from threat; 1: toward; -1: away).
Large points with error bars represent mean and standard error across participants and trials, and small points represent individual trials. Threats are sorted by speed with the ballistic rock on top from elephant (fastest) to spider (slowest). Orange: 1.5 s time-to-impact; brown: 5 s time-to-impact.

4.3.4 Computational Characteristics of Escape Behavior

Overall, these results strengthen the view that escaping humans face a complex decision-making problem. Statistically optimal decision algorithms, such as model-based planning, are time- and memory-consuming (Bach, 2017). Thus, defensive behavior, especially at a short defensive distance, has been suggested to be controlled by mechanisms that are solely based on scalar action values (also termed model-free) (LeDoux and Daw, 2018; Mobbs et al., 2020). These action values might be malleable from experience, but could even be hard-wired. Such characteristics have a potential to improve decision speed but reduce flexibility to adapt to changing environments. Hence, in the second part of E2, we perturbed the environment in non-natural ways to reveal the computational characteristics of the underlying mechanisms through behavior (cf. Appendix Table C.7). All trials in this part employed a predatory threat that had not occurred in the first part of E2 (panther; cf. Figure 4.4.A.).

The first characteristic we investigated was sensitivity to reinforcer devaluation. This addresses whether an action is suppressed when the outcome (e.g., escape to shelter) is suddenly no longer desirable. It distinguishes, for example, certain model-based from classical model-free mechanisms (Dickinson and Balleine, 1995). On half of the trials in a "force shield" block (cf. Figure 4.4.B.), we devalued the shelter by endowing participants with a protective force shield that built up around the participant at the start of a trial and then visually disappeared. In a tutorial, participants learned from experience that objects could not penetrate the force shield even when it was invisible, meaning it would protect just like the shelter. Participants did not encounter the panther during this tutorial. A devaluation-sensitive mechanism would stop escaping immediately when encountering the panther within the force shield, whereas an insensitive mechanism might need repeated exposure to the panther to learn to suppress escape (Dickinson and Balleine, 1995). Hence, the following analyses only refer to the first trial of each condition (subsequent trials are visualized in Appendix Figure C.12). We found that participants rarely escaped to shelter when attacked within the force shield compared to without, independent of time-to-impact (3.3% vs. 80%; E2-H1, $p < .0001$; cf. Figure 4.4.C.; Appendix Table C.7). When the force shield was active, they stayed at a comparable distance from shelter regardless of threat presence (5 m vs. 5.1 m; E2-H2). This demonstrates that escape initiation is sensitive to reinforcer devaluation and likely to be under model-based control. On the other hand, some secondary defensive behaviors showed a different picture. When within the force shield and a threat appeared, fruit picking rate over the remaining trial was suppressed compared to a no-threat trial (0.8 $s^{-}1$ vs. 1.1 $s^{-}1$; E2-H3, $p < .0001$; cf. Figure 4.4.C.). In turn, visual scanning over 0-1.5 s after threat appearance was increased, compared to a no-threat trial (45.0 °$s^{-}1$ vs. 25.6 °$s^{-}1$; E2-H4, $p < .05$; cf. Figure 4.4.C.). This was also confirmed with gaze elevation, which was extracted using eye-tracking, representing another measurement to assess rapid information-seeking behavior (cf. Appendix Table C.7). This suggests that the mechanisms controlling rapid information-seeking and foraging-suppression are partly insensitive to reinforcer devaluation.

The second characteristic we addressed was the flexibility to update behavior from experience (sometimes termed "Pavlovian" vs. "instrumental"). Escape depends on threat identity and trajectory, and participants often looked at the threat, presumably for threat identification. In a "Medusa" block, participants were instructed that they had to identify and suppress a particular movement (head orientation toward the threat) that would activate a lethal magical force. They were not instructed what this movement was. Most participants learned to suppress this action, and thus reduced mortality from the magical force from 90.0% (trial 1) to 18.5% (trial 7; E2-H5, $p < .0001$; cf. Figure 4.4.D.). In a post-hoc interview, 60.0% of participants reported the correct force-activating movement. The remaining 40% still appeared successful in reducing their mortality (from 83.3% to 30.0%; cf. Appendix Figure C.13), possibly by altering related movements (e.g., body orientation) or by learning the correct movement but not forming an awareness of it. Overall, these results indicate that human threat identification behavior preceding escape is malleable by experience rather than hardwired.

Finally, we addressed the flexibility of the action repertoire. In a "hands up" block, participants were instructed that some panthers were trained to stop attacking when humans raised both hands high above their head. This action is novel and non-natural in this context as participants would normally use their hands to aid escape or protect their torso, rather than raise them above the head. The potential presence of these specific panthers was indicated by a sign at the start of each trial. Participants were so efficient in utilizing this novel action that even in the first trial for each condition, they never escaped to shelter when the new action was signaled (0.0% vs. 86.7%; E2-H6, $p < .0001$; cf. Figures 4.4.E., C.14 for subsequent trials). Furthermore, in two exploratory trials in the last block of E2, participants were attacked by a panther in a visually different scenario without shelter. Unexpectedly, 26.7% of participants spontaneously evoked the non-natural hands-up action they had previously learned, despite a lack of instruction to do so, or proof of effectiveness in this new context. Taken together, these results confirm again that escape initiation is sensitive to reinforcer devaluation, and demonstrate that the action-selection mechanism can easily integrate novel instructed actions and later retrieve them in novel situations.

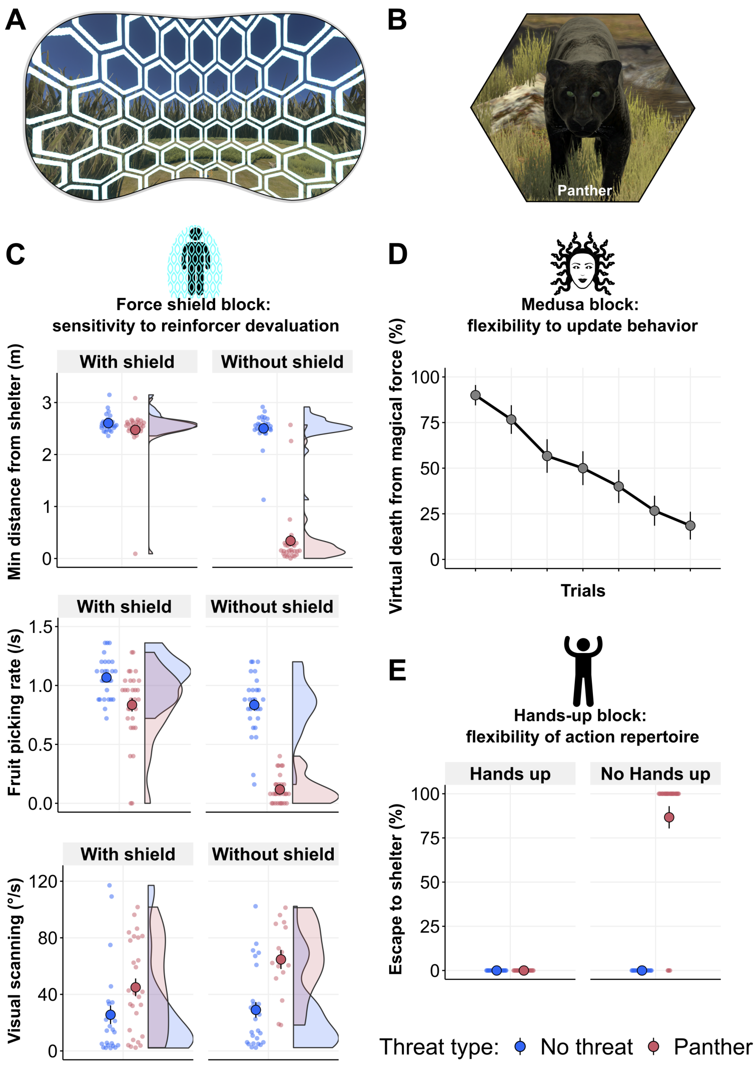

Figure 4.4: **Computational properties of the controllers underlying human escape behavior.** **(A)** Participant view of the force shield. **(B)** All trials in E2 part 2 used the panther as threat. **(C)** Minimum distance from the shelter (top), fruit picking rate over the entire trial after threat appearance (center), and visual scanning during 0-1.5 s after threat appearance (bottom) in the force shield block. **(D)** Percentage of virtual death from magical force over trials in the Medusa block. **(E)** Percentage of escape to shelter in the hands-up block.
Large points with error bars represent the mean and standard error across all participants and trials, and small points represent individual trials.

4.4 Discussion

While escape from immediate attack is sometimes depicted as a fixed reaction pattern ("fight or flight response"), the absence of observable and explicit deliberation does not imply a lack of dynamic and complex decision-making (Evans et al., 2019). We demonstrate that behavioral goals during human escape are dynamically updated, integrate time-to-impact with threat characteristics such as attack probability (feral vs. domestic) and expected trajectory (pursuit vs. ballistic interception), and are implemented in a manner that allows optimizing secondary goals (energy expenditure, behavioral affordances). Thus, in line with field observations in non-human animals (Evans et al., 2019), human escape responses are flexible and can integrate multiple variables such as spatial constraints of the environment and economic trade-offs even under strong timing constraints (Barrett and Finlay, 2018; Evans et al., 2019).

This flexibility poses a substantial challenge for computing accurate actions in limited time. Nevertheless, perturbance experiments show that escape decisions are controlled by a goal-directed mechanism that is sensitive to reinforcer devaluation in two separate tests. Furthermore, the action-selection mechanism can easily learn a novel non-natural action and integrate it into an enduring action repertoire. At the same time, general information-seeking and foraging-suppression behavior before escape appear insensitive to reinforcer devaluation and thus possibly rely on scalar action values, although threat-identification behavior can still be unlearned over time when leading to negative outcomes. Taken together, this suggests that the human brain might use different algorithms depending on the behavior required, or on the time constraints involved, which in our experiment were more pronounced for rapid information seeking than for the decision to escape. It is possible that with a time-to-impact shorter than the 1.5 s used here, humans would resort to simplified algorithms even for escape decisions.

Using third-person view and strategic computer games, we have previously shown that humans use goal-directed mechanisms to decide whether and when to approach a foraging opportunity under threat (Bach, 2017; Castegnetti et al., 2020), but that in doing so they may often employ approximate computations that rely on situation-specific heuristics (Korn and Bach, 2018, 2019). While we demonstrate that escape initiation depends on time-to-impact, we cannot yet quantitatively formalize the underlying computations and whether they represent or approximate statistical optimality. These are likely to involve additional and possibly threat-specific influences of physical distance, participant speed, foraging success, and the participants' actual escape speed which is variable across trials. One might argue that VR, although more realistic than third-person view computer games, is still insufficient to match the rich and large repertoire of defensive behaviors informally observed or imagined in real life. Notwithstanding, the robust elicitation of appropriate motor escape behaviors, the primary objective of the task, demonstrates the effectiveness of our experimental manipulations. Additionally, the use of task-irrelevant vocalization (e.g., shrieks, squeaks, screams) by the majority of the participants, and the occurrence of avoidance in the absence of threat, testifies to the realism and immersive nature of our VR environment. In turn, cybersickness in our experiments was lower than in typical VR experiments (Saredakis et al., 2020). Cybersickness, a constellation of bodily symptoms of discomfort such as eye strain, headache, and dizziness, driven by sensory integration processes (Kober et al., 2012; Weech et al., 2019), is negatively related to presence and immersion (Davis et al., 2014; Rebenitsch and Owen, 2016). Thus, low cybersickness further supports the notion of our VR environment being immersive.

Additionally, we found strong individual differences in escape decisions and their implementation. Across all threats and conditions, we found that generally fearful females (Wolpe and Lang, 1964) with a high fear of spiders (Zsido et al., 2018) initiated escape earlier, and left more space between themselves and the threat. Then the same personal characteristics on top of sensation seeking (Hoyle et al., 2002) jointly explained more than 30% of the between-person variance in head orientation where high sensation seekers oriented more, and people with high fearfulness and fear of spiders oriented less, toward the threat when it appeared.

As discussed in the general introduction (cf. Chapter 1 Section 1.4.3), a large-scale study using a 2D risky foraging task identified sex as the strongest predictor of performance among young people (Bach et al., 2020). Our findings revealed a similar pattern: males demonstrated higher performance than females (cf. Appendix C.2.1), likely due to maintaining a smaller distance from the threat and escaping later. Specifically, they were less cautious when they had more time but adjusted their behavior to match females when under time constraints. Notably, these cautiousness differences did not impact survival in either study. This supports the hypothesis that males are not inherently reckless but adjust their risk-taking to exploit lower-risk opportunities within high-risk environments. However, as these findings stem from post hoc analyses of the combined sample, they should be interpreted cautiously.

Spider phobia emerged as the strongest predictor of cautious behavior. Specifically, individuals with high spider phobia responded to threats by escaping more frequently with reduced flexibility, maintaining a greater distance, and orienting away from the threat. This is consistent with prior research, which identifies specific phobias as the only psychopathology reliably linked to avoidance biases in AAC tasks (Fricke and Vogel, 2020). Interestingly, our findings indicate that these effects are particularly strong for spider phobia compared to snake phobia, despite both threats being included in the task. The reasons for this difference remain uncertain. Although spider and snake phobias have similar lifetime prevalence rates (2% to 6%) and share evolutionary roots, some studies suggest that disgust may play a unique role in spider phobia (Cisler et al., 2009; Jong et al., 2002; Olatunji and Deacon, 2008). However, disgust sensitivity alone (Fergus and Valentiner, 2009) did not explain the variations in escape decisions observed in our study, implying that it may be the specific combination of fear and disgust that drives the heightened cautious behavior linked to spider phobia.

We found no general effect of self-reported anxiety (Ree et al., 2008), which aligns with expectations given the nature of our task. Since the task focused on circa-threat responses—situations in which an immediate threat is present — fear, rather than anxiety, appears to play a predominant role. Anxiety is typically associated with pre-encounter situations, where anticipatory responses are more relevant (Mobbs et al., 2020). However, the distinction between fear and anxiety is still contested (cf. Section 1.4.1 for more details).

In summary, our results provide an entry point for understanding how the healthy human brain computes and implements flexible and sophisticated escape decisions. This mechanistic and computational framework could offer a unique reference point for clinical research to investigate how these mechanisms might be altered in fear-related disorders (Bach, 2017; Yamamori and Robinson, 2023).

5

General discussion

Each experimental chapter included a discussion section that connected specific findings to existing research, addressing limitations and potential areas for improvement. In this concluding section, I will summarize each chapter's main points, examine their interconnections, and discuss broader limitations and future research directions.

5.1 Overview

In Chapter 1, I introduced the broad scope of risk-taking behavior, highlighting the complexity and unpredictability inherent in real-life decision-making and defining risk in this thesis as "first-order uncertainty"—the irreducible unpredictability in outcomes. This chapter also covered key aspects of risk preferences, exploring whether they are stable traits akin to personality or if they fluctuate dynamically in response to context. Additionally, I examined the possibility of both a general risk factor r and domain-specific influences. To investigate risk-taking, I introduced a range of assessment methods, from controlled experimental tasks to ecologically valid, naturalistic paradigms, including self-reported measures, and VR tasks. Each method, while insightful, has distinct strengths and limitations regarding precision and real-world applicability. Moreover, I discussed previous findings about individual differences that shape risk behavior, such as sensation-seeking, impulsivity, and sex differences, alongside the roles of anxiety and fear. Lastly, I presented computational frameworks, focusing on how theory-driven and data-driven approaches in computational psychiatry enable a more structured analysis of risk-taking behavior. These frameworks, specifically Prospect Theory and transdiagnostic psychiatric dimensions, provide a basis for understanding the computational underpinnings of risk preferences and individual differences. Through this multifaceted approach, the chapter sets the stage for a deeper exploration of risk-taking in the following chapters, examining how internal states and external factors converge to shape risk-related behavior.

In Chapter 2, I employed an economic gambling task alongside a computational prospect-theoretic model to assess patients' risk preferences, including risk and loss aversion, before and after a worry induction. Specifically, in-patients with MDD, with or without GA symptoms, and healthy controls had to choose between a certain monetary payoff and an uncertain gamble. There were no significant differences in risk and loss aversion between the three groups at baseline. After worry induction, patients with GA symptoms, but not the other groups, showed increased risk aversion. Crucially, these changes in decision-making were predominantly driven by anxiety rather than depression, as confirmed by psychiatric symptom scores. These findings suggest that decision-making disruptions in anxiety disorder may be driven by anxiety symptoms such as worry rather than causing them. This could shape etiological models, motivate standardization of emotional state in research on decision-making in anxiety disorders, and guide treatment strategies targeting reducing worry.

In Chapter 3, I explored which clinically relevant personality traits might influence cautious behavior using a cross-species validated AAC test, widely applied in pre-clinical anxiety disorder research. Participants chose whether and when to approach rewards under varying threat probability and magnitude. They also completed a psychiatric questionnaire battery with a known three-factor structure, consisting of Compulsive Behavior and Intrusive Thought (CIT), Anxious-Depression (AD), and Social-Withdrawal (SW). Using an exploration-confirmation approach across two large online samples, the results indicated that transdiagnostic compulsivity, rather than anxiety, is the strongest predictor of cautious behavior. Specifically, transdiagnostic compulsivity (i.e. CIT) was associated with decreased passive avoidance, increased behavioral inhibition, and reduced sensitivity to trial-by-trial threat features. High transdiagnostic compulsivity also implied an altered subjective model of threat and reward relations in the environment. Broad and unspecific associations were found between individual questionnaire scores and behavior, underscoring the value of transdiagnostic dimensions. Notably, there were no significant associations between behavior and transdiagnostic anxiety-depression (i.e. AD) or sex, and individual anxiety questionnaires were not among the best predictors of behavior. This study highlights that cautiousness in AAC tasks is comprised of two components, which are linked to transdiagnostic compulsivity in opposite ways, but not specifically or particularly strongly to self-reported trait anxiety. This finding challenges the traditional understanding of AAC tasks and provides a new view on cautious behavior in every-day situations.

In Chapter 4, I presented a comprehensive investigation of human escape responses to a broad range of biologically relevant threats using wireless 3D-VR in a large, immersive space. Thirteen naturally occurring threats, including predators, self-defending animals, and inanimate objects with realistic attack animations, were employed to observe how humans escape to shelter. Using a rigorous exploration-replication approach, the study uncovered a surprising blend of flexibility and rigidity in human escape decisions, offering new insights into the field. First, I found that human escape responses, even at close range, are dynamic rather than instinctive or hard-wired. Second, these decisions are based on a detailed assessment of threat identity and predicted behavior, rather than relying on broad features like a silhouette. Third, escape execution was found to integrate secondary behavioral goals. Fourth, I identified both external (threat-related) and internal (person-related) factors that influence escape decisions and their implementation. Key external factors included threat characteristics such as speed, attack probability, and expected trajectory. On the internal side, factors like spider phobia, general fearfulness, sensation seeking, and sex significantly impacted responses, while self-reported anxiety levels showed no link with general behavioral cautiousness. Finally, perturbation experiments suggested that distinct computational mechanisms underlie different types of threat-related behavior. Specifically, the decision-making algorithm underlying escape decisions demonstrated planning capabilities and the ability to integrate novel actions, while rapid information-seeking and foraging-suppression behaviors were only partially sensitive to devaluation. Taken together, this study provides steps toward a computational model of how the human brain rapidly solves survival challenges.

5.2 Implications for Cognitive Neuroscience

All three studies show that risk-taking behavior is modulated both by external and internal factors. In Chapter 2, internal factors include the induced worry state and the anxiety symptoms, while external factors include the framing of the economic lottery choice (cf. Appendix Results A.2). In Chapter 3, behavioral inhibition and passive avoidance were influenced by external factors such as threat magnitude and probability, and by internal factors like transdiagnostic compulsivity and other psychiatric traits. There was also an interaction between these factors. For instance, threat characteristics impacted behavior less in participants with high compulsivity. Finally, in Chapter 4, a number of behavioral readouts were influenced by external threat characteristics such as speed, attack probability, and trajectory, as well as internal factors like spider phobia, general fearfulness, sensation seeking, and sex.

The following sections will dive deeper into the findings, focusing on internal factors with an emphasis on individual differences, and integrate these insights for a more comprehensive understanding.

5.2.1 Anxiety

Avoidance of risky situations is widely regarded as a hallmark behavior of anxiety, yet empirical findings present a mixed picture. One explanation is that anxiety may impact risky decision-making through context-specific differences, or 'specific effects,' which only become evident under certain experimental conditions. When more general or 'broad effects' are observed, it becomes crucial to determine whether these effects are unique to anxiety or are influenced by overlap with other psychiatric traits, such as depression or compulsivity, which can similarly heighten caution. Anxiety's impact on risk-taking may also vary based on temporary emotional states rather than reflecting stable, trait-based tendencies.

In Chapter 2, anxiety increases risk aversion only in heightened emotional states in patients with high anxiety. This suggests that heightened risk aversion might arise during periods of acute stress or heightened emotionality but not in calmer states. Interestingly, this might be necessary to drive risk aversion in economic decisions, which typically lack a strong emotional context. This contrasts with risky foraging scenarios, which naturally evoke greater emotional arousal. However, it remains unclear whether a specific emotional state (e.g., worry) is necessary or if heightened general arousal suffices.

In Chapter 4, the impact of anxiety appears highly specific in foraging tasks: among more than 20 behavioral variables, only two were significantly associated with anxiety, yet these associations were aligned with previous work on a similar task (cf. AEP, Section 1.3.2). Specifically, both studies linked anxiety with earlier escape and greater escape success, particularly in response to distant threats. These findings suggest that trait anxiety plays a role in responses based on threat proximity, supporting theories that associate fear with temporally proximal or immediate threats and anxiety with temporally distal or further threats. However, it remains unclear whether anxiety is more closely related to the timing of the attack or the actual distance from the threat.

Another aspect often overlooked in theories of threat distance is the role of uncertainty in scenarios where threats are further away rather than imminent. In distant threats, the exact timing and trajectory of the potential attack might be more uncertain, creating a wider range of possible outcomes. Interestingly, escaping in the AEP is similar to cashing out in the BART (cf. tasks detailed in Section 1.3.2), and higher anxiety predicts less risk-taking in a high-uncertainty version of the BART, whereas anxiety is unrelated to risk-taking in low-uncertainty conditions (Smith et al., 2016). This is consistent with the evidence from Chapter 2 and other contexts where greater intolerance of uncertainty is reliably associated with increased anxiety (Dugas et al., 1997, 1998; Dugas and Ladouceur, 2000; Grupe and Nitschke, 2013; Mahoney and McEvoy, 2012; Mcevoy and Mahoney, 2011; Sandhu et al., 2023; Yook et al., 2010).

Nonetheless, in Chapter 3, some anxiety-related questionnaires correlated with approach-avoidance biases, but these connections were not specific to anxiety. Other psychiatric measures showed similar patterns and often explained more variance. This suggests that cautiousness in this category of tasks is not directly related to feelings of anxiety, or their representation in questionnaire measures. Another possibility is that (some) approach-avoidance tasks may not be as sensitive to distinguishing individual anxiety-related differences as previously assumed (Bach, 2021; Fricke and Vogel, 2020). Instead, they may highlight other important task characteristics.

5.2.2 Sensation-Seeking and Daringness

Notably, Chapter 3 also replicated a finding from Bach et al. (2020), who identified daringness as the best self-reported predictor of behavioral cautiousness in a similar AAC task with an open field component. Although the associations were not unique, daringness consistently emerged as the best individual questionnaire predictor. However, sensation-seeking, a trait related to daringness (cf. Section

1.4.2), was not a particularly important predictor of escape behaviors in the VR study in Chapter 4. In fact, it was only associated with a behavioral variable involved in the implementation of escape.

Whether daringness is truly distinct from sensation-seeking or simply a matter of terminology remains to be clarified. To the extent that they represent an overlapping concept, another possible reason for these differences could be the contrast in emphasis between reward and threat across studies. In the AAC tasks of Chapter 3 and Bach et al. (2020), daringness primarily reflects less cautious behavior in approaching rewards, while the VR study in Chapter 4 and the AEP in Fung et al. (2019) focus more on avoiding threats. Future research should examine whether these differences stem from a decoupling of reward-seeking and threat-avoidance (or rather, avoidance of threat-related uncertainty) or are influenced by other underlying factors.

5.2.3 Spider Phobia

Interestingly, spider phobia emerged as a strong predictor of cautious behavior in the VR study presented in Chapter 4. Individuals with high spider phobia responded to threats by escaping more frequently with reduced flexibility, maintaining a greater distance, and orienting away from the threat even for escape from non-spider threats. This aligns with previous research, including a systematic review identifying specific phobias as the only psychopathology reliably linked to avoidance biases in AAC tasks (Fricke and Vogel, 2020). In particular, our findings suggest that these effects are particularly pronounced for spider phobia compared to snake phobia, despite both threats being included in the task. The reasons for this difference remain uncertain. Although spider and snake phobias share evolutionary roots and exhibit similar lifetime prevalence rates (2% to 6%, Polák et al., 2020), some studies suggest that disgust may play a unique role in spider phobia (Cisler et al., 2009; Jong et al., 2002; Olatunji and Deacon, 2008).

5.2.4 Sex Differences

Sex differences in risk-taking are widely observed across species and are often attributed to evolutionary factors, with males engaging in higher-risk behaviors due to lower parental investment and greater reproductive variance. Human self-report studies largely align with this pattern, stating that males take more risks across various domains, including physical, financial, and everyday activities (Byrnes et al., 1999). Findings from behavioral risk-taking tasks are more nuanced on the other hand, suggesting that males may not be inherently more reckless but instead adjust their risk-taking more dynamically in low-risk situations. For instance, a large-scale study using a 2D risky foraging task identified sex as the strongest predictor of performance among young people, particularly in lower-risk conditions (Bach et al., 2020). Our findings in Chapter 4 revealed a similar pattern: males demonstrated higher performance than females (cf. Appendix C.2.1), likely due to maintaining a smaller distance from the threat and escaping later. Specifically, they were less cautious when they had more time, but they adjusted their behavior to match females under time constraints. Notably, these cautiousness differences did not impact survival in either study, supporting the hypothesis that males are not inherently reckless but instead exploit lower-risk opportunities within high-risk environments (Lewis et al., 2022).

However, we found no notable sex differences in Chapters 2 and 3. Regarding Chapter 2, it is possible that the sample size and statistical power were insufficient to detect subtle sex differences, particularly in a clinical population of anxious patients where other factors, such as symptom severity or treatment effects, might overshadow sex-related patterns. In Chapter 3, the lack of observed sex differences might be attributed to the experimental design, which relied on binary choices rather than continuous reward collection. This design may have limited the ability to capture sex-specific risk-taking strategies, such as dynamic adjustments in cautiousness.

5.2.5 Intermediate Conclusion

In conclusion, the findings in these studies highlight the complexity and specificity of the factors that influence risk-taking behavior, highlighting the interaction between internal characteristics, external contexts, and situational variables. Anxiety, while often linked to cautiousness, appears to exert its effects in context-specific ways, particularly under conditions of heightened emotional arousal or uncertainty. The differential roles of traits like daringness, sensation-seeking, and spider phobia further underscore the need to disentangle the contributions of overlapping psychological dimensions. In addition, the nuanced sex differences observed in some tasks point to adaptive strategies rather than generalized risk-taking tendencies.

5.3 Implications for Mental Health

Decision-making disruptions in mental disorders can be transient, shifting with the patient's emotional state. Such variability may partly explain inconsistent findings across studies, as decision-making may be influenced by factors like stress, mood, or anxiety at the time of assessment. This emphasizes the importance of distinguishing between state and trait influences in decision-making research. Adopting longitudinal or repeated-measure designs could deepen our understanding of how fluctuating internal states impact decision-making in clinical populations, offering a more nuanced, dynamic view of these processes over time.

A transdiagnostic approach also proves valuable in understanding risk and defensive behaviors, as single-diagnosis frameworks may overlook the unique ways in which overlapping symptom dimensions, such as compulsivity and anxiety, shape behavior. Furthermore, using multiple psychiatric questionnaires when linking symptoms with behaviors is crucial to avoid attributing nonspecific effects to specific traits. Relying on a single measure risks mistaking general associations for specific ones. Common correlations between task-related variables and psychiatric traits may reflect the limitations of frequently used yet rarely compared clinical questionnaires, underscoring the need for broader, multi-measure assessments to reveal a clearer picture of the connections between psychiatric symptoms and behavior.

The strong association between spider phobia and avoidance behaviors underscores the potential to enhance phobia-specific therapies, particularly exposure therapy, by addressing the unique interaction of fear and disgust in shaping threat responses. Spider phobia can still be elicited by virtual, rather than real, threats, highlighting the viability of using VR to replicate phobic triggers effectively. This reinforces the potential for VR-based exposure therapy, which can replicate realistic and dynamic threat scenarios in a safe, controlled, and immersive environment. By practicing adaptive responses in these scenarios, patients can confront their phobia, build resilience, and develop coping strategies that can be transferred to real-world contexts, potentially improving treatment efficacy and long-term outcomes.

5.4 Implications for Society

Interestingly, risk-taking is often praised in society, with those who take greater risks generally enjoying higher status. This may be because risk-taking tends to be more advantageous at the societal level than at the individual level. On a societal scale, risk-taking is often necessary for innovation and progress as an individual's success can propel collective advancement through new inventions or ideas that benefit the entire community. Moreover, the loss of an individual might be seen as inconsequential—or even beneficial—if their resources are redistributed in ways that support others.

From the individual's perspective, the potential negative consequences of risk-taking, such as bankruptcy or even death, are far more significant and personally devastating. This disparity raises an important ethical question: if society benefits from individuals taking risks, does it have a responsibility to support those who fail? Gaining a deeper understanding of the drivers behind risk-taking, as well as identifying circumstances or individuals where risk-taking becomes maladaptive or disruptive—whether through extreme risk aversion or excessive risk-seeking—could serve as a primary safety net. Such insights could help mitigate the negative impacts of risk-taking while still fostering societal progress.

5.5 Limitations

5.5.1 Limitations Related to the Samples

A primary limitation across all studies concerns the characteristics of the sample used, which may impact the generalizability of the findings.

In Chapter 2, the relatively small sample size may have limited the statistical power to detect more nuanced effects between the patient groups at baseline. Although the main effects were robust, subtle contrasts could have been missed. In addition, the focus of this chapter is on a specific clinical population, psychiatric inpatients with severe depression and comorbid anxiety, which may not fully represent the broader spectrum of individuals with milder symptoms or those in outpatient care. Thus, the findings may not generalize to populations experiencing different forms or severities of these conditions.

In Chapter 3, I relied on online recruitment, which introduces additional questions about sample representativeness, particularly compared to in-person samples. Online samples often differ in demographic and psychological characteristics (Gillan and Daw, 2016), which can influence the generalizability of the results to clinical and community populations that might respond differently in controlled laboratory settings. More studies with larger, more diverse, and clinically varied samples, including those recruited in person, would help validate and extend these findings.

The sample in Chapter 4 consisted primarily of university students with characteristics typical of WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations (Henrich et al., 2010). Furthermore, for ethical considerations, pre-screening was quite extensive, preventing anyone who experienced significant adverse events from participating. This may limit the extent to which our findings apply to individuals from different socio-demographic or cultural backgrounds. This may be especially important in risk-taking research, as cross-cultural studies have documented meaningful differences between populations (L'Haridon and Vieider, 2019; Rieger et al., 2015; Vieider et al., 2012). Therefore, future research would benefit from more inclusive sampling strategies that encompass a broader range of socio-cultural backgrounds to provide a fuller picture of how risk-taking behaviors manifest globally.

Despite these sample-specific limitations, this thesis benefits from the diversity of samples across studies. This diversity allows us to explore a broad spectrum of risk-related and defensive behaviors and offers a unique perspective by integrating findings from clinically affected individuals, online respondents, and healthy students. While each sample type has its limitations, this multi-sample approach strengthens the thesis by providing complementary insights that together enhance our understanding of how internal and external factors shape risk-taking across different populations and contexts.

5.5.2 Limitations Related to the Experimental Designs

As each study examines distinct aspects of risk-taking and defensive behavior—economic risk in Chapter 2, approach-avoidance conflict in Chapter 3, and escape behavior in Chapter 4—there is limited opportunity to directly compare findings across these domains. This restricts the ability to generalize conclusions regarding risk aversion, cautiousness, and escape tendencies across different contexts. Future studies could address this by integrating multiple risk domains within a unified experimental framework, which would allow for the assessment of whether internal (e.g., personality traits, anxiety) and external (e.g., threat type, framing) factors exert consistent influences across varied types of risk-taking behaviors. Additionally, longitudinal or repeated-measures designs could help clarify whether these behaviors are stable across contexts or vary based on situational demands.

Another limitation pertains to the choice of questionnaires in individual differences research. Our findings are only as robust as the questionnaires' ability to accurately measure the targeted constructs, particularly in the realm of anxiety research, where there is no consensus on the most appropriate measure. This issue is highlighted in Chapter 3, which underscores the risks of drawing broad conclusions about anxiety based on a single type of questionnaire. The results can vary significantly depending on the chosen questionnaire, raising questions about whether the observed effects genuinely reflect anxiety-related constructs or are influenced by confounding factors.

For example, in Chapter 3, the STAI-Trait and GAD-7 items mainly loaded onto an anxious-depression factor, while the STICSA-Trait items showed stronger loading on a compulsivity factor (cf. Appendix B.1.3 for hypotheses on these factor-loading differences). In addition, LSAS items that assess social anxiety are loaded onto a separate social withdrawal factor. These differences indicate that the findings derived from STICSA may differ considerably from those obtained with STAI or GAD-7. To mitigate these issues, I included multiple psychiatric measures in Chapters 2 and 3 to capture anxiety and other psychiatric traits more comprehensively. However, including multiple measures is not always feasible, especially in studies where individual differences are not the primary focus, as it requires substantial time and resources. In Chapter 4, for example, I had to limit the approach to a single anxiety measure due to practical constraints, although I included several other psychiatric questionnaires to capture a broader range of individual differences. This allowed some insight into psychiatric traits beyond anxiety, while keeping the assessment time manageable.

Moreover, findings in Chapter 3 highlight the advantages of using transdiagnostic dimensions over traditional phenotyping, as this approach reduces the dependence on any one questionnaire and allows for a broader assessment of psychiatric traits. However, transdiagnostic approaches require a high subject-to-variable ratio to reliably capture dimensional constructs, making them difficult to implement in studies with limited sample sizes. This issue is particularly relevant in clinical research, where the recruitment of large homogeneous samples can be challenging. Future studies could explore adaptive sampling methods or machine-learning techniques to optimize the accuracy of transdiagnostic measures in smaller samples.

5.5.3 Limitations Related to Computational Psychiatry

While computational psychiatry has shown and continues to show, great promise in bridging the gap between neurobiology, computational neuroscience, and clinical psychiatry, enhancing our understanding of the (ab)normal human behavior, it is not without its limitations. One of the core issues is that the limitations inherent in the measurements themselves persist. The quality of results is constrained by the quality of the methods; therefore, computational approaches cannot compensate for poorly designed experiments.

Computational psychiatry also carries a high risk of drawing unjustified conclusions for several reasons. In theory-driven approaches, presenting a single well-fitting model is not sufficient. Models need to be compared against alternatives, and simulations are necessary to falsify each model properly (Palminteri et al., 2017). Moreover, many studies report that the reliability of parameter estimates is often poor to moderate; however, various methods can be employed to improve this (cf. reviews Karvelis et al., 2023; Zorowitz et al., 2023).

Data-driven computational methods also face significant challenges due to the high dimensionality of data and the multitude of analytical techniques available. These conditions can lead to problems such as multiple comparisons and overfitting, which are not always adequately controlled. When the number of predictors exceeds the number of available data points, it becomes easy to create models that produce seemingly perfect predictions, which can often be misleading (Huys et al., 2016; Rutledge et al., 2019).

Furthermore, computational models relevant to psychiatry should incorporate neurobiological constraints, which makes them subject to the limitations of current empirical neuroscience. While efforts have been made to model even the most subjective human experiences, such as happiness (Rutledge et al., 2014), computational models still struggle to comprehend qualia—defined as the phenomenal properties of experience, such as the "fearfulness" of fear. As a result, some computational frameworks deliberately avoid addressing qualia altogether (Bach and Dayan, 2017).

Ultimately, the most significant challenge may be the complexity of real-world behavior and its interactions within a highly intricate and interconnected environment. Our current mathematical tools might fall short, and while we anticipate breakthroughs in computational research, the full scope of these complexities could remain elusive, perhaps even unsolvable.

5.6 Future Directions

A significant portion of this work has focused on examining between-person differences in behavior, which overlooks important within-person variations. Understanding within-person dynamics could provide insight into how individuals modulate their risk and defensive behaviors in response to shifting internal states (e.g. stress, sensation-seeking, Lydon-Staley et al., 2020) and external contexts (e.g., increasing threat levels). This approach would allow future studies to capture more nuanced, situationally responsive behaviors that may not be apparent when only comparing individuals.

Another area not addressed in this thesis—but with a well-established foundation in classical defense theories—is the role of physiological arousal. In several theoretical frameworks, such as the defense cascade model, physiological arousal is considered a primary trigger for activating defensive behaviors (Kozlowska et al., 2015). The integration of psychophysiology and neuroscience has enabled translational insights from animal studies into human research, enhancing our understanding of the neurobiological mechanisms behind threat responses. However, while physiological responses to threat have been extensively studied in tightly controlled laboratory conditions, there is a notable gap in connecting these physiological measures to more naturalistic defensive behaviors, as seen in real-world or VR environments. Incorporating physiological data into studies with dynamic and ecologically valid scenarios could enrich our understanding of threat responses and highlight unique response patterns or representations that are obscured in artificial contexts.

Moreover, VR provides an opportunity to explore these responses in an immersive and realistic context, which may reveal how physiological reactions interact with environmental cues in ways that are distinct from reactions to abstract or static threats. Integrating physiological measures, such as heart rate, galvanic skin response, or eye tracking, into VR-based studies on escape and defensive behavior could offer deeper insights into how arousal shapes rapid decision-making in realistic settings. This approach could also clarify whether certain physiological markers consistently predict defensive actions, providing a multilayered understanding of how physiological and psychological components jointly influence behavior under threat.

Finally, this thesis did not address the neural underpinnings of the risk-related behaviors examined. The neural correlates of risk aversion, cautiousness, and escape behavior could provide foundational insights into the brain regions and networks involved in processing different types of risk. Incorporating neuroimaging techniques, such as Electroencephalography or Optically-Pumped Magnetometers Magnetoencephalography (Brookes et al., 2022), could help map the neural circuits engaged during risk-taking and defensive behaviors, and explore how these circuits vary across contexts (e.g., economic decisions vs. immediate escape responses). Future studies could benefit from using neural and physiological measures in combination, examining how the brain and body coordinate during risk-related decisions in both controlled and naturalistic settings. This would not only enrich our understanding of the neural architecture supporting these behaviors but could also identify how specific neural pathways or networks respond differentially across risk contexts, thereby enhancing our ability to translate findings from animal research to human behavior.

5.7 Concluding remarks

Overall, the studies in this thesis collectively contribute to understanding risk-taking and defensive behaviors such as risk aversion, cautiousness, and escape behaviors by showing how internal factors, such as anxiety and compulsivity, and external factors, such as threat characteristics and task framing, interact to shape decision-making. Anxiety amplifies risk aversion as a state-dependent process and may influence escapes through specific effects, while compulsivity contributes to a more rigid, trait-like cautiousness. The distinct role of spider phobia, as revealed in the VR study, highlights how specific psychopathologies uniquely shape cautious behavior, underscoring the importance of tailored approaches in understanding threat responses. Crucially, all studies validate findings by replicating prior research and expanding knowledge, ensuring new insights build on reliable evidence rather than fragmented findings, while providing a deeper understanding of how internal and external factors converge to shape risk-related behaviors across diverse scenarios.

Bibliography

Abivardi, A., Khemka, S., and Bach, D. R. (2020). Hippocampal representation of threat features and behavior in a human approach–avoidance conflict anxiety task. *The Journal of Neuroscience*, 40(35):6748–6758.

Adams, R. A., Huys, Q. J. M., and Roiser, J. P. (2016). Computational psychiatry: Towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery and Psychiatry*, 87(1):53–63.

Adolphs, R. (2013). The biology of fear. *Current Biology*, 23(2):R79–R93.

Anusic, I. and Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology*, 110(5):766–781.

Arendasy, M. E., Hergovich, A., Sommer, M., and Bognar, B. (2005). Dimensionality and construct validity of a video-based, objective personality test for the assessment of willingness to take risks in road traffic. *Psychological Reports*, 97(1):309–320.

Argyriou, E., Um, M., Carron, C., and Cyders, M. A. (2018). Age and impulsive behavior in drug addiction: A review of past research and future directions. *Pharmacology Biochemistry and Behavior*, 164:106–117.

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders : DSM-5*. American Psychiatric Publishing, Washington DC, 5th edition.

Aupperle, R. L. and Paulus, M. P. (2010). Neural systems underlying approach and avoidance in anxiety disorders. *Dialogues in Clinical Neuroscience*, 12(4):517–531.

Aupperle, R. L., Sullivan, S., Melrose, A. J., Paulus, M. P., and Stein, M. B. (2011). A reverse translational approach to quantify approach-avoidance conflict in humans. *Behavioural Brain Research*, 225(2):455–463.

Aven, T. (2012). The risk concept-historical and recent development trends. *Reliability Engineering and System Safety*, 99:33–44.

Aven, T. and Renn, O. (2009). On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, 12(1):1–11.

Aven, T., Renn, O., and Rosa, E. A. (2011). On the ontological status of the concept of risk. *Safety Science*, 49(8–9):1074–1079.

Aylward, J., Hales, C., Robinson, E., and Robinson, O. J. (2020). Translating a rodent measure of negative bias into humans: the impact of induced anxiety and unmedicated mood and anxiety disorders. *Psychological Medicine*, 50:237–246.

Aylward, J., Valton, V., Ahn, W. Y., Bond, R. L., Dayan, P., Roiser, J. P., and Robinson, O. J. (2019). Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nature Human Behaviour*, 3:1116–1123.

Bach, D. R. (2015). Anxiety-like behavioural inhibition is normative under environmental threat-reward correlations. *PLoS Computational Biology*, 11(12):1–20.

Bach, D. R. (2017). The cognitive architecture of anxiety-like behavioral inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1):18–29.

Bach, D. R. (2021). Cross-species anxiety tests in psychiatry: pitfalls and promises. *Molecular Psychiatry*, 27(1):154–163.

Bach, D. R. and Dayan, P. (2017). Algorithms for survival: A comparative perspective on emotions. *Nature Reviews Neuroscience*, 18(5):311–319.

Bach, D. R., Guitart-Masip, M., Packard, P. A., Miró, J., Falip, M., Fuentemilla, L., and Dolan, R. J. (2014). Human hippocampus arbitrates approach-avoidance conflict. *Current Biology*, 24(12):1435.

Bach, D. R., Hoffmann, M., Finke, C., Hurlemann, R., and Ploner, C. J. (2019). Disentangling hippocampal and amygdala contribution to human anxiety-like behavior. *Journal of Neuroscience*, 39(43):8517–8526.

Bach, D. R., Korn, C. W., Vunder, J., and Bantel, A. (2018). Effect of valproate and pregabalin on human anxiety-like behaviour in a randomised controlled trial. *Translational Psychiatry*, 8.

Bach, D. R., Moutoussis, M., Bowler, A., and Dolan, R. J. (2020). Predictors of risky foraging behaviour in healthy young people. *Nature Human Behaviour*, 4(8):832–843.

Bados, A., Gómez-Benito, J., and Balaguer, G. (2010). The state-trait anxiety inventory, trait version: Does it really measure anxiety? *Journal of Personality Assessment*, 92(6):560–567.

Baek, K., Kwon, J., Chae, J. H., Chung, Y. A., Kralik, J. D., Min, J. A., Huh, H., Choi, K. M., Jang, K. I., Lee, N. B., Kim, S., Peterson, B. S., and Jeong, J. (2017). Heightened aversion to risk and loss in depressed patients with a suicide attempt history. *Scientific Reports*, 7:1–13.

Barrett, L. F. and Finlay, B. L. (2018). Concepts, goals and the control of survival-related behaviors. *Current Opinion in Behavioral Sciences*, 24:172–179.

Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., and Whiteford, H. A. (2014). The global burden of anxiety disorders in 2010. *Psychological Medicine*, 44:2363–2374.

Beard, C., Millner, A. J., Forgeard, M. J., Fried, E. I., Hsu, K. J., Treadway, M. T., Leonard, C. V., Kertz, S. J., and Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine*, 46:3359–3369.

Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*.

Beck, A. T., Emery, G., and Greenberg, R. L. (1985). *Anxiety disorders and phobias: A cognitive perspective*. Basic Books.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4(6):561–571.

Beck, J. G., Stanley, M. A., Averill, P. M., Baldwin, L. E., and Deagle III, E. A. (1998). The utility of the fear survey schedule-iii: An extended replication. *Journal of Anxiety Disorders*, 12(1):81–96.

Bedder, R. L., Vaghi, M. M., Dolan, R. J., and Rutledge, R. B. (2023). Risk taking for potential losses but not gains increases with time of day. *Scientific Reports*, 13(1):1–13.

Beddington, J., Cooper, C. L., Field, J., Goswami, U., Huppert, F. A., Jenkins, R., Jones, H. S., Kirkwood, T. B. L., Sahakian, B. J., and Thomas, S. M. (2008). The mental wealth of nations. *Nature*, 455:1057–1060.

Berenbaum, H., Thompson, R. J., and Pomerantz, E. M. (2007). The relation between worrying and concerns: The importance of perceived probability and cost. *Behaviour Research and Therapy*, 45:301–311.

Berg, J. M., Latzman, R. D., Bliwise, N. G., and Lilienfeld, S. O. (2015). Parsing the heterogeneity of impulsivity: A meta-analytic review of the behavioral implications of the upps for psychopathology. *Psychological Assessment*, 27(4):1129–1146.

Biedermann, S. V., Biedermann, D. G., Wenzlaff, F., Kurjak, T., Nouri, S., Auer, M. K., Wiedemann, K., Briken, P., Haaker, J., Lonsdorf, T. B., and Fuss, J. (2017). An elevated plus-maze in mixed reality for studying human anxiety-related behavior. *BMC Biology*, 15(1):1–14.

Bieling, P. J., Antony, M. M., and Swinson, R. P. (1998). The state–trait anxiety inventory, trait version: Structure and content re-examined. *Behaviour Research and Therapy*, 36(7-8):777–788.

Blair, K. S., Otero, M., Teng, C., Geraci, M., Ernst, M., Blair, R. J., Pine, D. S., and Grillon, C. (2017). Reduced optimism and a heightened neural response to everyday worries are specific to generalized anxiety disorder, and not seen in social anxiety. *Psychological medicine*, 47:1806–1815.

Blais, A.-R. and Weber, E. U. (2006). A domain-specific risk-taking (dospert) scale for adult populations. *Judgment and Decision Making*, 1(1):33–47.

Blanchard, C. D., Hynd, A. L., Minke, K. A., Minemoto, T., and Blanchard, R. J. (2001). Human defensive behaviors to threat scenarios show parallels to fear and anxiety-related defense patterns of non-human mammals. *Neuroscience & Biobehavioral Reviews*, 25(7-8):761–770.

Blanchard, R. J., Yudko, E. B., Rodgers, R. J., and Blanchard, D. C. (1993). Defense system psychopharmacology: an ethological approach to the pharmacology of fear and anxiety. *Behavioural brain research*, 58(1-2):155–165.

Bornovalova, M. A., Daughters, S. B., Hernandez, G. D., Richards, J. B., and Lejuez, C. W. (2005). Differences in impulsivity and risk-taking propensity between primary users of crack cocaine and primary users of heroin in a residential substance-use program. *Experimental and Clinical Psychopharmacology*, 13(4):311–318.

Brand, J. A., Henry, J., Melo, G. C., Wlodkowic, D., Wong, B. B. M., and Martin, J. M. (2023). Sex differences in the predictability of risk-taking behavior. *Behavioral Ecology*, 34(1):108–116.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., and Mushtaq, F. (2019). Studying human behavior with virtual reality: The unity experiment framework. *Behavior Research Methods*, 52:455–463.

Brookes, M. J., Leggett, J., Rea, M., Hill, R. M., Holmes, N., Boto, E., and Bowtell, R. (2022). Magnetoencephalography with optically pumped magnetometers (opm-meg): the next generation of functional neuroimaging. *Trends in Neurosciences*, 45:621–634.

Brown, A. L., Imai, T., Vieider, F., and Camerer, C. F. (2021). Meta-analysis of empirical estimates of loss-aversion. *SSRN Electronic Journal*, 1757282.

Buelow, M. T. and Barnhart, W. R. (2018). Test-retest reliability of common behavioral decision making tasks. *Archives of Clinical Neuropsychology*, 33(1):125–129.

Buelow, M. T. and Blaine, A. L. (2015). The assessment of risky decision making: A factor analysis of performance on the iowa gambling task, balloon analogue risk task, and columbia card task. *Psychological Assessment*, 27(3):777–785.

Buelow, M. T. and Suhr, J. A. (2009). Construct validity of the iowa gambling task. *Neuropsychology Review*, 19(1):102–114.

Burnette, C. B., Luzier, J., Bennett, B. L., Weisenmuller, C., Kerr, P., and Keener, J. (2022a). The tension between ethics and rigor when using amazon mturk for eating disorder research: Response to commentaries on burnette et al. (2021).

Burnette, C. B., Luzier, J. L., Bennett, B. L., Weisenmuller, C. M., Kerr, P., Martin, S., Keener, J., and Calderwood, L. (2022b). Concerns and recommendations for using amazon mturk for eating disorder research. *International Journal of Eating Disorders*, 55:263–272.

Butler, G. and Mathews, A. (1983). Cognitive processes in anxiety. *Advances in Behaviour Research and Therapy*, 5:51–62.

Butler, G. and Mathews, A. (1987). Anticipatory anxiety and risk perception. *Cognitive Therapy and Research*, 11:551–565.

Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125(3):367–383.

Campbell, M. L., Oldenburg, C., Thomas, S., Wainwright, K., and Ruffman, T. (2016). The situated fear questionnaire: Using emotional events to measure contextual individual differences in the appraisal and experience of fear. *Personality and Individual Differences*, 98:86–91.

Canessa, N., Crespi, C., Baud-Bovy, G., Dodich, A., Falini, A., Antonellis, G., and Cappa, S. F. (2017). Neural markers of loss aversion in resting-state brain activity. *NeuroImage*, 146:257–265.

Canessa, N., Crespi, C., Motterlini, M., Baud-Bovy, G., Chierchia, G., Pantaleo, G., Tettamanti, M., and Cappa, S. F. (2013). The functional and structural neural basis of individual differences in loss aversion. *Journal of Neuroscience*, 33(36):14307–14317.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., and Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2):119–137.

Castegnetti, G., Tzovara, A., Khemka, S., Melinščak, F., Barnes, G. R., Dolan, R. J., and Bach, D. R. (2020). Representation of probabilistic outcomes during risky decision-making. *Nature Communications*, 11(1):1–11.

Cavanagh, K. and Davey, G. C. L. (2000). The development of a measure of individual differences in disgust. In *Proceedings of the British Psychological Society*, Winchester, UK. British Psychological Society.

Chaloupka, E. C., Kang, J., Mastrangelo, M. A., and Donnelly, M. S. (1997). Cardiorespiratory and metabolic responses during forward and backward walking. *Journal of Orthopaedic and Sports Physical Therapy*, 25(5):302–306.

Chandler, J., Mueller, P., and Paolacci, G. (2014). Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46:112–130.

Chapman, J., Snowberg, E., Wang, S., and Camerer, C. (2018). Loss attitudes in the u.s. population: Evidence from dynamically optimized sequential experimentation (dose).

Charness, G., Garcia, T., Offerman, T., and Villeval, M. C. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? *Journal of Risk and Uncertainty*, 60(2):99–123.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of economic behavior & organization*, 83(1):50–58.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior and Organization*, 87:43–51.

Charpentier, C. J., Aylward, J., Roiser, J. P., and Robinson, O. J. (2017). Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. *Biological Psychiatry*, 81:1014–1022.

Charpentier, C. J., De Martino, B., Sim, A. L., Sharot, T., and Roiser, J. P. (2016a). Emotion-induced loss aversion and striatal-amygdala coupling in low-anxious individuals. *Social Cognitive and Affective Neuroscience*, 11(4):569–579.

Charpentier, C. J., Hindocha, C., Roiser, J. P., and Robinson, O. J. (2016b). Anxiety promotes memory for mood-congruent faces but does not alter loss aversion. *Scientific Reports*, 6.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., and Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *The Lancet Psychiatry*, 3(3):243–250.

Chumbley, J. R., Krajbich, I., Engelmann, J. B., Russell, E., Uum, S. V., Koren, G., and Fehr, E. (2014). Endogenous cortisol predicts decreased loss aversion in young men. *Psychological Science*, 25:2102–2105.

Cisler, J. M., Olatunji, B. O., and Lohr, J. M. (2009). Disgust sensitivity and emotion regulation potentiate the effect of disgust propensity on spider fear, blood-injection-injury fear, and contamination fear. *Journal of Behavior Therapy and Experimental Psychiatry*, 40:219–229.

Condon, D. M. and Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43(1):52–64.

Corporation, H. (2023). Vive pro eye specs & user guide.

Craske, M. G., Stein, M. B., Eley, T. C., Milad, M. R., Holmes, A., Rapee, R. M., and Wittchen, H.-U. (2017). Anxiety disorders. *Nature Reviews Disease Primers*, 3:17024.

Cross, C. P., Copping, L. T., and Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, 137(1):97–130.

Cross, C. P., Cyrenne, D.-L. M., and Brown, G. R. (2013). Sex differences in sensation-seeking: A meta-analysis. *Scientific Reports*, 3(1):2486.

Crump, M. J., McDonnell, J. V., and Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PLoS ONE*, 8.

Cryan, J. F. and Sweeney, F. F. (2011). Using appropriate animal models to guide clinical drug development the age of anxiety: role of animal models of anxiolytic action in drug discovery. *British Journal of Pharmacology*, 164(4).

Cuthbert, B. N. (2022). Research domain criteria (rdoc): Progress and potential. *Current Directions in Psychological Science*, 31(2):107–114.

Cuthbert, B. N. and Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of rdoc. *BMC Medicine*, 11(1):126.

Cuthbert, B. N. and Kozak, M. J. (2013). Constructing constructs for psychopathology: The nimh research domain criteria. *Journal of Abnormal Psychology*, 122(3):928–937.

Daly, M. and Wilson, M. (1983). *Sex, evolution, and behavior*. PWS, Boston, second edition. edition.

Datta, S. R., Anderson, D. J., Branson, K., Perona, P., and Leifer, A. (2019). Computational neuroethology: A call to action. *Neuron*, 104(1):11–24.

Davis, M., Walker, D. L., Miles, L., and 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.

Davis, S., Nesbitt, K., and Nalivaiko, E. (2014). A systematic review of cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment*, volume 19, pages 1–9. ACM.

de Wit, S. J., Alonso, P., Schweren, L., et al. (2014). Multicenter voxel-based morphometry mega-analysis of structural brain scans in obsessive-compulsive disorder. *American Journal of Psychiatry*, 171(3):340–349.

Deakin, J., Aitken, M., Robbins, T., and Sahakian, B. J. (2004). Risk taking during decision-making in normal volunteers changes with age. *Journal of the International Neuropsychological Society*, 10(4):590–598.

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, 63:453–482.

DeBattista, C., Kinrys, G., Hoffman, D., et al. (2011). The use of referenced-eeg (reeg) in assisting medication selection for the treatment of depression. *Journal of Psychiatric Research*, 45(1):64–75.

Dickinson, A. and Balleine, B. (1995). Motivational control of instrumental action. *Current Directions in Psychological Science*, 4(5):162–167.

Dinsmoor, J. A. (1977). Escape, avoidance, punishment: Where do we stand? *Journal of the Experimental Analysis of Behavior*, 28(1):83–95.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dolan, R. J. and Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2):312–325.

Dreber, A., Rand, D. G., Wernerfelt, N., Garcia, J. R., Vilar, M. G., Lum, J. K., and Zeckhauser, R. (2011). Dopamine and risk choices in different domains: Findings among serious tournament bridge players. *Journal of Risk and Uncertainty*, 43(1):19–38.

Dugas, M. J., Freeston, M. H., and Ladouceur, R. (1997). Intolerance of uncertainty and problem orientation in worry. *Cognitive Therapy and Research*, 21:593–606.

Dugas, M. J., Gagnon, F., Ladouceur, R., and Freeston, M. H. (1998). Generalized anxiety disorder: A preliminary test of a conceptual model. *Behaviour Research and Therapy*, 36:215–226.

Dugas, M. J. and Ladouceur, R. (2000). Treatment of gad. *Behavior Modification*, 24:635–657.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J., Harris, W. A., Lowry, R., McManus, T., Chyen, D., et al. (2008). Youth risk behavior surveillance–united states, 2007. *MMWR: Morbidity & Mortality Weekly Report*, 57.

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4):281–295.

Ernst, M., Plate, R. C., Carlisi, C. O., Gorodetsky, E., Goldman, D., and Pine, D. S. (2014). Loss aversion and 5htt gene variants in adolescent anxiety. *Developmental Cognitive Neuroscience*, 8:77–85.

Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., and Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. *Nature*, 525(7568):243–246.

Evans, D. A., Stempel, A. V., Vale, R., and Branco, T. (2019). Cognitive control of escape behaviour. *Trends in Cognitive Sciences*, 23(5):334–348.

Evans, D. A., Stempel, A. V., Vale, R., Ruehle, S., Lefler, Y., and Branco, T. (2018). A synaptic threshold mechanism for computing escape decisions. *Nature*, 558:590–594.

Fanselow, M. S. and Lester, L. S. (1988). A functional behavioristic approach to aversively motivated behavior: Predatory imminence as a determinant of the topography of defensive behavior. In Bolles, R. C. and Beecher, M. D., editors, *Evolution and learning*, pages 185–212. Lawrence Erlbaum Associates.

Feng, Y., Krahé, C., Koster, E. H., Lau, J. Y., and Hirsch, C. R. (2022). Cognitive processes predict worry and anxiety under different stressful situations. *Behaviour Research and Therapy*, 157:104168.

Fergus, T. A. and Valentiner, D. P. (2009). The disgust propensity and sensitivity scale-revised: An examination of a reduced-item version. *Journal of Anxiety Disorders*, 23(5):703–710.

Figner, B., Mackinlay, R. J., Wilkening, F., and Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the columbia card task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3):709–730.

First, M. B. and Gibbon, M. (2004). *The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II).*, pages 134–143. John Wiley & Sons, Inc.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., and Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a short version. *Psychological Assessment*, 14(4):485–496.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., and Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, 3(10).

Frey, R., Richter, D., Schupp, J., Hertwig, R., and Mata, R. (2021). Identifying robust correlates of risk preference: A systematic approach using specification curve analysis. *Journal of Personality and Social Psychology*, 120(2):538–557.

Fricke, K. and Vogel, S. (2020). How interindividual differences shape approach-avoidance behavior: Relating self-report and diagnostic measures of interindividual differences to behavioral measurements of approach and avoidance. *Neuroscience and Biobehavioral Reviews*, 111:30–56.

Friston, K. J., Stephan, K. E., Montague, P. R., and Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry*, 1(2):148–158.

Fung, B. J., Qi, S., Hassabis, D., Daw, N., and Mobbs, D. (2019). Slow escape decisions are swayed by trait anxiety. *Nature Human Behaviour*, 3(7):702–708.

Garner, D. M., Olmsted, M. P., Bohr, Y., and Garfinkel, P. E. (1982). The eating attitudes test: psychometric features and clinical correlates. *Psychological Medicine*, 12(4):871–878.

Gillan, C. M. and Daw, N. D. (2016). Taking psychiatry research online. *Neuron*, 91(1):19–23.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., and Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *ELife*, 5:1–34.

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., Fineberg, N. A., Sahakian, B. J., and Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry*, 75(8):631–638.

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., and Wit, S. D. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *American Journal of Psychiatry*, 168:718–726.

Giorgetta, C., Grecucci, A., Zuanon, S., Perini, L., Balestrieri, M., Bonini, N., Sanfey, A. G., and Brambilla, P. (2012). Reduced risk-taking behavior as a trait feature of anxiety. *Emotion*, 12(6):1373–1383.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Goetz, A. R., Lee, H.-Y., and Cougle, J. R. (2013). Factors associated with health-related avoidance: The role of disgust and contamination sensitivity. *Journal of Behavior Therapy and Experimental Psychiatry*, 44(4):416–421.

Golding, J. F. (1998). Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Research Bulletin*, 47(5):507–516.

Golding, J. F. (2006). Predicting individual differences in motion sickness susceptibility by questionnaire. *Personality and Individual Differences*, 41(2):237–248.

Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M., and Mainen, Z. F. (2014). Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, 17(11):1455–1462.

Gray, J. A. and McNaughton, N. (2000). *The neuropsychology of anxiety: an enquiry into the functions of the septohippocampal system*. Oxford UP.

Gromer, D., Kiser, D. P., and Pauli, P. (2021). Thigmotaxis in a virtual human open field test. *Scientific Reports*, 11(1).

Gros, D. F., Antony, M. M., Simms, L. J., and McCabe, R. E. (2007). Psychometric properties of the state-trait inventory for cognitive and somatic anxiety (sticsa): Comparison to the state-trait anxiety inventory (stai). *Psychological Assessment*, 19(4):369–381.

Grupe, D. W. and Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, 14:488–501.

Habig, B., Chiyo, P. I., and Lahti, D. C. (2017). Male risk-taking is related to number of mates in a polygynous bird. *Behavioral Ecology*, 28(2):541–548.

Hadlaczky, G., Hökby, S., Mkrtchian, A., Wasserman, D., Balazs, J., Machín, N., Sarchiapone, M., Sisask, M., and Carli, V. (2018). Decision-making in suicidal behavior: The protective role of loss aversion. *Frontiers in Psychiatry*, 9.

Haidt, J., McCauley, C., and Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, 16(5):701–713.

Haines, N., Kvam, P., Beauchaine, T., and Turner, B. M. (2020). Learning from the reliability paradox: How theoretically informed generative models can advance the social, behavioral, and brain sciences. *PsyArXiv*. Preprint available at OSF.

Halligan, P. W. and David, A. S. (2001). Cognitive neuropsychiatry: towards a scientific psychopathology. *Nature Reviews Neuroscience*, 2(3):209–215.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, 32(1):50–55.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry*, 23(1):56–62.

Harmer, C. J., Cowen, P. J., and Goodwin, G. M. (2011). Efficacy markers in depression. *Journal of Psychopharmacology*, 25:1148–1158.

Harris, B. (1979). Whatever happened to little albert? *American psychologist*, 34(2):151.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (redcap)-a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2):377–381.

Hauser, T. U., Skvortsova, V., De Choudhury, M., and Koutsouleris, N. (2022). The promise of a model-based psychiatry: Building computational models of mental ill health. *The Lancet Digital Health*, 4(11):e816–e828.

Hayes, S., Hirsch, C. R., Krebs, G., and Mathews, A. (2010). The effects of modifying interpretation bias on worry in generalized anxiety disorder. *Behaviour Research and Therapy*, 48:171–178.

Haynes, J. (1895). Risk as an economic factor. *The Quarterly Journal of Economics*, 9(4):409.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83.

Hergovich, A., Arendasy, M. E., Sommer, M., and Bognar, B. (2007). The vienna risk-taking test - traffic: A new measure of road traffic risk-taking. *Journal of Individual Differences*, 28(4):198–204.

Hertwig, R., Wulff, D. U., and Mata, R. (2019). Three gaps and what they may mean for risk preference. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1766).

Hirsch, C. R., Krahé, C., Whyte, J., Loizou, S., Bridge, L., Norton, S., and Mathews, A. (2018). Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. *Journal of Consulting and Clinical Psychology*, 86:1017–1030.

Hirsch, C. R. and Mathews, A. (1997). Interpretative inferences when reading about emotional events. *Behaviour Research and Therapy*, 35:1123–1132.

Hirsch, C. R. and Mathews, A. (2012). A cognitive model of pathological worry. *Behaviour Research and Therapy*, 50:636–646.

Hirsch, C. R., Mathews, A., Lequertier, B., Perman, G., and Hayes, S. (2013). Characteristics of worry in generalized anxiety disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, 44:388–395.

Hirsch, C. R., Meeten, F., Krahé, C., and Reeder, C. (2016). Resolving ambiguity in emotional disorders: The nature and role of interpretation biases. *Annual Review of Clinical Psychology*, 12:281–305.

Hockey, G. R. J., Maule, A. J., Clough, P. J., and Bdzola, L. (2000). Effects of negative mood states on risk in everyday decision making. *Cognition and Emotion*, 14(6):823–855.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5).

Hong, R. Y. (2007). Worry and rumination: Differential associations with anxious and depressive symptoms and coping behavior. *Behaviour Research and Therapy*, 45:277–290.

Hopkins, A. K., Gillan, C., Roiser, J., Wise, T., and Sidarus, N. (2022). Optimising the measurement of anxious-depressive, compulsivity and intrusive thought and social withdrawal transdiagnostic symptom dimensions. *PsyArXiv*.

Horvath, P. and Zuckerman, M. (1993). Sensation seeking, risk appraisal, and risky behavior. *Personality and individual differences*, 14(1):41–52.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., and Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences*, 32(3):401–414.

Huys, Q. J. M., Browning, M., Paulus, M. P., and Frank, M. J. (2021a). Advances in the computational understanding of mental illness. *Neuropsychopharmacology*, 46(1):3–19.

Huys, Q. J. M., Browning, M., Paulus, M. P., and Frank, M. J. (2021b). Advances in the computational understanding of mental illness. *Neuropsychopharmacology*, 46(1):3–19.

Huys, Q. J. M., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3):404–413.

Huys, Q. J. M., Moutoussis, M., and Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Networks*, 24(6):544–551.

Hyman, S. E. (2007). Can neuroscience be integrated into the dsm-v? *Nature Reviews Neuroscience*, 8(9):725–732.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research domain criteria (rdoc): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7):748–751.

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA - Journal of the American Medical Association*, 318(13):1215–1216.

Iosifescu, D., Neborsky, R., and Valuck, R. (2016). The use of the psychiatric electroencephalography evaluation registry (peer) to personalize pharmacotherapy. *Neuropsychiatric Disease and Treatment*, 12:2131–2142.

Jazayeri, M. and Afraz, A. (2017). Navigating the neural space in search of the neural code. *Neuron*, 93(5):1003–1014.

Jong, P. J. D., Peters, M., and Vanderhallen, I. (2002). Disgust and disgust sensitivity in spider phobia: Facial emg in response to spider and oral disgust imagery. *Journal of Anxiety Disorders*, 16:477–493.

Josef, A. K., Richter, D., Samanez-Larkin, G. R., Wagner, G. G., Hertwig, R., and Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology*, 111(3):430–450.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2).

Kallai, J., Makany, T., Csatho, A., Karadi, K., Horvath, D., Kovacs-Labadi, B., Jarai, R., Nadel, L., and Jacobs, J. W. (2007). Cognitive and affective aspects of thigmotaxis strategy in humans. *Behavioral Neuroscience*, 121(1):21–30.

Karvelis, P., Paulus, M. P., and Diaconescu, A. O. (2023). Individual differences in computational psychiatry: A review of current challenges. *Neuroscience and Biobehavioral Reviews*, 148.

Kastrinogiannis, A. and Lonsdorf, T. B. (2023). Navigating the approach-avoidance matrix: Exploring individual differences in a novel virtual reality foraging task. *Society for Psychotherapy Research*.

Katyal, S., Huys, Q. J., Dolan, R. J., and Fleming, S. M. (2023). How underconfidence is maintained in anxiety and depression. *Nature Human Behaviour*, 7(3):352–361.

Kendler, K. S., Aggen, S. H., Knudsen, P. G., Røysamb, E., Neale, M. C., and Reichborn-Kjennerud, T. (2011). The structure of genetic and environmental risk factors for syndromal and subsyndromal common dsm-iv axis i and all axis ii disorders. *American Journal of Psychiatry*, 168.

Kendler, K. S., Gardner, C. O., Gatz, M., and Pedersen, N. L. (2007). The sources of co-morbidity between major depression and generalized anxiety disorder in a swedish national twin sample. *Psychological Medicine*, 37:453–462.

Kertz, S. J., Bigda-Peyton, J. S., Rosmarin, D. H., and Björgvinsson, T. (2012). The importance of worry across diagnostic presentations: Prevalence, severity and associated symptoms in a partial hospital setting. *Journal of Anxiety Disorders*, 26:126–133.

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., and Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the united states. *International Journal of Methods in Psychiatric Research*, 21:169–184.

Khemka, S., Barnes, G., Dolan, R. J., and Bach, D. R. (2017). Dissecting the function of hippocampal oscillations in a human anxiety model. *Journal of Neuroscience*, 37(29):6869–6876.

Khurana, A., Romer, D., Betancourt, L. M., Brodsky, N. L., Giannetta, J. M., and Hurt, H. (2015). Experimentation versus progression in adolescent drug use: A test of an emerging neurobehavioral imbalance model. *Development and Psychopathology*, 27(3):901–913.

Khurana, A., Romer, D., Betancourt, L. M., and Hurt, H. (2017). Working memory ability and early drug use progression as predictors of adolescent substance use disorders. *Addiction*, 112(7):1220–1228.

Khurana, A., Romer, D., Betancourt, L. M., and Hurt, H. (2018). Modeling trajectories of sensation seeking and impulsivity dimensions from early to late adolescence: Universal trends or distinct sub-groups? *Journal of Youth and Adolescence*, 47(9):1992–2005.

Kim, H. K., Park, J., Choi, Y.-J., and Choe, M. (2018). Development of a virtual reality sickness questionnaire (vrsq). *International Journal of Human-Computer Interaction*, 34(10):976–993.

King, J. E., Weiss, A., and Sisco, M. M. (2008). Aping humans: Age and sex effects in chimpanzee (pan troglodytes) and human (homo sapiens) personality. *Journal of Comparative Psychology*, 122(4):418–427.

Kircanski, K., Thompson, R. J., Sorenson, J. E., Sherdell, L., and Gotlib, I. H. (2015). Rumination and worry in daily life: Examining the naturalistic validity of theoretical constructs. *Clinical Psychological Science*, 3(6):926–939.

Klaus, F., Chumbley, J. R., Seifritz, E., Kaiser, S., and Hartmann-Riemer, M. (2020). Loss aversion and risk aversion in non-clinical negative symptoms and hypomania. *Frontiers in Psychiatry*, 11:1–7.

Klorman, R., Weerts, T., Hastings, J., Mela, B., and Lang, P. (1974). Psychometric description of some specific-fear questionnaires. *Behavior Therapy*, 5(3):401–409.

Knowles, K. A. and Olatunji, B. O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the state-trait anxiety inventory. *Clinical Psychology Review*, 82:101928.

Kober, S. E., Kurzmann, J., and Neuper, C. (2012). Cortical correlate of spatial presence in 2d and 3d interactive virtual reality: An eeg study. *International Journal of Psychophysiology*, 83(3):365–374.

Korn, C. W. and Bach, D. R. (2018). Heuristic and optimal policy computations in the human brain during sequential decision-making. *Nature Communications*, 9(1):1–11.

Korn, C. W. and Bach, D. R. (2019). Minimizing threat via heuristic and optimal policies recruits hippocampus and medial prefrontal cortex. *Nature Human Behaviour*, 3:733–745.

Korn, C. W., Vunder, J., Miró, J., Fuentemilla, L., Hurlemann, R., and Bach, D. R. (2017). Amygdala lesions reduce anxiety-like behavior in a human benzodiazepine-sensitive approach–avoidance conflict test. *Biological Psychiatry*, 82:522–531.

Kourtesis, P., Korre, D., MacPherson, S. E., and Autefage, V. (2023). Validation of the cybersickness questionnaire (csq-vr): A two-factor model that accurately assesses cybersickness in virtual reality. *Journal of Neuroscience Methods*, 383:109718.

Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., Derks, E. M., Fleischhacker, W. W., and Hasan, A. (2016). Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *The Lancet Psychiatry*, 3(10):935–946.

Kozlowska, K., Walker, P., McLean, L., and Carrive, P. (2015). Fear and the defense cascade: Clinical implications and management.

Lahey, B. B., Moore, T. M., Kaczkurkin, A. N., and Zald, D. H. (2021). Hierarchical models of psychopathology: empirical support, implications, and remaining issues. *World Psychiatry*, 20(1):57–63.

Lahey, B. B., Rathouz, P. J., Applegate, B., Tackett, J. L., and Waldman, I. D. (2010). Psychometrics of a self-report version of the child and adolescent dispositions scale. *Journal of Clinical Child and Adolescent Psychology*, 39(3):351–361.

Lane, N. E. and Kennedy, R. S. (1988). A new method for quantifying simulator sickness: Development and application of the simulator sickness questionnaire (ssq).

Lauriola, M., Panno, A., Levin, I. P., and Lejuez, C. W. (2014). Individual differences in risky decision making: A meta-analysis of sensation seeking and impulsivity with the balloon analogue risk task. *Journal of Behavioral Decision Making*, 27(1):20–36.

Lauriola, M., Russo, P. M., Lucidi, F., Violani, C., and Levin, I. P. (2005). The role of personality in positively and negatively framed risky health decisions. *Personality and Individual Differences*, 38(1):45–59.

Leahy, R. L., Tirch, D. D., and Melwani, P. S. (2012). Processes underlying depression: Risk aversion, emotional schemas, and psychological flexibility. *International Journal of Cognitive Therapy*, 5:362–379.

LeDoux, J. and Daw, N. D. (2018). Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nature Reviews Neuroscience*, 19(5):269–282.

LeDoux, J. E., Moscarello, J., Sears, R., and Campese, V. (2017). The birth, death and resurrection of avoidance: A reconceptualization of a troubled paradigm. *Molecular Psychiatry*, 22(1):24–36.

Lee, J., Rouault, M., and Wyart, V. (2023). Compulsivity is linked to maladaptive choice variability but unaltered reinforcement learning under uncertainty. *BioRxiv*, 914:1–20.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., and Pedulla, C. M. (2003). Evaluation of the balloon analogue risk task (bart) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence*, 26(4):475–479.

Lejuez, C. W., Richards, J. B., Read, J. P., Kahler, C. W., Ramsey, S. E., Stuart, G. L., Strong, D. R., and Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, 8(2):75–84.

Lewis, G., Srinivasan, R., Roiser, J., Blakemore, S.-J., Flouri, E., and Lewis, G. (2022). Risk-taking to obtain reward: sex differences and associations with emotional and depressive symptoms in a nationally representative cohort of uk adolescents. *Psychological Medicine*, 52(13):2805–2813.

Liebowitz, M. R. (1987). Delineating social phobia. *British Journal of Psychiatry*, 150(5):718–719.

Lighthall, N. R., Mather, M., and Gorlick, M. A. (2009). Acute stress increases sex differences in risk seeking in the balloon analogue risk task. *PLoS ONE*, 4(7).

Lighthall, N. R., Sakaki, M., Vasunilashorn, S., Nga, L., Somayajula, S., Chen, E. Y., Samii, N., and Mather, M. (2012). Gender differences in reward-related decision processing under stress. *Social Cognitive and Affective Neuroscience*, 7(4):476–484.

Lilienfeld, S. O. and Treadway, M. T. (2016). Clashing diagnostic approaches: Dsm-icd versus rdoc. *Annual Review of Clinical Psychology*, 12:435–463.

Lin, C.-H., Song, T.-J., Chen, Y.-Y., Lee, W.-K., and Chiu, Y.-C. (2013). Reexamining the validity and reliability of the clinical version of the iowa gambling task: Evidence from a normal subject group. *Frontiers in Psychology*, 4.

Loh, E., Kurth-Nelson, Z., Berron, D., Dayan, P., Duzel, E., Dolan, R., and Guitart-Masip, M. (2017). Parsing the role of the hippocampus in approach-avoidance conflict. *Cerebral Cortex*, 27(1):201–215.

Loree, A. M., Lundahl, L. H., and Ledgerwood, D. M. (2015). Impulsivity as a predictor of treatment outcome in substance use disorders: Review and synthesis. *Drug and Alcohol Review*, 34(2):119–134.

Lydon-Staley, D. M., Falk, E. B., and Bassett, D. S. (2020). Within-person variability in sensation-seeking during daily life: Positive associations with alcohol use and self-defined risky behaviors. *Psychology of Addictive Behaviors*, 34:257–268.

L'Haridon, O. and Vieider, F. M. (2019). All over the map: A worldwide comparison of risk preferences. *Quantitative Economics*, 10(1):185–215.

Maccrimmon, K. R. and Wehrung, D. A. (1990). Characteristics of risk taking executives. *Science*, 36(4).

MacKillop, J., Weafer, J., Gray, J. C., Oshri, A., Palmer, A., and de Wit, H. (2016). The latent structure of impulsivity: impulsive choice, impulsive action, and impulsive personality traits. *Psychopharmacology*, 233(18):3361–3370.

MacLeod, C. and Cohen, I. L. (1993). Anxiety and the interpretation of ambiguity: A text comprehension study. *Journal of Abnormal Psychology*, 102:238–247.

Mahoney, A. E. and McEvoy, P. M. (2012). A transdiagnostic examination of intolerance of uncertainty across anxiety and depressive disorders. *Cognitive Behaviour Therapy*, 41:212–222.

Maia, T. V. and Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, 14(2):154–162.

Maner, J. K., Richey, J. A., Cromer, K., Mallott, M., Lejuez, C. W., Joiner, T. E., and Schmidt, N. B. (2007). Dispositional anxiety and risk-avoidant decision-making. *Personality and Individual Differences*, 42(4):665–675.

Marin, R. S., Biedrzycki, R. C., and Firinciogullari, S. (1991). Reliability and validity of the apathy evaluation scale. *Psychiatry Research*, 38(2):143–162.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Marsch, L. A. (2021). Digital health data-driven approaches to understand human behavior. *Neuropsychopharmacology*, 46(1):191–196.

Martin, C. A., Kelly, T. H., Rayens, M. K., Brogli, B. R., Brenzel, A., Smith, W. J., and Omar, H. A. (2002). Sensation seeking, puberty, and nicotine, alcohol, and marijuana use in adolescence. *Journal of the American academy of child & adolescent psychiatry*, 41(12):1495–1502.

Martino, B. D., Camerer, C. F., and Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8):3788–3792.

Mason, O., Linney, Y., and Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophrenia Research*, 78(2–3):293–296.

Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, 44:1–23.

Mata, R., Frey, R., Richter, D., Schupp, J., and Hertwig, R. (2018). Risk preference: A view from psychology. *Journal of Economic Perspectives*, 32(2):155–172.

Mata, R., Josef, A. K., and Hertwig, R. (2016). Propensity for risk taking across the life span and around the globe. *Psychological Science*, 27(2):231–243.

Matchett, G. and Davey, G. C. L. (1991). A test of a disease-avoidance model of animal phobias. *Behaviour Research and Therapy*, 29(1):91–94.

Mcevoy, P. M. and Mahoney, A. E. J. (2011). To be sure, to be sure: Intolerance of uncertainty mediates symptoms of various anxiety disorders and depression.

McNaughton, N., DeYoung, C. G., and Corr, P. J. (2016). *Approach/Avoidance*, pages 25–49. Elsevier Inc.

Meyer, T. J., Miller, M. L., Metzger, R. L., and Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, 28(6):487–495.

Miranda, R. and Mennin, D. S. (2007). Depression, generalized anxiety disorder, and certainty in pessimistic predictions about the future. *Cognitive Therapy and Research*, 31:71–82.

Mishra, S. (2014). Decision-making under risk: Integrating perspectives from biology, economics, and psychology. *Personality and Social Psychology Review*, 18(3):280–307.

Mishra, S. and Lalumière, M. L. (2011). Individual differences in risk-propensity: Associations between personality and behavioral measures of risk. *Personality and Individual Differences*, 50(6):869–873.

Mitte, K. (2007). Anxiety and risky decision-making: The role of subjective probability and subjective costs of negative events. *Personality and Individual Differences*, 43(2):243–253.

Mkrtchian, A., Valton, V., and Roiser, J. P. (2023). Reliability of decision-making and reinforcement learning computational parameters. *Computational Psychiatry*, 7(1):30–46.

Mobbs, D., Headley, D. B., Ding, W., and Dayan, P. (2020). Space, time, and fear: Survival computations along defensive circuits. *Trends in Cognitive Sciences*, 24(3):228–241.

Mobbs, D. and Kim, J. J. (2015). Neuroethological studies of fear, anxiety, and risky decision-making in rodents and humans. *Current Opinion in Behavioral Sciences*, 5:8–15.

Mobbs, D. and LeDoux, J. (2018). Editorial overview: Survival behaviors and circuits. *Current Opinion in Behavioral Sciences*, 24:168–171.

Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., Dolan, R. J., and Frith, C. D. (2007). When fear is near: Threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science*, 317(5841):1079–1083.

Mobbs, D., Wise, T., Suthana, N., Guzmán, N., Kriegeskorte, N., and Leibo, J. Z. (2021). Promises and challenges of human computational ethology. *Neuron*, 109(14):2224–2238.

Moffitt, T. E., Caspi, A., Taylor, A., Kokaua, J., Milne, B. J., Polanczyk, G., and Poulton, R. (2010). How common are common mental disorders? evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychological Medicine*, 40:899–909.

Mohammadi, M. R., Pourdehghan, P., Mostafavi, S. A., Hooshyari, Z., Ahmadi, N., and Khaleghi, A. (2020). Generalized anxiety disorder: Prevalence, predictors, and comorbidity in children and adolescents. *Journal of Anxiety Disorders*, 73.

Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1):72–80.

Montgomery, S. A. and Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, 134:382–389.

Moritz, S. and Jelinek, L. (2009). Inversion of the "unrealistic optimism" bias contributes to overestimation of threat in obsessive-compulsive disorder. *Behavioural and Cognitive Psychotherapy*, 37:179–193.

Moritz, S. and Pohl, R. F. (2009). Biased processing of threat-related information rather than knowledge deficits contributes to overestimation of threat in obsessive-compulsive disorder. *Behavior Modification*, 33:763–777.

Morris, R. W., Quail, S., Griffiths, K. R., Green, M. J., and Balleine, B. W. (2015). Corticostriatal control of goal-directed action is impaired in schizophrenia. *Biological Psychiatry*, 77:187–195.

Moutoussis, M., Garzón, B., Neufeld, S., Bach, D. R., Rigoli, F., Goodyer, I., Bullmore, E., Fonagy, P., Jones, P., Hauser, T., et al. (2021). Decision-making ability, psychopathology, and brain connectivity. *Neuron*, 109(12):2025–2040.

Moutoussis, M., Shahar, N., Hauser, T. U., and Dolan, R. J. (2018). Computation in psychotherapy, or how computational psychiatry can aid learning-based psychological therapies. *Computational Psychiatry*, 2:50–73.

Newman, M. G., Llera, S. J., Erickson, T. M., Przeworski, A., and Castonguay, L. G. (2013). Worry and generalized anxiety disorder: A review and theoretical synthesis of evidence on nature, etiology, mechanisms, and treatment. *Annual Review of Clinical Psychology*, 9:275–297.

Nicholson, N., Soane, E., Fenton-O'Creevy, M., and Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research*, 8(2):157–176.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin*, 101(2):259–282.

Nolen-Hoeksema, S., Wisco, B. E., and Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3:400–424.

Olatunji, B. O., Broman-Fulks, J. J., Bergman, S. M., Green, B. A., and Zlomke, K. R. (2010). A taxometric investigation of the latent structure of worry: Dimensionality and associations with depression, anxiety, and stress. *Behavior Therapy*, 41:212–228.

Olatunji, B. O. and Deacon, B. (2008). Specificity of disgust sensitivity in the prediction of fear and disgust responding to a brief spider exposure. *Journal of Anxiety Disorders*, 22:328–336.

Olatunji, B. O., Williams, N. L., Lohr, J. M., and Sawchuk, C. N. (2007). The disgust scale: Item analysis, factor structure, and suggestions for refinement. *Psychological Assessment*, 19(3):281–297.

O'Neil, E. B., Newsome, R. N., Li, I. H. N., Thavabalasingam, S., Ito, R., and Lee, A. C. H. (2015). Examining the role of the human hippocampus in approach-avoidance decision making using a novel conflict paradigm and multivariate functional magnetic resonance imaging. *Journal of Neuroscience*, 35(45):15039–15049.

Onnela, J.-P. and Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*, 41(7):1691–1696.

Orsini, C. A., Willis, M. L., Gilbert, R. J., Bizon, J. L., and Setlow, B. (2016). Sex differences in a rat model of risky decision making. *Behavioral Neuroscience*, 130(1):50–61.

Overman, W. H. (2004). Sex differences in early childhood, adolescence, and adulthood on cognitive tasks that rely on orbital prefrontal cortex. *Brain and Cognition*, 55(1):134–147.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*, 21(6):425–433.

Panayiotou, G., Karekla, M., and Mete, I. (2014). Dispositional coping in individuals with anxiety disorder symptomatology: Avoidance predicts distress. *Journal of Contextual Behavioral Science*, 3:314–321.

Patton, J. H., Stanford, M. S., and Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, 51(6):768–774.

Patzelt, E. H., Kool, W., Millner, A. J., and Gershman, S. J. (2019). The transdiagnostic structure of mental effort avoidance. *Scientific Reports*, 9.

Pawlowski, B., Atwal, R., and Dunbar, R. I. M. (2008). Sex differences in everyday risk-taking behavior in humans. *Evolutionary Psychology*.

Payzan-LeNestour, E. and Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS computational biology*, 7(1):e1001048.

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., and Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, 1(11):803–809.

Pennings, J. M. E. and Smidts, A. (2000). Assessing the construct validity of risk attitude. *Management Science*, 46(10):1337–1348.

Penninx, B. W., Nolen, W. A., Lamers, F., Zitman, F. G., Smit, J. H., Spinhoven, P., Cuijpers, P., Jong, P. J. D., Marwijk, H. W. V., Meer, K. V. D., Verhaak, P., Laurant, M. G., Graaf, R. D., Hoogendijk, W. J., Wee, N. V. D., Ormel, J., Dyck, R. V., and Beekman, A. T. (2011). Two-year course of depressive and anxiety disorders: Results from the netherlands study of depression and anxiety (nesda). *Journal of Affective Disorders*, 133:76–85.

Polák, J., Sedláčková, K., Landová, E., and Frynta, D. (2020). Faster detection of snake and spider phobia: revisited. *Heliyon*, 6(5).

Polák, M., Mikuláš, M., Janovcová, M., Sedláčková, L., Landová, E., and Frynta, D. (2020). The snake detection hypothesis and the role of fear in visual perception: A meta-analysis. *Frontiers in Psychology*, 11:1–16.

Qi, S., Hassabis, D., Sun, J., Guo, F., Daw, N., and Mobbs, D. (2018). How cognitive and reactive fear circuits optimize escape decisions in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 115(12):3186–3191.

Reavis, R. and Overman, W. H. (2001). Adult sex differences on a decision-making task previously shown to depend on the orbital prefrontal cortex. *Behavioral Neuroscience*, 115(1):196–206.

Rebenitsch, L. and Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2):101–125.

Ree, M. J., French, D., MacLeod, C., and Locke, V. (2008). Distinguishing cognitive and somatic dimensions of state and trait anxiety: Development and validation of the state-trait inventory for cognitive and somatic anxiety (sticsa). *Behavioural and Cognitive Psychotherapy*, 36(3):313–332.

Ree, M. J., MacLeod, C., French, D., and Locke, V. (2000). The state–trait inventory for cognitive and somatic anxiety: Development and validation. In *Annual meeting of the Association for the Advancement of Behavior Therapy, New Orleans, LA*.

Reinhold, J. A., Mandos, L. A., Rickels, K., and Lohoff, F. W. (2011). Pharmacological treatment of generalized anxiety disorder. *Expert Opinion on Pharmacotherapy*, 12:2457–2467.

Reynolds, B. W., Basso, M. R., Miller, A. K., Whiteside, D. M., and Combs, D. (2019). Executive function, impulsivity, and risky behaviors in young adults. *Neuropsychology*, 33(2):212–221.

Rieger, M. O., Wang, M., and Hens, T. (2015). Risk preferences around the world. *Management Science*, 61(3):637–648.

Roberti, J. W., Storch, E. A., and Bravata, E. (2003). Further psychometric support for the sensation seeking scale-form v. *Journal of Personality Assessment*, 81(3):291–292.

Roberts, E. K., Spielberger, C. D., Phillips, W. M., and Pettit, J. W. (2016). Validation of the state-trait inventory for cognitive and somatic anxiety (sticsa) in an adolescent population. In *Annual meeting of the Association for Behavioral and Cognitive Therapies*. ABCT.

Robinson, O. J., Vytal, K., Cornwell, B. R., and Grillon, C. (2013). The impact of anxiety upon cognition: perspectives from human threat of shock studies. *Frontiers in Human Neuroscience*, 7:1–21.

Rodgers, R. J., Cao, B. J., Dalvi, A., and Holmes, A. (1997). Animal models of anxiety: An ethological perspective. *Brazilian Journal of Medical and Biological Research*, 30(3):289–304.

Rodrigues, J., Studer, E., Streuber, S., Meyer, N., and Sandi, C. (2020). Locomotion in virtual environments predicts cardiovascular responsiveness to subsequent stressful challenges. *Nature Communications*, 11(1).

Rouault, M., Seow, T., Gillan, C. M., and Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry*, 84(6):443–451.

Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., Davison, C., Demić, E., Esteban-Serna, C., Friedemann, M., et al. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, 4(6):622–633.

Ruscio, A. M. and Borkovec, T. D. (2004). Experience and appraisal of worry among high worriers with and without generalized anxiety disorder. *Behaviour Research and Therapy*, 42:1469–1482.

Rutledge, R. B., Chekroud, A. M., and Huys, Q. J. M. (2019). Machine learning and big data in psychiatry: toward clinical applications. *Current Opinion in Neurobiology*, 55:152–159.

Rutledge, R. B., Skandali, N., Dayan, P., and Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being (si methods). *Proceedings of the National Academy of Sciences*, 111(33):12252–12257.

Rutledge, R. B., Smittenaar, P., Zeidman, P., Brown, H. R., Adams, R. A., Lindenberger, U., Dayan, P., and Dolan, R. J. (2016). Risk taking for potential reward decreases across the lifespan. *Current Biology*, 26(12):1634–1639.

Sandhu, T. R., Xiao, B., and Lawson, R. P. (2023). Transdiagnostic computations of uncertainty: towards a new lens on intolerance of uncertainty. *Neuroscience & Biobehavioral Reviews*, 148:105123.

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., and Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience*, 14:1–19.

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., and Grant, M. (1993). Development of the alcohol use disorders identification test (audit): Who collaborative project on early detection of persons with harmful alcohol consumption-ii. *Addiction*, 88(6):791–804.

Schad, D. J., Jünger, E., Sebold, M., Garbusow, M., Bernhardt, N., Javadi, A. H., Zimmermann, U. S., Smolka, M. N., Heinz, A., Rapp, M. A., and Huys, Q. J. M. (2014). Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. *Frontiers in Psychology*, 5:1450.

Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives*, 32(2):135–154.

Schoevers, R. A., Beekman, A. T., Deeg, D. J., Jonker, C., and Tilburg, W. V. (2003). Comorbidity and risk-patterns of depression, generalised anxiety disorder and mixed anxiety-depression in later life: Results from the amstel study. *International Journal of Geriatric Psychiatry*, 18:994–1001.

Schoevers, R. A., Deeg, D. J., Tilburg, W. V., and Beekman, A. T. (2005). Depression and generalized anxiety disorder: Co-occurrence and longitudinal patterns in elderly patients. *American Journal of Geriatric Psychiatry*, 13:31–39.

Scholl, J. and Klein-Flügge, M. (2018). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. *Behavioural Brain Research*, 355:56–75.

Schonberg, T., Fox, C. R., and Poldrack, R. A. (2011). Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in Cognitive Sciences*, 15(1):11–19.

Schuetz, I. and Fiehler, K. (2022). Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research*, 15(3):1–18.

Schulreich, S., Heekeren, H. R., and Gerhardt, H. (2016). Incidental fear cues increase monetary loss aversion. *Emotion*, 16(3):402–412.

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P. J., Dalley, J. W., Flor, H., Gallinat, J., et al. (2010). The imagen study: Reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry*, 15(12):1128–1139.

Schwebel, D. C., Severson, J., Ball, K. K., and Rizzo, M. (2006). Individual difference factors in risky driving: The roles of anger/hostility, conscientiousness, and sensation-seeking. *Accident Analysis & Prevention*, 38(4):801–810.

Sege, C. T., Bradley, M. M., and Lang, P. J. (2018). Avoidance and escape: Defensive reactivity and trait anxiety. *Behaviour Research and Therapy*, 104:62–68.

Seow, T. X. F., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., McDonough, M., and Gillan, C. M. (2020). A dimensional investigation of error-related negativity (ern) and self-reported psychiatric symptoms. *International Journal of Psychophysiology*, 158:340–348.

Seow, T. X. F., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., O'Connell, R., and Gillan, C. M. (2021). Model-based planning deficits in compulsivity are linked to faulty neural representations of task structure. *Journal of Neuroscience*, 41(30):6539–6550.

Seow, T. X. F. and Gillan, C. M. (2020). Transdiagnostic phenotyping reveals a host of metacognitive deficits implicated in compulsivity. *Scientific Reports*, 10(1):1–11.

Sevinc, V. and Berkman, E. T. (2020). Disentangling the effects of working memory and perceived closeness on cybersickness in virtual reality. *Cognitive Research: Principles and Implications*, 5(1):1–12.

Seyfarth, R. M. and Cheney, D. L. (2003). Meaning and emotion in animal vocalizations. *Annals of the New York Academy of Sciences*, 1000:32–55.

Shackman, A. J. and Fox, A. S. (2016). Contributions of the central extended amygdala to fear and anxiety. *Journal of Neuroscience*, 36(31):8050–8063.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge University Press.

Shapiro, D. N., Chandler, J., and Mueller, P. A. (2013). Using mechanical turk to study clinical populations. *Clinical Psychological Science*, 1(2):213–220.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (2016). Facts and fears: Understanding perceived risk. In *The perception of risk*, pages 175–191. Routledge.

Smith, A. R., Ebert, E. E., and Broman-Fulks, J. J. (2016). The relationship between anxiety and risk taking is moderated by ambiguity. *Personality and Individual Differences*, 95:40–44.

Smith, R., Kirlic, N., Stewart, J. L., Touthang, J., Kuplicki, R., Khalsa, S. S., Feinstein, J., Paulus, M. P., and Aupperle, R. L. (2021). Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: A computational modelling approach. *Journal of Psychiatry and Neuroscience*, 46(1):E74–E87.

Smoski, M. J., Lynch, T. R., Rosenthal, M. Z., Cheavens, J. S., Chapman, A. L., and Krishnan, R. R. (2008). Decision-making and risk aversion among depressive adults. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(4):567–576.

Sokol-Hessner, P., Camerer, C. F., and Phelps, E. A. (2013). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Social Cognitive and Affective Neuroscience*, 8:341–350.

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., and Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *PNAS*, 106.

Sokol-Hessner, P., Raio, C. M., Gottesman, S. P., Lackovic, S. F., and Phelps, E. A. (2016). Acute stress does not affect risky monetary decision-making. *Neurobiology of Stress*, 5:19–25.

Sokol-Hessner, P. and Rutledge, R. B. (2019a). The psychological and neural basis of loss aversion. *Current Directions in Psychological Science*, 28(1):20–27.

Sokol-Hessner, P. and Rutledge, R. B. (2019b). The psychological and neural basis of loss aversion. *Current Directions in Psychological Science*, 28:20–27.

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences*, 23(8):699–714.

Sookman, D. and Pinard, G. (2002). *Overestimation of Threat and Intolerance of Uncertainty in Obsessive Compulsive Disorder*, pages 63–89. Pergamon.

Spielberger, C. D. (1983). *Manual for the State–Trait Anxiety Inventory (Form Y)*. Mind Garden, Palo Alto, CA.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder. *Archives of Internal Medicine*, 166(10):1092–1097.

Sporrer, J. K., Brookes, J., Hall, S., Zabbah, S., Serratos Hernandez, U. D., and Bach, D. R. (2023). Functional sophistication in human escape. *IScience*, 26(11):1–14.

Starcevic, V., Berle, D., Milicevic, D., Hannan, A., Lamplugh, C., and Eslick, G. D. (2007). Pathological worry, anxiety disorders and the impact of co-occurrence with depressive and other anxiety disorders. *Journal of Anxiety Disorders*, 21:1016–1027.

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., and Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, 43(2):476–493.

Steimer, T. (2002). The biology of fear-and anxiety-related behaviors. *Dialogues in Clinical Neuroscience*, 4(3):231–249.

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., and Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, 16(7):966–973.

Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., Heinz, A., Huys, Q. J. M., Owen, M. J., Binder, E. B., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Montague, P. R., Schnyder, U., Wang, X. J., and Breakspear, M. (2016). Charting the landscape of priority problems in psychiatry, part 1: Classification and diagnosis. *The Lancet Psychiatry*, 3(1):77–83.

Stephan, K. E., Friston, K. J., and Frith, C. D. (2009). Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia bulletin*, 35(3):509–527.

Stephenson, M. T., Hoyle, R. H., Palmgreen, P., and Slater, M. D. (2003). Brief measures of sensation seeking for screening and large-scale surveys. *Drug and Alcohol Dependence*, 72(3):279–286.

Stigler, G. J. and Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, 67(2):76–90.

Struijs, S. Y., Lamers, F., Rinck, M., Roelofs, K., Spinhoven, P., and Penninx, B. W. J. H. (2018). The predictive value of approach and avoidance tendencies on the onset and course of depression and anxiety disorders. *Depression and Anxiety*, 35(6):551–559.

Struijs, S. Y., Lamers, F., Vroling, M. S., Roelofs, K., Spinhoven, P., and Penninx, B. W. J. H. (2017). Approach and avoidance tendencies in depression and anxiety disorders. *Psychiatry Research*, 256:475–481.

Stöber, J. (1997). Trait anxiety and pessimistic appraisal of risk and chance. *Personality and Individual Differences*, 22:465–476.

Tallis, F., Eysenck, M., and Mathews, A. (1992). A questionnaire for the measurement of nonpathological worry. *Personality and Individual Differences*, 13(2):161–168.

ter Meulen, W. G., Draisma, S., van Hemert, A. M., Schoevers, R. A., Kupka, R. W., Beekman, A. T., and Penninx, B. W. (2021). Depressive and anxiety disorders in concert–a synthesis of findings on comorbidity in the nesda study. *Journal of Affective Disorders*, 284:85–97.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie*, 20(4):410–433.

Tolchinsky, A. (2013). *The Development of a Self-Report Questionnaire to Measure Problematic Video Game Play and its Relationship to Other Psychological Phenomena*. PhD thesis, University of South Florida.

Tolin, D. F., Woods, C. M., and Abramowitz, J. S. (2003). Relationship between obsessive beliefs and obsessive-compulsive symptoms. *Cognitive Therapy and Research*, 27(6):657–669.

Tom, S. M., Fox, C. R., Trepel, C., and Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, pages 515–518.

Tversky, A. and Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., and De Martino, B. (2017). Compulsivity reveals a novel dissociation between action and confidence. *Neuron*, 96(2):348–354.e4.

van den Bos, R., Homberg, J., and de Visser, L. (2013). A critical review of sex differences in decision-making tasks: Focus on the iowa gambling task. *Behavioural Brain Research*, 238:95–108.

van den Bos, R., Taris, R., Scheppink, B., de Haan, L., and Verster, J. C. (2014). Salivary cortisol and alpha-amylase levels during an assessment procedure correlate differently with risk-taking measures in male and female police recruits. *Frontiers in Behavioral Neuroscience*, 7.

van der Walt, W. H. and Wyndham, C. H. (1973). An equation for prediction of energy expenditure of walking and running. *Journal of applied physiology*, 34(5):559–563.

van Meurs, B., Wiggert, N., Wicker, I., and Lissek, S. (2014). Maladaptive behavioral consequences of conditioned fear-generalization: A pronounced, yet sparsely studied, feature of anxiety pathology. *Behaviour Research and Therapy*, 57:29–37.

van Overveld, W. J. M., de Jong, P. J., Peters, M. L., Cavanagh, K., and Davey, G. C. L. (2006). Disgust propensity and disgust sensitivity: Separate constructs that are differentially related to specific fears. *Personality and Individual Differences*, 41(7):1241–1252.

Videlier, M., Cornette, R., Bonneaud, C., and Herrel, A. (2015). Sexual differences in exploration behavior in xenopus tropicalis? *Journal of Experimental Biology*, 218(11):1733–1739.

Vieider, F. M., Chmura, T., and Martinsson, P. (2012). Risk attitudes, development, and growth: Macroeconomic evidence from experiments in 30 countries. Technical report. Standard-Nutzungsbedingungen.

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R. N., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., and Bullmore, E. T. (2015). Disorders of compulsivity: a common bias towards learning habits. *Molecular Psychiatry*, 20:345–352.

Vîslă, A., Stadelmann, C., Watkins, E., Zinbarg, R. E., and Flückiger, C. (2022). The relation between worry and mental health in nonclinical population and individuals with anxiety and depressive disorders: A meta-analysis. *Cognitive Therapy and Research*, 46:480–501.

Walasek, L., Mullett, T. L., and Stewart, N. (2018). A meta-analysis of loss aversion in risky contexts. *SSRN Electronic Journal*.

Walz, N., Mühlberger, A., and Pauli, P. (2016). A human open field test reveals thigmotaxis related to agoraphobic fear. *Biological Psychiatry*, 80(5):390–397.

Wang, S. W., Filiba, M., and Camerer, C. F. (2010). Dynamically optimized sequential experimentation (dose) for estimating economic preference parameters. *NBER Working Paper*, pages 1–41.

Wang, X. J. and Krystal, J. H. (2014). Computational psychiatry. *Neuron*, 84:638–654.

Watkins, E. R. and Roberts, H. (2020). Reflecting on rumination: Consequences, causes, mechanisms and treatment of rumination. *Behaviour Research and Therapy*, 127.

Watts, A. L., Greene, A. L., Bonifay, W., and Fried, E. I. (2023). A critical evaluation of the p-factor literature. *Nature Reviews Psychology*.

Weafer, J., Baggott, M. J., and De Wit, H. (2013). Test-retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention. *Experimental and Clinical Psychopharmacology*, 21(6):475–481.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., and Schnurr, P. P. (2013). The ptsd checklist for dsm-5 (pcl-5) – lec-5 and extended criterion a [measurement instrument].

Weber, E. U., Blais, A. R., and Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4):263–290.

Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: A review. *Frontiers in Psychology*, 10:1–19.

Weller, J. A., Levin, I. P., and Bechara, A. (2010). Do individual differences in iowa gambling task performance predict adaptive decision making for risky gains and losses? *Journal of Clinical and Experimental Neuropsychology*, 32(2):141–150.

Westen, D. and Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 69:875–899.

White, T. L., Lejuez, C. W., and de Wit, H. (2008). Test-retest characteristics of the balloon analogue risk task (bart). *Experimental and Clinical Psychopharmacology*, 16(6):565–570.

Wise, T. and Dolan, R. J. (2020). Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nature Communications*, 11(1):1–13.

Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B., and Marquand, A. F. (2015). From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience & Biobehavioral Reviews*, 57:328–349.

Wolpe, J. and Lang, P. J. (1964). A fear survey schedule for use in behaviour therapy. *Behaviour Research and Therapy*, 2(1):27–30.

Xu, P., Dam, N. T. V., van Tol, M. J., Shen, X., Cui, Z., Gu, R., Qin, S., Aleman, A., Fan, J., and jia Luo, Y. (2020). Amygdala–prefrontal connectivity modulates loss aversion bias in anxious individuals. *NeuroImage*, 218.

Yamamori, Y. and Robinson, O. J. (2023). Computational perspectives on human fear and anxiety. *Neuroscience & Biobehavioral Reviews*, 144:104959.

Yilmaz Balban, M., Cafaro, E., Saue-Fletcher, L., Washington, M. J., Bijanzadeh, M., Lee, A. M., Chang, E. F., and Huberman, A. D. (2021). Human responses to visually evoked threat. *Current Biology*, 31(3):601–612.e3.

Yook, K., Kim, K. H., Suh, S. Y., and Lee, K. S. (2010). Intolerance of uncertainty, worry, and rumination in major depressive disorder and generalized anxiety disorder. *Journal of Anxiety Disorders*, 24:623–628.

Zorowitz, S., Solis, J., Niv, Y., and Bennett, D. (2023). Inattentive responding can induce spurious associations between task behaviour and symptom measures. *Nature Human Behaviour*, 7:1667–1681.

Zsido, A. N., Arato, N., Inhof, O., Janszky, J., and Darnai, G. (2018). Short versions of two specific phobia measures: The snake and the spider questionnaires. *Journal of Anxiety Disorders*, 54:11–16.

Zuckerman, M. (1979). Sensation seeking and risk taking. In *Emotions in personality and psychopathology*, pages 161–197. Springer.

Zuckerman, M. (2007a). *Sensation seeking and risky behavior*. American Psychological Association.

Zuckerman, M. (2007b). The sensation seeking scale v (sss-v): Still reliable and valid. *Personality and Individual Differences*, 43(5):1303–1305.

Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12(1):63–70.

Öhman, A. (2008). Fear and anxiety: Overlaps and dissociations. *Handbook of Emotions*, pages 709–729.

A

Study 1 Supplements

A.1 Supplemental Methods

A.1.1 Correlation between Risk Preference Estimates

The three baseline estimates were not correlated with each other across participants, suggesting that they are underlined by distinct processes (λ and ρ: $r = -0.11$, $p = 0.40$; λ and μ: $r = 0.04$, $p = 0.78$; μ and ρ: $r = 0.03$, $p = 0.79$). Noteworthily, risk aversion (ρ) and choice consistency (μ) were correlated after the worry induction ($r = -0.31$, $p < 0.05$). This correlation was mainly driven by HC ($r = -0.45$, $p < 0.05$) and not by MDD+GA ($r = -0.29$, $p = 0.27$) or MDD-GA ($r = -0.14$, $p = 0.51$). There was no correlation post induction between ρ and λ ($r = -0.11$, $p = 0.39$) or λ and μ ($r = -0.04$, $p = 0.73$).

A.1.2 Power Analyses

We conducted a post-hoc power analysis using G*Power (version 3.1.9.7), to calculate achieved power based on α, sample size, and effect size. These analyses were informed by the effect size reported in existing literature, specifically referencing Charpentier et al. (2017). They reported higher risk aversion in pathologically anxious individuals compared to controls ($t(46) = 2.49$, $p = 0.016$, Cohen's $d = 0.72$). Applying this effect size to our sample sizes ($N_{\text{MDD+GA}} = 16$, $N_{\text{MDD-GA}} = 24$) and setting an alpha error probability at 5%, we estimated a post hoc power of approximately 71% for detecting a one-tailed difference between the two independent means.

We also conducted a sensitivity analysis for the results of the 2x2 repeated-measures ANOVA. The parameters were set as follows: an alpha error probability of 5%, power of 95%, a total sample size of 40 (with 2 groups), 2 measurements, a correlation among repeated measures of 0.85 (calculated from the data), and a nonsphericity correction factor of 1. The critical F value is 4.10, and the minimum effect size f is 0.16 which we exceeded as our main interaction of interest.

These estimations allow contextualizing the power of our results, particularly regarding the lack of baseline differences between MDD+GA and MDD-GA groups. However, it is important to note that this power estimate is based on the external validity of the effect size from Charpentier et al. (2017) and may not fully represent our study context. Additionally, we acknowledge that our use of post-hoc rather than a priori power analysis, necessitated by external constraints preventing further data collection, has its limitations.

A.2 Supplemental Results

A.2.1 Propensity to Gamble

Across all participants, the propensity to choose the gamble instead of the certain reward was on average 50% ± 2% (baseline) and 47% ± 2% (post-induction). There were no significant group differences at baseline (cf. Figure A.1). We found a trend-like interaction between time points (baseline and induction) and patient groups ($F(1, 38) = 3.97$, $p = 0.054$). Post-hoc t-tests revealed that MDD+GA gambled significantly less after the worry induction than MDD-GA ($t = -2.51$, $p < 0.05$). There was no difference between MDD and HC ($t = -0.65$, $p = 0.52$). When separating trial types, we found a trend-like interaction between time points and patient groups in gain-only trials ($F(1, 38) = 3.06$, $p = 0.089$) such that MDD+GA gambled significantly less after the worry induction than MDD-GA ($t = -2.69$, $p < 0.01$). This was not the case for mixed-only trials where no interaction and no difference between groups were found.

Finally, when adding the type of gamble (i.e. mixed and gain only) as within-subjects factor in addition to time points and patient group, a trend level interaction arose between type of gamble and patient groups but not time point ($F(1, 114) = 3.42$, $p = 0.067$) such that MDD+GA patients gambled significantly less than MDD-GA patients on gain-only decisions ($t = -3.11$, $p < 0.001$) but the propensity to gamble on mixed gamble trials did not differ between patient groups ($t = -1.2$,

$p = 0.24$).

A.2.2 Propensity to Gamble in Gain Trials

When adding the most distinctive anxiety scores (HAM-A) as a within-subjects factor, a significant interaction between timepoint and anxiety in the propensity to gamble was revealed in gain-only trials ($t(35) = -2.08$, $p < 0.05$) and the interaction between timepoint and patient group was no longer significant ($t(35) = -0.20$, $p = 0.84$). No other effect was significant.

However, when adding the most distinctive depressive symptom score (HAM-D) as a within-subjects factor in addition to anxiety, condition, and timepoint, the interaction between timepoint and anxiety was no longer significant ($t(28) = -1.08$, $p = 0.29$). There was also no significant effect when only HAM-D was added as a within-subjects factor.

A.2.3 Effect of Anxiety on the Propensity to Gamble

The difference in the propensity to gamble in gain-only trials from baseline to the post-induction was correlated with HAM-A ($r^2 = 0.35$), BDI ($r^2 = 0.30$), and PSWQ ($r^2 = 0.28$) in MDD+GA patients but not in MDD-GA or HC (both $r^2 < 0.10$). Given HAM-A's strong positive correlations with depressive scores (e.g., HAM-A and HAM-D: $r^2 = 0.44$; cf. Figure A.3), we controlled for them using partial correlations. The correlation with HAM-A was still strong even when controlling for depressive scores as measured by MADRS ($r^2 = 0.31$), BDI ($r^2 = 0.18$), and HAM-D ($r^2 = 0.19$). Inversely, the correlation with BDI was less strong when controlling for anxiety as measured by HAM-A ($r^2 = 0.12$).

A.2.4 Replication of Charpentier et al. (2017)

We fully replicated previous findings and found enhanced risk aversion in MDD+GA compared to MDD-GA (MDD+GA: 0.87 vs. MDD-GA: 0.95, $t = -2.36$, $p < 0.05$) and to HC (MDD+GA: 0.87 vs. HC: 0.94, $t = -2.21$, $p < 0.05$) but not between MDD and HC (MDD: 0.92 vs. HC: 0.94, $t = -0.76$, $p = 0.45$). Additionally, there was no difference between groups in loss aversion (MDD+GA: 1.77 vs. MDD-GA: 1.62, $t = 0.58$, $p = 0.56$; MDD+GA: 1.77 vs. HC: 1.58, $t = 0.7$, $p = 0.48$) or choice consistency (MDD+GA: 1.68 vs. MDD-GA: 1.44, $t = 1.26$, $p = 0.21$; MDD+GA: 1.68 vs. HC: 1.68, $t = 0$, $p = 1$).

Regarding the propensity to gamble, we conceptually replicate their findings. Where they found a trend-level difference, we found a significant group difference such that MDD+GA patients gambled less than MDD-GA (MDD+GA: 0.4 vs. MDD-GA: 0.52, $t = -2.71$, $p < 0.01$) and HC (MDD+GA: 0.4 vs. HC: 0.51, $t = -2.66$, $p < 0.01$). There was no difference between MDD and HC (MDD: 0.47 vs. HC: 0.51, $t = -1.14$, $p = 0.26$). When adding the type of gamble (i.e. mixed and gain only) as a within-subjects factor, we only found a trend-like interaction between trial type and patient group ($F(1, 118) = 3.50$, $p = 0.064$), and no interaction when comparing MDD+GA and HC ($F(1, 115) = 1.53$, $p = 0.22$). This only partially replicates their significant interaction when comparing anxiety patients with HC. However, we also found that MDD+GA patients gambled significantly less than MDD-GA patients (MDD+GA: 0.36 vs. MDD-GA: 0.58, $t = -3.11$, $p < 0.005$) and HC (MDD+GA: 0.36 vs. HC: 0.55, $t = -2.79$, $p < 0.01$) on gain-only decisions but not on mixed gamble trials (MDD+GA: 0.44 vs. MDD-GA: 0.51, $t = -1.2$, $p = 0.24$; MDD+GA: 0.44 vs. HC: 0.53, $t = -1.42$, $p = 0.16$).

A.3 Supplemental Tables

Table A.1: Comorbidities for each group acquired through the Structured Clinical Interview for DSM-IV (SKID-I and SKID-I). In parentheses is the DSM-IV code of each diagnosis.

Diagnosis	MDD, Single episode, Severe, Without psychotic features (296.23)	MDD, Single episode, Severe, With psychotic features (296.24)	MDD, Recurrent, Moderate (296.32)	MDD, Recurrent, Severe, Without psychotic features (296.33)	MDD, Recurrent, Severe, With psychotic features (296.34)	Adjustment disorder with depressed mood (309.0)
N/A	MDD+GA: 3, MDD-GA: 9	MDD+GA: 1, MDD-GA: 1		MDD+GA: 5, MDD-GA: 7	MDD+GA: 1, MDD-GA: 2	HC: 1
Generalized anxiety disorder (300.02)	MDD+GA: 1			MDD+GA: 1		
Dysthymic Disorder (300.4)				MDD+GA: 1		
Cannabis-related disorders, abuse (304.30)	MDD-GA: 1, MDD+GA: 1 (with 300.01 Panic disorder without agoraphobia)					
Panic disorder with agoraphobia (300.21)			MDD-GA: 1		MDD+GA: 1 (with 300.4 Dysthymic disorder)	
Social phobia (300.23)				MDD-GA: 1, MDD+GA: 1 (with 300.4 Dysthymic Disorder, 301.9 personality disorder 301.6 dependent personality disorder)		
Eating disorder (307.50)				MDD-GA: 1		
Feeding disorder of infancy (307.59)				MDD+GA: 1		

Table A.2: **Summary of results to replicate Charpentier et al., 2017.** In our study, we have two control groups (healthy control and depressed control corresponding to columns 3 and 4 respectively).
Anx: Anxious (MDD+GA in the current study), HC: healthy controls, Dep: Depressed (MDD-GA in the current study).

Findings	Charpentier et al., 2017 ($N_{\text{Anx}} = 25$; $N_{\text{HC}} = 23$)	Current Study ($N_{\text{Anx}} = 16$; $N_{\text{HC}} = 23$)	Current Study ($N_{\text{Anx}} = 16$; $N_{\text{Dep}} = 25$)
Enhanced risk aversion	$t(46) = 2.5, p = 0.02$	$t(78) = 2.21, p = 0.03$	$t(78) = 2.36, p = 0.02$
Unchanged loss aversion	$t(46) = 0.14, p = 0.89$	$t(78) = 0.70, p = 0.48$	$t(78) = 0.58, p = 0.56$
Decreased propensity to gamble	$t(46) = -1.43, p = 0.16$	$t(78) = -2.66, p = 0.01$	$t(78) = -2.71, p = 0.008$
Gamble type by group interaction	$F(1, 46) = 5.20, p = 0.03$	$F(1, 115) = 1.53, p = 0.22$	$F(1, 118) = 3.50, p = 0.064$
Reduced gambling on gain-only trials	$t(46) = -2.73, p = 0.01$	$t(78) = -2.79, p = 0.007$	$t(78) = -3.11, p = 0.003$
Unchanged gambling on mixed gamble trials	$t(46) = -0.39, p = 0.70$	$t(78) = -1.42, p = 0.16$	$t(78) = -1.20, p = 0.24$

A.4 Supplemental Figures

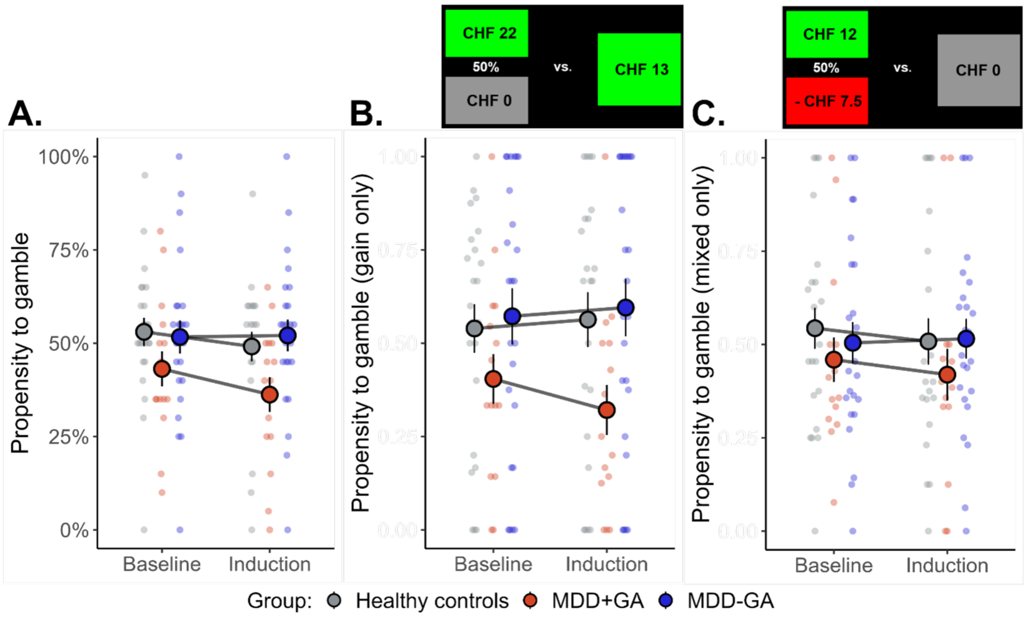

Figure A.1: **Reduced propensity to gamble in MDD+GA after worry induction compared to MDD-GA and healthy controls.** Mean estimates of propensity to gamble on all trials **(A)**, on gain-only trials **(B)**, and on mixed-only trials **(C)** plotted separately for MDD+GA, MDD-GA, and HC. There were no significant group differences in the propensity to gamble on all trials (MDD+GA: 43 vs. MDD-GA: 52, $t = -1.32$, $p = 0.20$; MDD: 48 vs. HC: 53, $t = -0.95$, $p = 0.35$), on gain-only trails (MDD+GA: 40 vs. MDD-GA: 57, $t = -1.67$, $p = 0.1$; MDD: 51 vs. HC: 54, $t = -0.41$, $p = 0.69$), and on mixed-only trials (MDD+GA: 46 vs. MDD-GA: 50, $t = -0.55$, $p = 0.58$; MDD: 49 vs. HC: 54, $t = -0.83$, $p = 0.41$).

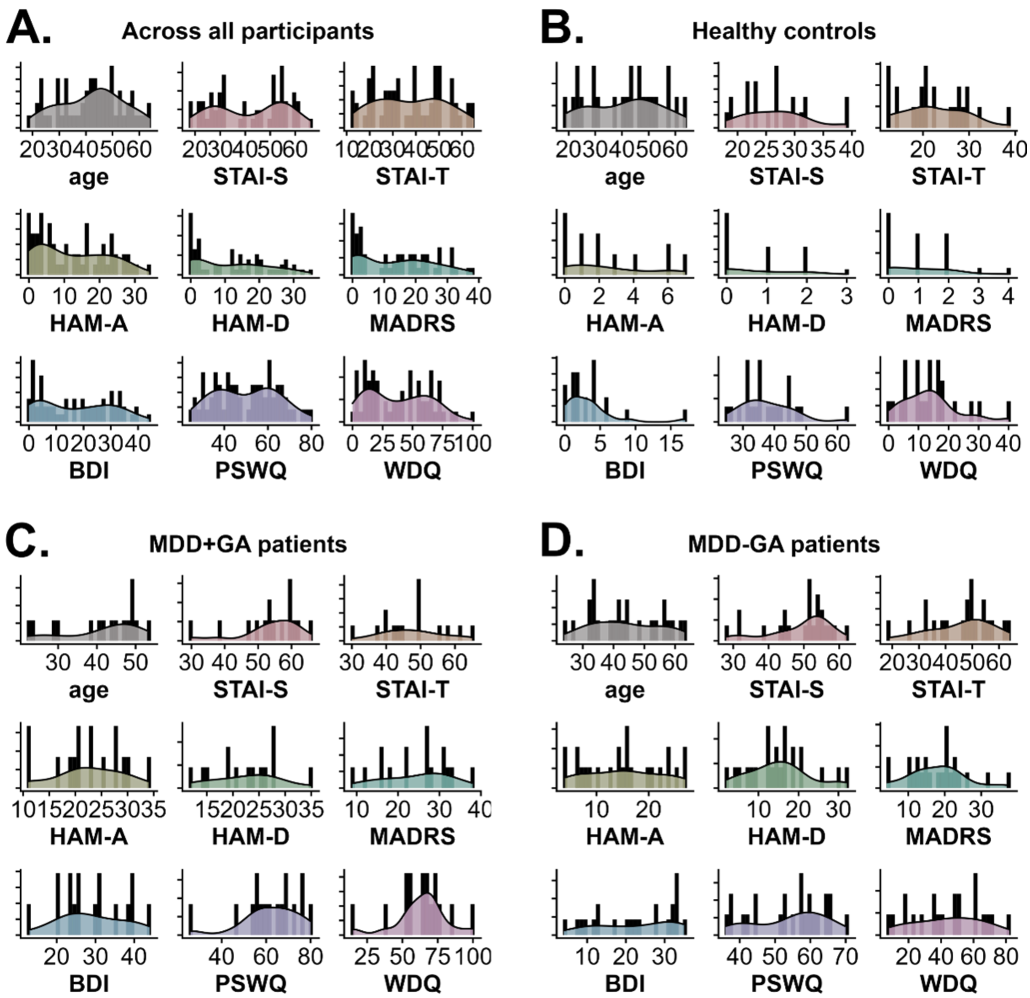

Figure A.2: **Distribution of questionnaire scores across all participants (A)**, healthy controls **(B)**, MDD+GA patients **(C)**, and MDD-GA patients **(D)**.

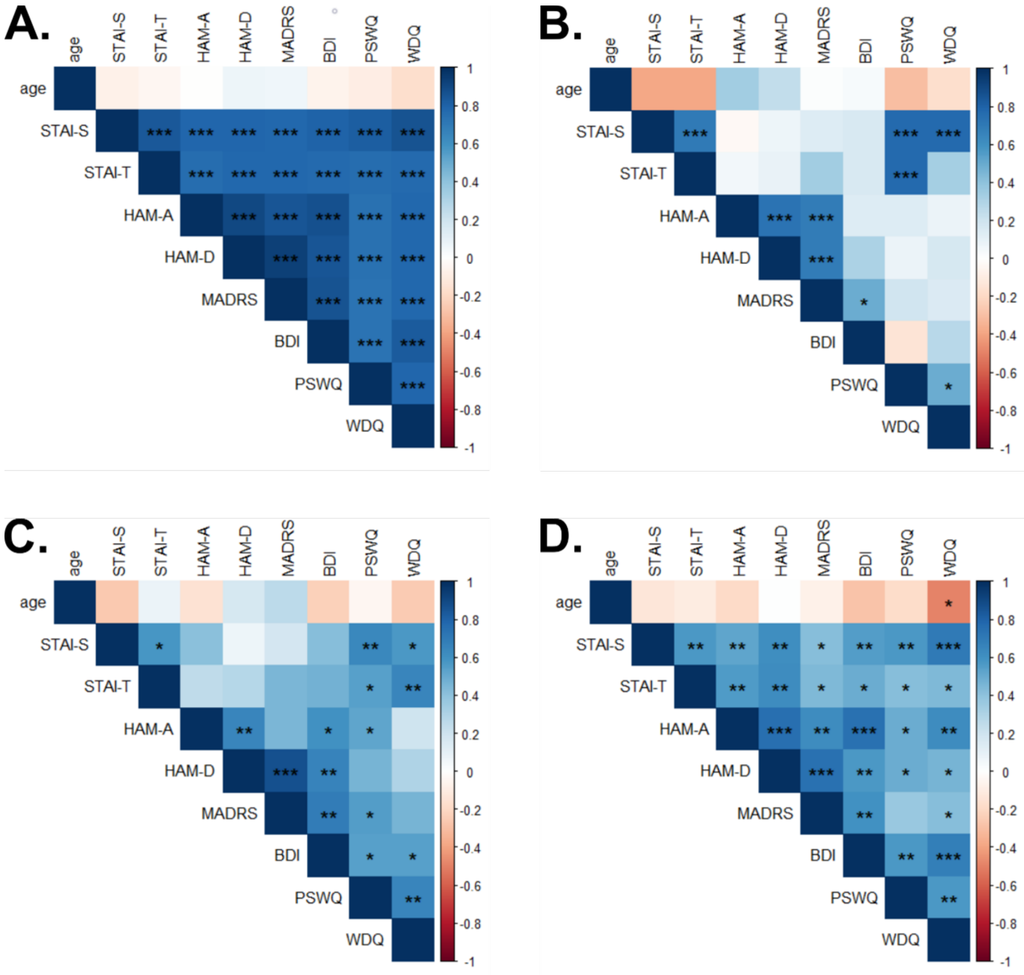

Figure A.3: **Correlation between questionnaire scores across all participants (A)**, healthy controls **(B)**, MDD+GA patients **(C)**, and MDD-GA patients **(D)**.

B

Study 2 Supplements

B.1 Supplemental Methods

B.1.1 Attention Checks

To identify inattentive participants, I included several attention checks within the questionnaires. Namely:

- "If you were paying attention to the previous questions, please select "A lot" as your answer." in OCI-R.
- "It is important that you carefully read the following options, choose "Yes" below." in the Schizo.
- "If you were focusing on the statements above, tick "A little"". in STICSA.

Each statement was formulated such that participants could not easily search for a recurrent keyword. Participants were told that missing the attention check questions would disqualify them from obtaining any potential bonuses they may have won during the task.

B.1.2 Pre-processing

Data were pre-processed according to a pre-defined pipeline. To avoid response latencies being biased by lapses or extreme values, they were only included if they fell into response windows of 150 ms $<$ approach latency $<$ 2,000 ms and 0 ms $<$ return latency $<$ 2,000 ms, as in previous work (Bach, 2015, 2017), and were then log-transformed.

B.1.3 Incorporating Three Additional Questionnaires into the Factor Analysis

I introduced three additional questionnaires—daringness (CADS; Lahey et al., 2010), sensation seeking (BSSS; Hoyle et al., 2002), and trait anxiety (STICSA; Ree et al., 2008)—into our factor analysis to examine their effects on the established 3-factor latent structure. The inclusion of these additional questionnaires did not alter the existing structure. Notably, I observed near-perfect correlations between the loadings and factor scores from the factor analysis with the original set of questionnaires as in Rouault et al. (2018) and Gillan et al. (2016), and from the analysis that included the new questionnaires (cf. Figures B.3-B.6).

All three questionnaires showed high loadings on the CIT factor and only marginal loadings on the AD and SW factors. While it was expected that sensation seeking and daringness would align with the CIT factor, trait anxiety was more surprising. However, several reasons might explain why trait anxiety was more strongly associated with CIT than AD. First, STICSA contains quite a few items that are linked to the intrusive nature of anxiety-related thoughts (e.g. item 10: "*I can't get some thought out of my mind.*", item 16: "*I keep busy to avoid uncomfortable thoughts.*", item 17: "*I cannot concentrate without irrelevant thoughts intruding.*" and item 19: "*I worry that I cannot control my thoughts as well as I would like to.*". Second, the AD factor might relate more closely to apathy and depression rather than anxiety. Indeed, generalized anxiety as assessed by the STAI and which mapped more strongly onto AD than CIT, correlated more with apathy (discovery: r=.70, confirmation: r=.55) and depression (discovery: r=.79, confirmation: r=.72) than with STICSA scores (discovery: r=.62, confirmation: r=.68; cf. Figure B.3).

B.1.4 Second-order Principal Component Analysis

To test whether the effect of the CIT symptom dimension could be explained by a general psychopathology factor broader than CIT, I tested the effect of a second-order factor. Second-order factor analysis can highlight higher-order, broad-spectrum factors by analyzing associations between the oblique first-order factors (Caspi et al., 2014; Lahey et al., 2021).

To this end, I conducted a Principal Component Analysis (PCA) using the `prcomp()` function from the `Stats` package on the factor scores from the combined sample. I then tested the effect of the first component of the resulting PCA on behavior. While this second-order principal component significantly predicted behavior, it did not explain behavior better than the CIT symptom dimension. Indeed, behavior only explained 13.49% variance in this second-order component, compared to 37.42% in CIT. In fact, it performed worse than 6 individual questionnaire scores, namely those assessing alcoholism, OCD, impulsivity, schizotypy, sensation seeking, and daringness (cf. Table B.3).

B.2 Supplemental Results

B.2.1 Elevated Anxiety Levels in the Online Sample

The absence of anxiety's effect cannot be attributed to low anxiety levels within the online sample. Indeed, our current online participants exhibited higher anxiety scores compared to those in a previous in-lab study conducted by Sporrer et al. (2023). Specifically, the in-lab study reported mean trait anxiety scores (Ree et al., 2008) of 27.93 (SD = 5.57) for the discovery sample and 29.80 (8.89) for the exploratory sample. In comparison, the online sample presented considerably higher scores, with means of 39.38 (15.68) and 48.99 (14.66), respectively.

B.2.2 Inattentive Responding and Data Validity

It is worth noting that the distribution of clinical scores differs significantly across waves in our study. Specifically, the confirmation sample exhibits more unusual patterns compared to the discovery sample. For instance, scales assessing OCD and Generalized Anxiety lack the expected positive skew, and the Alcoholism scale displays bimodal distributions, both of which deviate from typical findings. Such atypical distributions have been linked to inattentive responding in questionnaires.

Zorowitz et al. (2023) demonstrated that inattentive responding can produce spurious associations between task behavior and symptom measures. However, there are several factors indicating that in the present case, inattentive responding is unlikely to explain the lack of specificity in the pattern of cognitive deficits associated with CIT and other psychiatric questionnaires in our study.

First, Zorowitz et al. (2023) reported that nearly no significant (spurious) correlations emerged among symptom measures with more symmetric distributions. In our study, the strongest effects were observed in the discovery sample, where clinical scores followed typical patterns and were more symmetric (cf. Table 3.1, Figure B.2). Second, Zorowitz et al. (2023) noted a marked reduction in significant correlations after excluding inattentive responders, a result unlikely to be due to reduced statistical power. In our experiments, I employed conservative, pre-registered exclusion criteria, combining several task-behavior and self-report measures (cf. Methods 3.2). This approach, as noted by Zorowitz et al. (2023), is one of the best ways to prevent spurious correlations. Additionally, rigorous participant screening does not appear to introduce overcontrol bias, as inattentive responding is independent of psychopathology (Zorowitz et al., 2023). Finally, Zorowitz et al. (2023) highlighted that false-positive rates for spurious behavioral-symptom correlations actually increase with sample size (i.e. keeping the inattentive responders) due to an increase in measurement bias rather than measurement noise. In our study, when all participants were included in the analysis, the results of interest were smaller or insignificant compared to those excluding inattentive respondents. Based on these additional tests and the successful replication of results, our findings are unlikely to be driven by false-positive correlations due to inattentive responding.

B.2.3 Interindividual Differences in Subjective Prior Assumptions

To investigate the subjective prior assumption that the presence of tokens alerts the predator, participants completed a predator exposure task. Here, optimal behavior according to the task statistics was to make an exposure attempt early in the trial, regardless of token appearance. Interestingly, post hoc analysis on the combined sample shows that the opposite pattern is true for people with high IQ who tried to expose the robber more frequently after the token appeared ($\beta = -2.1, t(1,961) = -2.06, p < .05$). A one standard deviation increase in IQ score is linked to a 76.98 ms later approach. This suggests that people with high IQ might assume that the presence of tokens alerts the predator.

B.2.4 Interindividual Differences in Threat Memory

To further investigate which aspect of threat memory is influenced by factor scores, I extracted each participant's slope and intercept from a linear regression between the estimated catch rates and the threat levels, indicating threat memory precision and bias, respectively. I then used these regressors as dependent variables in another linear regression with the factor scores or demographics as predictors.

In this post-hoc exploratory analysis, CIT was associated with higher bias ($\beta = .21, t(997) = 6.70, p < .0001$) and lower precision ($\beta = -.08, t(997) = -2.54, p < .01$). The inverse was true with people with high IQ who had a lower bias ($\beta = -.12, t(1002) = -4.08, p < .001$) but a higher precision ($\beta = .09, t(1002) = 2.86, p < .005$).

I then split the data into subsets comprising of either participant scoring in the 25% top or the 75% bottom of CIT scores. I then repeated our analyses in these sub-samples. The 25% highest CIT scorers were not able to dissociate between the threats and reported similar catch rates ($F(2, 249) = 1.59, p > .05$) in contrast to the 75% lowest CIT scores who could ($F(2, 749) = 37.01, p < .0001$).

B.3 Supplemental Tables

Table B.1: **Demographic and questionnaire scores within each sample.** In parentheses is the standard deviation from the mean.

Variable	Discovery Sample	Confirmation Sample
N	315	690
Female	149	338
Age	36.40 (11.01)	33.41 (9.89)
Generalized Anxiety	42.21 (10.80)	44.79 (8.36)
Eating Disorders	14.12 (9.83)	17.45 (9.92)
Apathy	35.39 (9.15)	38.20 (6.95)
Alcoholism	9.71 (9.35)	17.16 (9.41)
Depression	40.32 (10.04)	44.71 (8.13)
OCD	25.32 (16.45)	35.19 (13.73)
Social Anxiety	54.60 (31.53)	66.13 (28.27)
IQ	7.40 (3.55)	6.55 (2.91)
Impulsivity	62.12 (12.90)	69.30 (10.32)
Schizotypy	15.01 (9.20)	21.01 (8.51)
Sensation seeking	2.97 (0.84)	3.31 (0.72)
Trait anxiety	39.38 (15.68)	48.99 (14.66)
Daringness	2.55 (0.79)	2.89 (0.55)

Table B.2: **Behavioral results within each sample.** In green are the data validation criteria that must be fulfilled to progress to the next analysis step. The GLMM and LMM included a 3 x 6 factorial design with threat level (low/medium/high) and potential loss (0-5 tokens). The models contained all possible polynomial terms, but I only reported linear contrasts for each factor or interaction. The p-values are not corrected for multiple comparisons and are presented as a heuristic guide only. In parentheses is the standard deviation from the beta mean.

Variable	Discovery Sample	Confirmation Sample
Approach choices		
Threat level	$\beta = -.358, (.02), t(1, 44964) = 209.56, p < .0001$	$\beta = -.173, (.02), t(1, 98925) = 85.24, p < .0001$
Potential loss	$\beta = -2.329\,(.04),\ t(1, 44964) = 3644.72,\ p < .0001$	$\beta = -1.003\,(.03),\ t(1, 98925) = 1418.03,\ p < .0001$
Interaction: Threat level x Potential loss	$\beta = .052\,(.06),\ t(1, 44964) = .69,\ p > .05$	$\beta = .018\,(.05),\ t(1, 98925) = .16,\ p > .05$
Approach latency		
Threat level	$\beta = .010\,(.00),\ t(1, 32333) = 11.8,\ p < .01$	$\beta = .003\,(.00),\ t(1, 80454) = 3.89,\ p < .05$
Potential loss	$\beta = .043\,(.00),\ t(1, 32333) = 104.41,\ p < .0001$	$\beta = .011\,(.00),\ t(1, 80454) = 32.81,\ p < .0001$
Interaction: Threat level x Potential loss	$\beta = .002\,(.01),\ t(1, 32333) = .08,\ p > .05$	$\beta = -.005\,(.00),\ t(1, 80454) = 1.9,\ p > .05$
Withdrawal latency		
Threat level	$\beta = -.047\,(.01),\ t(1, 22022) = 42.29,\ p < .0001$	$\beta = -.062\,(.00),\ t(1, 51285) = 173.98,\ p < .0001$
Potential loss	$\beta = -.023\,(.01),\ t(1, 22022) = 4.63,\ p < .05$	$\beta = -.012\,(.01),\ t(1, 51285) = 3.06,\ p > .05$
Interaction: Threat level x Potential loss	$\beta = -.02\,(.02),\ t(1, 22022) = 1.21,\ p > .05$	$\beta = .002\,(.01),\ t(1, 51285) = .04,\ p > .05$

Table B.3: **Effect of each questionnaire score (above black line) and symptom dimension (below black line) on approach choices and latencies.** The last three questionnaire scores (i.e. sensation seeking, trait anxiety, and daringness) in grey were not included in the factorial analysis to calculate the symptom dimensions. This table is based on the combined sample. The p-values are not corrected for multiple comparisons and are presented as a heuristic guide only. Cf. methods on how the explained variance was estimated.

Questionnaire	Approach choices	Approach Latencies	Explained variance (%)
Generalized Anxiety	$\beta = .11, (.05), t(1, 143657) = 4.08, p < .05$	$\beta = .02, (.01), t(1, 112602) = 8.82, p < .01$	2.13
Eating Disorders	$\beta = .32, (.07), t(1, 143800) = 22.97, p < .0001$	$\beta = .05, (.01), t(1, 112722) = 26.26, p < .0001$	5.57
Apathy	$\beta = .17, (.06), t(1, 143514) = 8.90, p < .01$	$\beta = .03, (.01), t(1, 112472) = 12.59, p < .001$	4.39
Alcoholism	$\beta = .39, (.06), t(1, 143369) = 45.86, p < .0001$	$\beta = .09, (.01), t(1, 112447) = 106.72, p < .0001$	17.46
Depression	$\beta = .25, (.05), t(1, 143371) = 22.75, p < .0001$	$\beta = .05, (.01), t(1, 112363) = 45.79, p < .0001$	8.94
OCD	$\beta = .37, (.06), t(1, 143657) = 42.96, p < .0001$	$\beta = .10, (.01), t(1, 112602) = 144.62, p < .0001$	21.05
Social Anxiety	N.s.	$\beta = .06, (.01), t(1, 112385) = 41.29, p < .0001$	4.76
Impulsivity	$\beta = .34, (.05), t(1, 143514) = 40.26, p < .0001$	$\beta = .07, (.01), t(1, 112472) = 83.68, p < .0001$	16.73
Schizotypy	$\beta = .42, (.06), t(1, 143514) = 56.02, p < .0001$	$\beta = .07, (.01), t(1, 112493) = 73.5, p < .0001$	14.63
Sensation seeking	$\beta = .43, (.06), t(1, 143371) = 59.92, p < .0001$	$\beta = .06, (.01), t(1, 112436) = 49.85, p < .0001$	15.06
Trait anxiety	$\beta = .34, (.06), t(1, 143219) = 35.60, p < .0001$	$\beta = .08, (.01), t(1, 112331) = 73.50, p < .0001$	12.64
Daringness	$\beta = .42, (.05), t(1, 143657) = 65.74, p < .0001$	$\beta = .08, (.01), t(1, 112665) = 112.07, p < .0001$	21.19
CIT	$\beta = .59, (.05), t(1, 143083) = 122.20, p < .0001$	$\beta = .13, (.01), t(1, 112208) = 251.89, p < .0001$	37.42

Table B.4: **Items from the daringness subscale of the CADS (Lahey et al., 2010).** Participants have to select of the options following these instructions "When you answer these questions, please think about the last 12 months and tick the box that you feel best describes you." The options are: "Not at all", "Just a little", "Pretty much/pretty often", and "Very much/very often"

Questions
1) Are you daring and adventurous?
2) Do you like rough games and sports?
3) Do you enjoy doing things that are risky or dangerous?
4) Do you like things that are exciting and loud?
5) Are you brave?

Table B.5: **Items from the Brief Sensation Seeking Scale (BSSS; Hoyle et al., 2002).** Participants have to select of the options following these instructions "Please read each statement carefully and select the option which best describes you. We are interested only in your likes or feelings, not how others feel about these things or how one is supposed to feel." The options are: "Strongly disagree", "Disagree", "Neither disagree nor agree", "Agree", and "Strongly agree".

Questions
1) I would like to explore strange places.
2) I get restless when I spend too much time at home.
3) I like to do frightening things.
4) I like wild parties.
5) I would like to take off on a trip with no pre-planned routes or timetables.
6) I prefer friends who are excitingly unpredictable.
7) I would like to try bungee jumping.
8) I would love to have new and exciting experiences, even if they are illegal.

B.4 Supplemental Figures

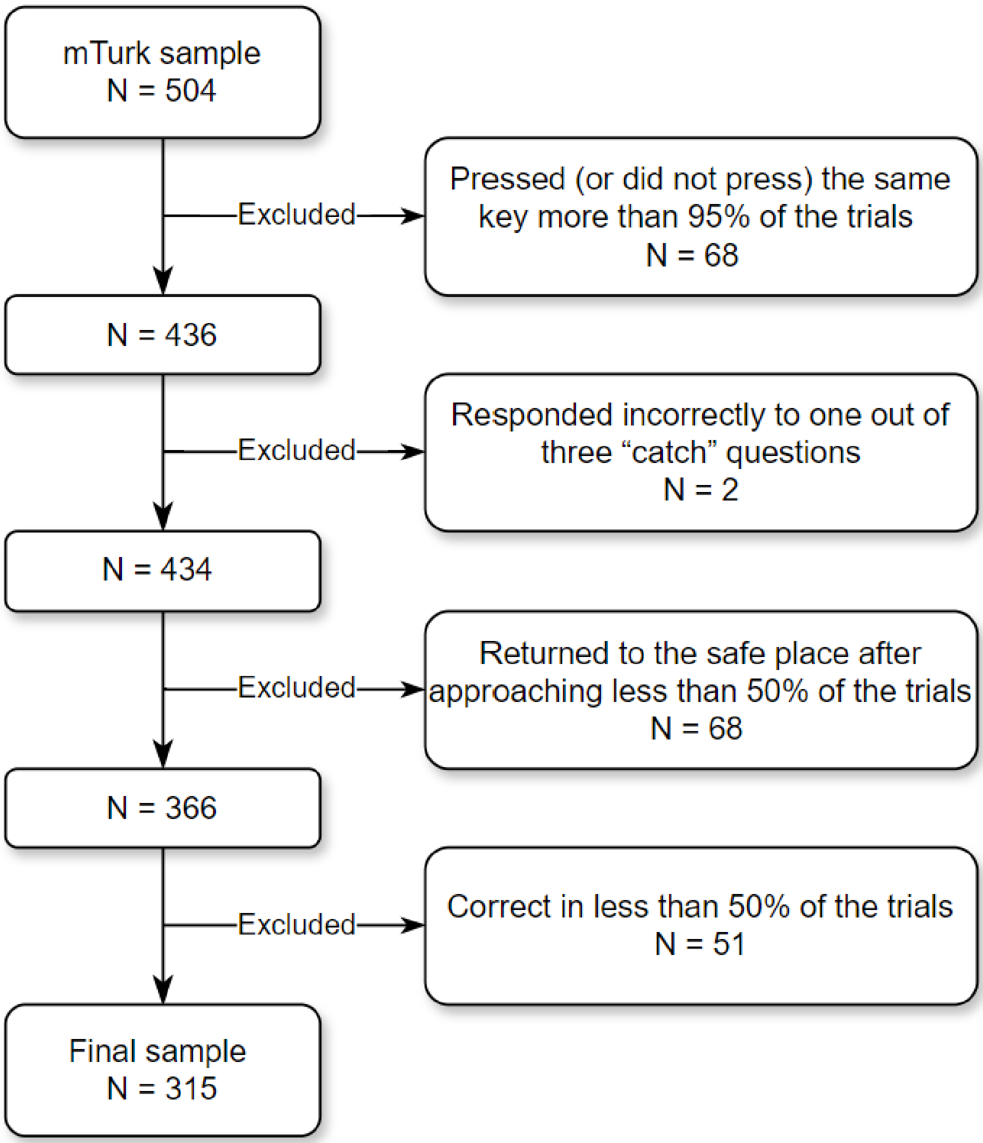

Figure B.1: **Exclusion flowchart for the first exploration experiment.** The same criteria were applied to the second confirmation experiment.

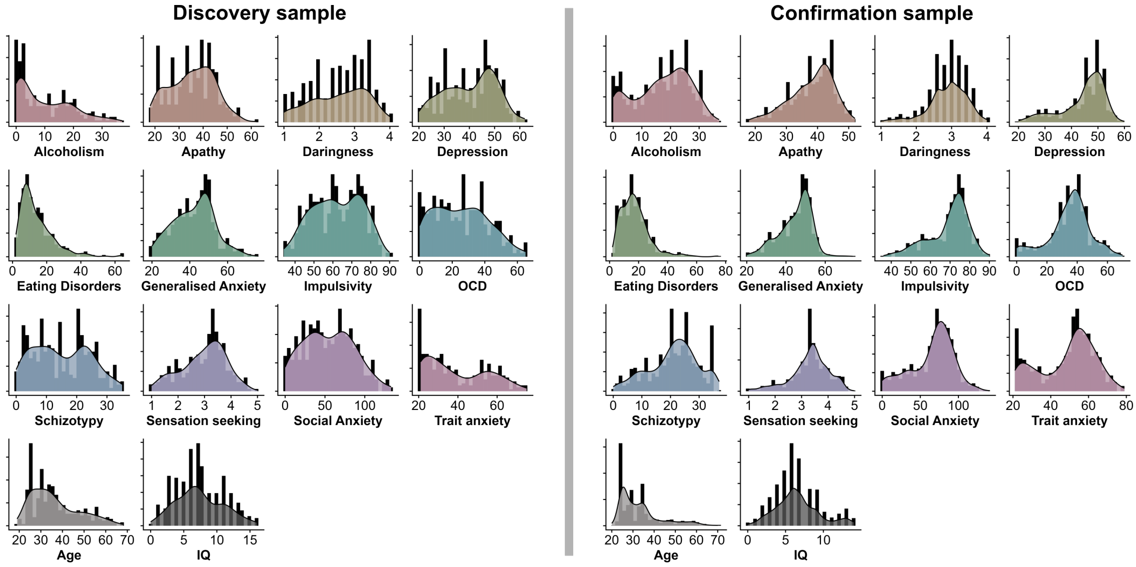

Figure B.2: **Distribution of questionnaire scores in both samples**.

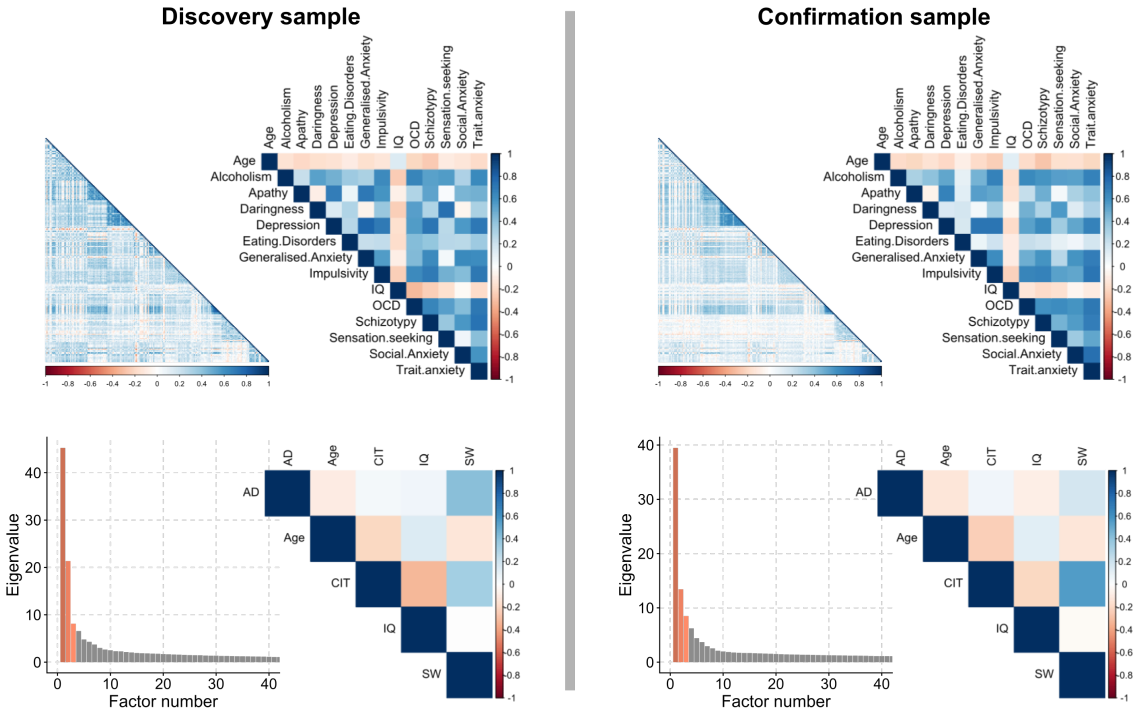

Figure B.3: **Correlation matrices between questionnaires in both samples**. Specifically, the correlation between questionnaire items (top left), questionnaire scores and demographics (top right), and the three factors and demographics (bottom right). In the bottom left are the eigenvalues from the factor analysis revealing the three-factor solution that best accounted for our data. The color scale indicates the correlation coefficient.

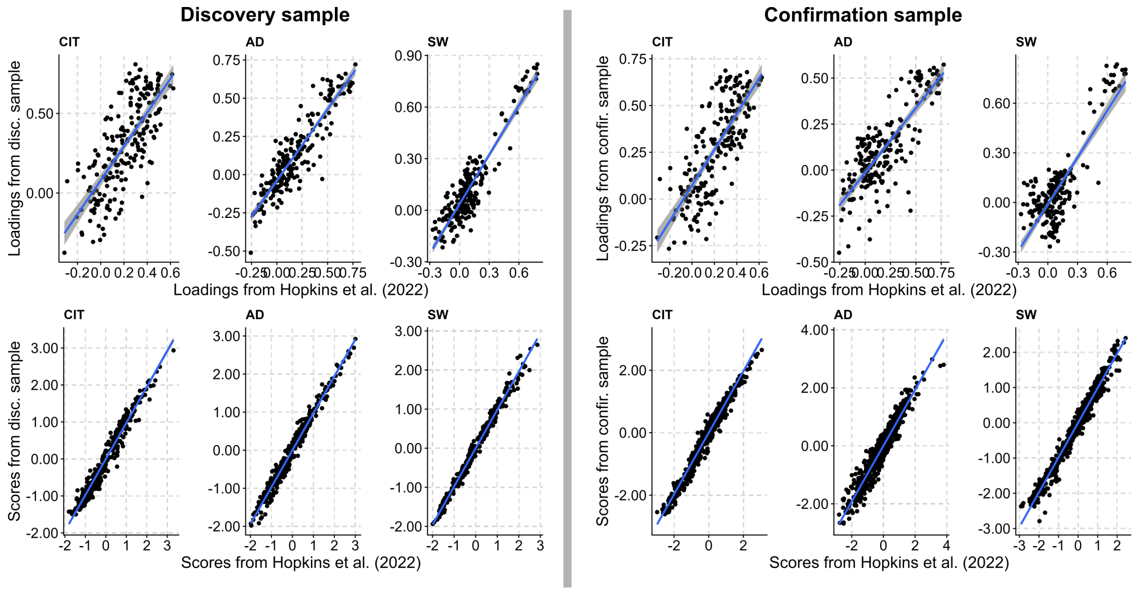

Figure B.4: **Correlation matrices between questionnaires in both samples**. Specifically, the correlation between questionnaire items (top left), questionnaire scores and demographics (top right), and the three factors and demographics (bottom right). In the bottom left are the eigenvalues from the factor analysis revealing the three-factor solution that best accounted for our data. The color scale indicates the correlation coefficient.

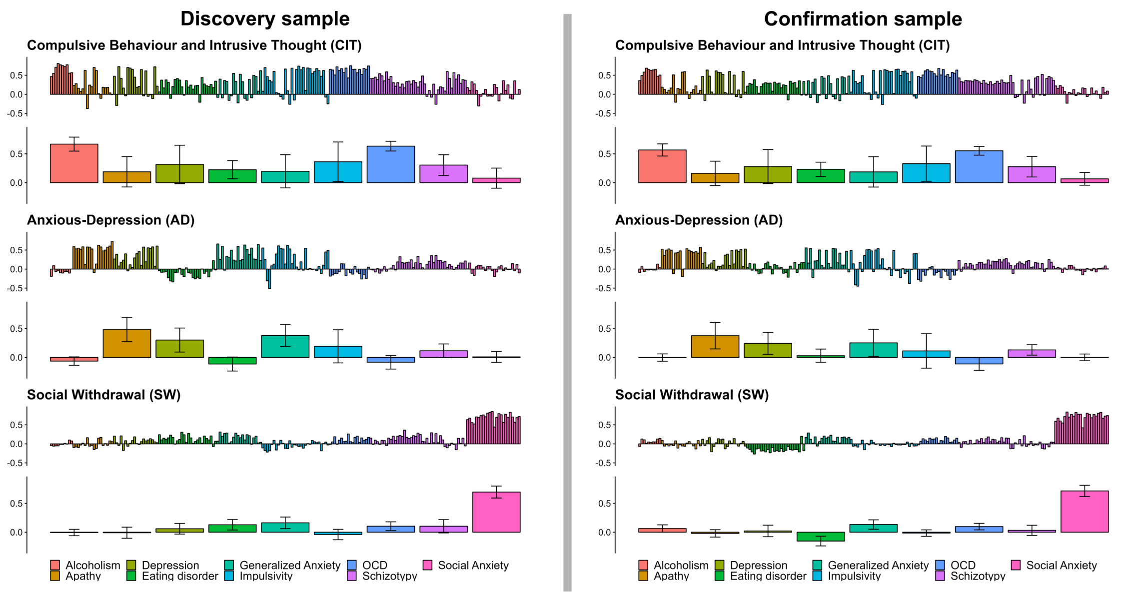

Figure B.5: **Questionnaire loadings onto each factor in both samples, color-coded by questionnaire**. The top figures of each factor detail the loadings of the individual questionnaire items, the bottom figures summarise the loadings at the questionnaire score level. The error bars represent the standard deviation of the mean over item loadings

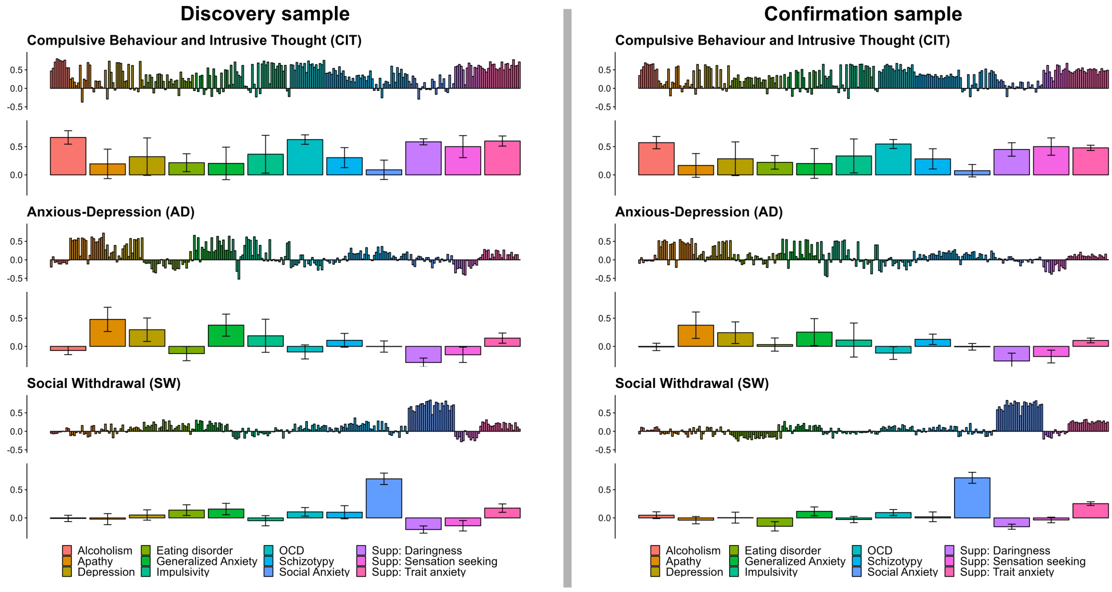

Figure B.6: **Loadings of all questionnaires (originals and additional) onto each factor in both samples**, color-coded by questionnaire. The top figures of each factor detail the loadings of the individual questionnaire items, the bottom figures summarise the loadings at the questionnaire score level. The error bars represent the standard deviation of the mean over item loadings.

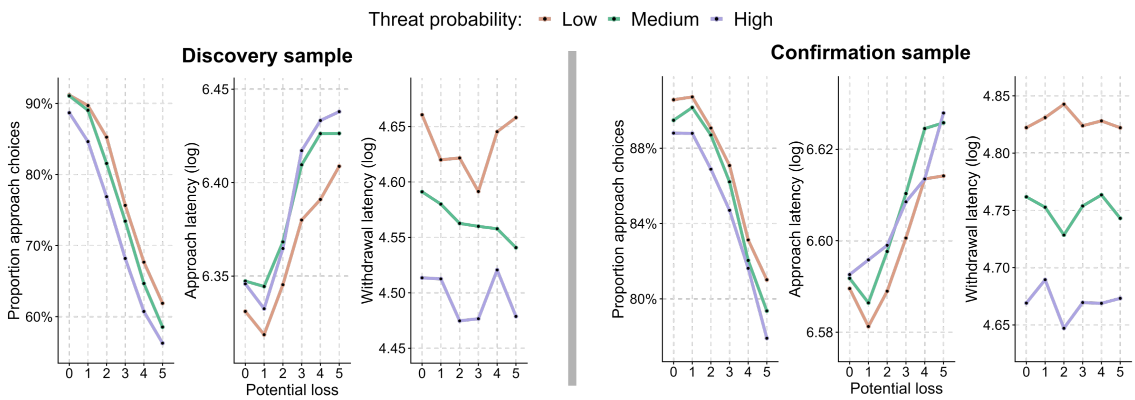

Figure B.7: **Behavioral results in both samples**. Proportion of approach-avoidance decisions, indexing passive avoidance (left), approach latency, indexing behavioral inhibition (center), and withdrawal latency (right).

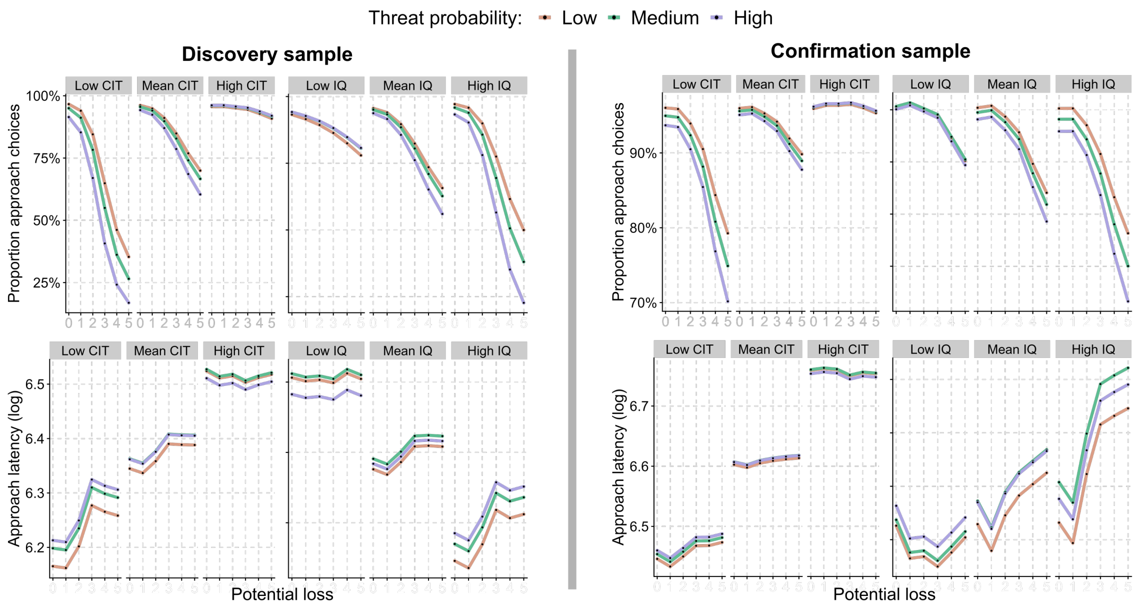

Figure B.8: **Estimated behavioral output according to main behavioral predictors in both samples**. Estimated marginal means of approach choice (top) and latency (bottom) depend on CIT symptom dimension scores (left) or IQ (right) while other predictors are kept fixed. *Low CIT/IQ: -1.5, Mean CIT/IQ: 0, and High CIT/IQ: +1.5*

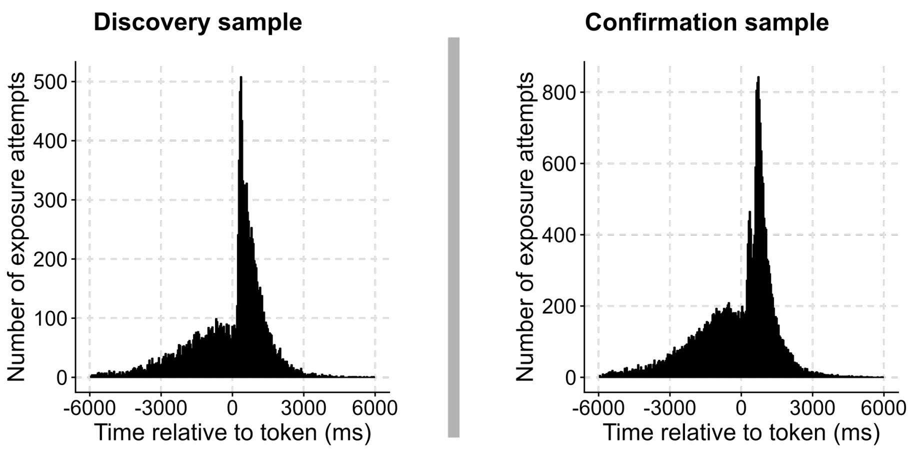

Figure B.9: **Time of threat exposure attempts relative to token appearance in both samples.**

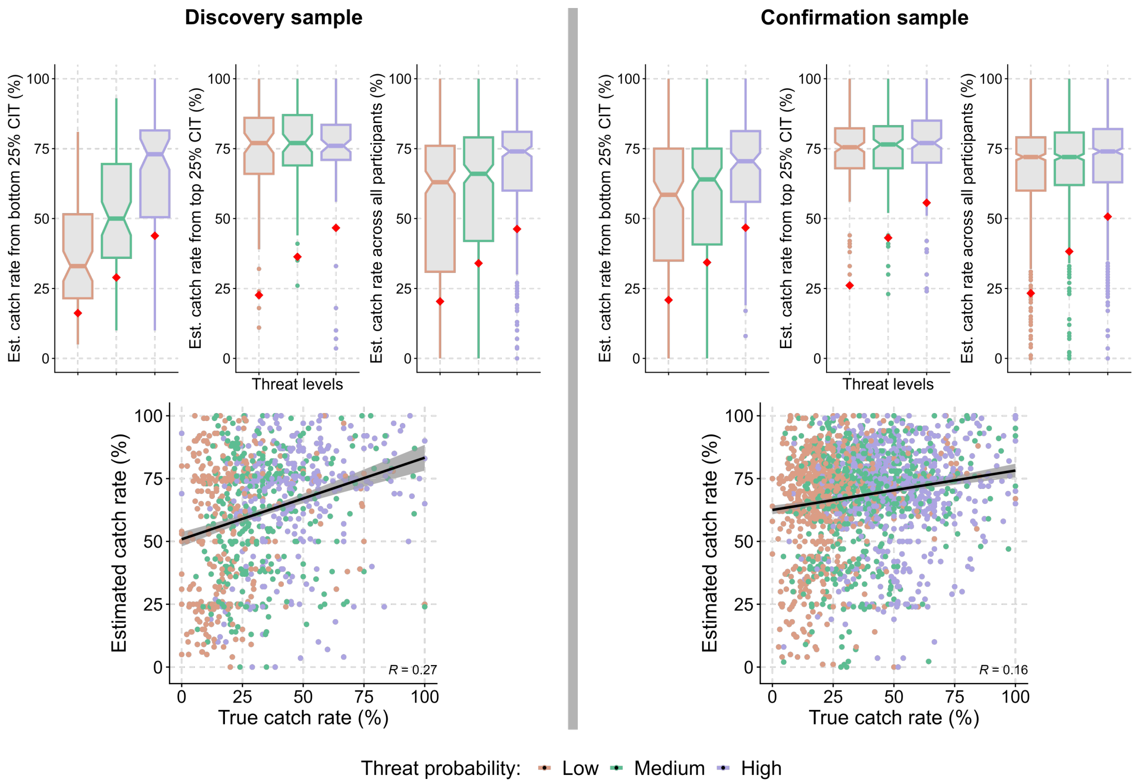

Figure B.10: **Biased reported threat memory in both samples**. Across participants, the estimated catch rates depended on the true catch rate which had to be learned during the experiment (bottom). CIT is linked to biased threat memory such that the top 25% CIT scorers (center top) cannot distinguish between different threat levels and overestimate their probabilities. While the bottom 25% CIT scores (left top) and all participants (right top) distinguished the threats better, they still overestimated the threat probabilities. Actual threat rates for each level are denoted by red diamonds. Est.: estimated.

C

Study 3 Supplements

C.1 Supplemental Methods

C.1.1 Equipment

We used an HTC Vive Pro Eye HMD, Windows PC with an Intel i7 9700K CPU and Nvidia RTX 2080Ti GPU. Vive controllers were held in each hand, and Vive Trackers were attached to the waist and feet to allow for real-time body tracking. The VR headsets included built-in eye tracking and a microphone positioned on the underside of the HMD.

To calibrate the eye tracking before starting E2, participants performed the HTC VIVE Pro Eye built-in calibration provided by the SRanipal software development kit (SDK). First, this procedure assists the participant in properly adjusting the head-mounted display for a snug fit and fine-tuning the lenses to accommodate their specific inter-pupillary distance (i.e. distance between the centers of eye pupils). Subsequently, the participant is presented with a series of five calibration positions to focus on consecutively ("track the dot"). Once all five positions have been fixated upon, the procedure ends. The manufacturer reports an accuracy of 0.5° - 1.1° (Corporation, 2023) which has been confirmed in a recent study that found an accuracy of 1.10° using real-world data (Schuetz and Fiehler, 2022).

C.1.2 Questionnaire Selection

We selected these scales because they were short and easy to fill in, have suitable psychometric properties and have been used in previous research.

Fear

To assess diverse fears, we selected the widely used FSS-III (Wolpe and Lang, 1964). Its two main criticisms relate to its poor discriminant validity between patients with specific anxiety disorders (Beck et al., 1998) and the conflation of fear and anxiety. Whereas the first criticism appeared less relevant in the context of the present study, the second might raise questions. Fear is often seen as a short-lived emotion which motivates escape from an imminent threat, and anxiety as a longer response to an ambiguous or uncertain threat which might happen in the future (Adolphs, 2013; Steimer, 2002; Öhman, 2008). The Situated Fear Questionnaire (SFQ) was developed to better distinguish those two concepts (Campbell et al., 2016). However, we are not aware of investigations on the factorial structure of the SFQ or extensive validation data, which is why it was not retained for the present study.

Animal phobias

Specific animal phobias are globally the most frequent mental illness (Steel et al., 2014). Snakes and spiders are especially potent in eliciting strong negative emotions, even in non-clinical populations (Polák et al., 2020). To measure the specific phobia related to snakes, we selected the SNAQ-12 (Zsido et al., 2018) which is a short version of the well-known Snake Questionnaire (SNAQ; Klorman et al., 1974). While reducing its length from 30 to 12 items, it retains good psychometric values such as internal consistency ($\alpha = .88$) and provides a cut-off score with an optimal balance between sensitivity and specificity. Thus, participants scoring above 8 should be considered potentially snake phobic. Similarly, to measure the specific phobia of spiders, we utilized the SPQ-12 (Zsido et al., 2018) which is a short version of the Spider Questionnaire (SPQ; Klorman et al., 1974). It also reduces the length from 31 to 12 items and retains good psychometric values ($\alpha = .9$) with a cut-off at 7 which would suggest a risk of developing spider phobia. In contrast to FSS, both scales have a high discriminant validity between phobic and non-phobic individuals (Zsido et al., 2018). Furthermore, they also have very good test-retest reliability. We did not include specific scales related to other animals as they have been found to be highly correlated with the animal sub-scale from the FSS-III (Matchett and Davey, 1991).

Disgust

The Disgust Propensity and Sensitivity Scale Revised (DPSS-R; van Overveld et al., 2006) based on the original DPSS (Cavanagh and Davey, 2000) aims to assess a general tendency to respond with disgust to any given situation by measuring the frequency of disgust experiences (i.e. disgust propensity) and the emotional impact of disgust experiences (i.e. disgust sensitivity). Furthermore, the DPSS-R has good predictive validity as it corresponds well with disgust-induced avoidance in behavioral experiments (van Overveld et al., 2006). Thus, we decided to select the DPSS-12 (Fergus and Valentiner, 2009) which is a shorter version of the DPSS-R (van Overveld et al., 2006). While it reduces the number of items from 16 to 12, it provides even stronger internal consistency (α for disgust propensity = .83, α for disgust sensitivity = .80). It was replicated in both clinical and non-clinical samples and provides an index of the subject's tendency to feel disgust that generalized across contexts and is not limited to the three dimensions mentioned above (Goetz et al., 2013). Another one of the most commonly used questionnaires to assess disgust is the Disgust Scale Revised (DS-R; Olatunji et al., 2007) based on the original DS (Haidt et al., 1994). The DS-R is a 25-item scale that measures the participant's level of disgust about three core dimensions, including core disgust, animal reminder disgust, and contamination-based disgust. However, the DS-R measures disgust for specific elicitors including some unrealistic scenarios (e.g., "eating monkey meat"). It therefore does not give any indication of whether they appraise these experiences more negatively, which is why it was not retained for our experiments.

Anxiety

To measure anxiety, we selected the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA; Ree et al., 2008, 2000) that assesses cognitive (i.e., thought processes such as intrusive thoughts) and somatic dimensions (i.e., symptoms like sweating or trembling) to better discriminate between the different components of anxiety with good discriminant validity. This scale has been validated in clinical and nonclinical samples demonstrating its excellent internal consistency (α = .87), reliability, and construct validity as a purer measure of anxiety (Gros et al., 2007; Roberts et al., 2016). The STICSA was created (Ree et al., 2008, 2000) to counter the lack of discriminant validity between anxiety and depression but keep the theoretical formulation of state and trait anxiety of one of the most long-standing and popular measures to assess anxiety which is the State-Trait Anxiety Inventory (STAI; Spielberger, 1983). Indeed, an important point consistently reported in the literature regards its tendency to measure confounding depressive symptoms rather than anxiety and its inability to distinguish them (Bados et al., 2010; Bieling et al., 1998). The STICSA Trait scale correlated more highly with another measure of anxiety (DASS-A) than with the STAI Trait scale, and the STAI Trait correlated more highly with a measure of depression (DASS-D) than with the STICSA Trait (Gros et al., 2007).

Daringness

Daringness is the best predictor in a risky foraging task (Bach et al., 2020). This study was conducted in young people and used the daringness items from the Child and Adolescent Dispositions Scale (CADS; Lahey et al., 2010). For our current experiments, we sought to use an adequate scale for adults. A similar questionnaire often used in adults is the Sensation Seeking Scale (SSS-V; Zuckerman, 2007b) which includes four subscales: Thrill and Adventure-Seeking, Experience Seeking, Disinhibition, and Boredom Susceptibility. Based on these four dimensions, a short version of the SSS was created by reducing its length from 40 to 8 items. The Brief Sensation Seeking Scale (BSSS; Hoyle et al., 2002) includes two items representing each aspect of sensation seeking and maintains good psychometric characteristics ($\alpha = .76$).

Cybersickness

Cybersickness is part of virtual reality-induced symptoms and effects (VRISE) and is considered to be a subtype of motion sickness induced by immersion into virtual reality (Davis et al., 2014; Saredakis et al., 2020). There is limited consensus about the symptoms provoked by VR as the biological mechanisms are unknown. One of the most commonly used measures is the Simulator Sickness Questionnaire (SSQ; Lane and Kennedy, 1988) which was created to assess motion sickness for simulator systems. It indicates three constructs of simulator sickness: Nausea, Disorientation, and Oculomotor dimension, along with a second-order more general factor concerning total severity. Despite its extensive use, the SSQ has been widely criticized for its psychometric qualities and applicability in VR. In response, two scales were developed that selected those symptoms from SSQ that were found to be most relevant for VR (Sevinc and Berkman, 2020). First, based on a factor analysis-based method, the Virtual Reality Sickness Questionnaire (VRSQ) includes nine symptoms from the original SSQ to indicate Oculomotor and Disorientation constructs (Kim et al., 2018). Secondly, based on an item-response theory approach, the Cybersickness Questionnaire (CSQ-VR) retains nine symptoms from the SSQ in two factors: Dizziness and Difficulty in Focusing, and uses a scoring method based on item weights (Kourtesis et al., 2023). While both subscales have equally good psychometric qualities, there is so far limited work using these scales. This is why we decided to include all the items from the SSQ, which allows computing both subscales.

Cyber sickness and motion sickness susceptibility are subject to large variability but older age and being male act as protective factors. To monitor predictors of potential adverse effects in the VR, we included a short version of the Motion Sickness Susceptibility Questionnaire (MSSQ; Golding, 1998, 2006). This scale assesses previous experiences of motion sickness in different contexts (e.g., cars, aircraft, funfair rides) during childhood and adulthood. We did not use this scale as an exclusion criterion because evidence is still lacking regarding its predictive validity in VR experiments. Existing data on MSSQ's predictive validity were derived from laboratory motion experiments without VR (Golding, 2006).

As video game usage might impact the susceptibility of experiencing cybersickness, we sought to include a questionnaire assessing video game habits. The Video Game Usage Questionnaire was created to measure the weekly amount of video game play whether it is the average number of hours played, the average duration of each session, etc. (Tolchinsky, 2013).

C.1.3 Threat Selection

We selected the threats based on six principles.

First, we sought to sample from a variety of threat classes, as there is a suggestion they might engage different action-selection mechanisms. We considered five broad and overlapping threat classes: conspecific, predatory, defensive, disgust-relevant, and inanimate. There are several reasons to motivate this wide range of threats. First, there is some evidence in animals that suggests different classes of threats engage different neural mechanisms. For example, learning to predict predatory or conspecific threat from conditioned stimuli involves distinct neural pathways. Secondly, these different threats may motivate different behaviors; for example, self-defending but not predatory animals may cease their aggression when an intruder retreats. We specifically sought to include domestic animals which may again evoke specific behaviors since a common goal for a human is to dominate a domestic animal rather than to escape from it. We did not include self-defending prey animals; an example of this is the mobbing behavior of smaller monkeys towards predating chimpanzees However, such threats may be quite habitat specific. Also, modeling them in VR would require the player to engage in aggressive behavior in the first place, which we decided was too difficult to implement. To motivate the class of disgust-related threats, some animals inflict no or limited acute damage to humans but are commonly feared or avoided in the population in modern-day times. This is evidenced by the prevalence of specific phobias of such animals (e.g., spider phobia with a prevalence of 3.5%), and by non-clinical self-report surveys 8–11. Factor analysis of cross-cultural fear ratings across many animal species has suggested three factors: fear of non-dangerous animals, fear of dangerous animals (including predators and self-defending animals, such as lions or snakes, respectively), and fear of a third category including spiders, worms, and slugs. Several among this latter category of animals were shown to evoke disgust to a similar or greater degree than fear.

Secondly, our primary research goal was to characterize cognitive or neural mechanisms that may be pre-programmed to control behavior towards particular threats. The emergence of pre-programmed mechanisms is likely to involve evolution of behavioral controllers themselves, or the evolution of learning systems that allow acquiring behavioral control during an individual's lifetime; a process that is perfected over timespans of hundreds of thousands of years, as exemplified by the acquisition of skills for tool use. A plethora of threats present in modern civilization and warfare were only developed in the past few centuries; therefore, to avoid an arbitrary cut-off date we focused exclusively on threats that emerged during pre-historic times and were present throughout all or most of human history. For each class of pre-historic threats, the next requirement was to select relevant species or instantiations. We approached this from two angles. First, we drew on contemporary and historical sources to assess the amount of damage inflicted by a particular threat and sought to prioritize those that inflicted the highest rate of deaths or injuries. Secondly, for threats that occur in phobias or are feared in surveys, we used the fear ratings as an additional or primary relevance criterion.

As the fourth selection principle, we sought to elicit a wide range of motor behaviors. To this end, beyond manipulating proximity and attack mode for each of the threats, we also endeavored to include threats of different physical sizes.

Fifth, threats that have been the subject of previous research in humans or other animals were prioritized. This allows any new evidence for behaviors to contribute to a larger field of study and allows for comparisons and validation against existing research.

Lastly, the implementation difficulty of the threat was considered, as the development of some otherwise relatively similar threats may have relied upon more complex texturing and animation work (e.g., the coat of a leopard versus that of a panther, or a long-haired versus a short-haired dog breed). These considerations led us to develop and include the following threats within five categories (cf. Table C.2).

C.2 Supplemental Results

C.2.1 Replicating Sex Differences

Bach et al. (2020) found that sex was the most significant predictor of cautious behavior in a 2D computerized risky foraging task, accounting for 17% of the variance in performance, measured by tokens retained after predator activation. In a post hoc analysis across our combined sample, sex explained around 8% of the variance in a similar measure of performance (cf. Table C.8). This sex difference was not attributable to males collecting fruits more quickly or starting to gather them earlier than females.

Additionally, sex was linked to the minimum distance maintained from the threat, with males keeping a smaller distance from the predator in both studies (in-lab study: 9%; VR studies: 14%). Uniquely in our VR study, we also observed that males tended to delay their escape more than females ($r^2 = .10, p < .05$). This tendency was also not due to faster escape speeds compared to females.

Importantly, males escaped later especially when they had more time (i.e. in the 5 seconds time-to-impact condition), allowing them to take advantage of lower-risk conditions. Indeed, on top of the individual significant effect of sex and time-to-impact, there were significant interactions in linear mixed effect models for escape initiation time ($\beta = .67 \pm .27, t(57) = 2.46, p < .05$), minimum threat distance ($\beta = -2.58 \pm .77, t(57) = -3.34, p < .005$), and the number of fruits collected per trial ($\beta = .81 \pm .32, t(57) = 2.56, p < .05$). In sum, males were less cautious when they had time (in the VR study) or when potential losses were small (in the in-lab study), but adapted their behavior to the same level as females when under time constraints or potential loss increased. Despite acting less cautiously in certain scenarios, these behavioral differences did not translate to higher mortality rates for males in either study, indicating that their riskier strategies did not compromise survival.

However, while these findings align with previously replicated results, it is important to note that I did not pre-select these hypotheses; instead, I analyzed the combined sample directly. This post hoc approach may introduce potential biases, and therefore, these results should be interpreted with caution, as further pre-registered studies are needed to confirm the robustness of these associations.

C.2.2 Replicating Effect of Anxiety

Fung et al. (2019) found that trait anxiety, measured by the STAI-Y, was associated with earlier escape responses for distant threats, but not for closer threats, in a simple 2D task (detailed in Section 1.3.2). Specifically, they did not observe a main effect of anxiety or an interaction between anxiety and close threat type, but they did find a significant interaction between anxiety and distant threat type.

While our experimental design is slightly different, we partially replicate these findings. We found that trait anxiety was correlated with escape initiation time ($r = -.34, r^2 = .11, p < .01$), indicating that a higher anxiety score was associated with earlier escape (cf. Table C.8). This correlation was not significant in the short time-to-impact condition (closer threats) but was significant in the longer time-to-impact condition (distant threats, $r = -.29, r^2 = .08, p < .05$). Using a linear mixed-effects model, we observed significant main effects of anxiety ($\beta = -.61 \pm .23, t(73.05) = -2.67, p < .01$) and time to impact ($\beta = -1.65 \pm 0.14, t(56) = -11.80, p < .0001$), but their interaction only reach trend-like significance ($\beta = .27 \pm .14, t(56) = 1.91, p = .0615$).

Additionally, Fung et al. (2019) found that trait anxiety is related to how successfully participants escaped the predators. Notably, they found a significant interaction between anxiety score and predator type such that anxiety was positively related to escape success in the distant predator condition. However, they found no overall effect of anxiety. Similarly, we also found a significant interaction between anxiety and threat timing ($\beta = .03 \pm .01, t(56) = 2.09, p < .05$) and an overall small but significant correlation of trait anxiety on survival rates ($r = .28, r^2 = .08, p < .05$). (cf. Table C.8).

The effect of anxiety appears to be highly specific. Among over 20 behavioral variables, only two showed significant associations with anxiety, yet these aligned with previous findings from another risky-foraging task that differed substantially in presentation. Specifically, in both studies, anxiety was linked to earlier escape and higher escape success, particularly in response to distant rather than imminent threats.

The differences between our studies may stem from variations in experimental conditions. Our study included only two conditions (1.5s and 5s time to impact), represented by 7 threats with varying speeds, which led to differences in actual distances. In contrast, Fung et al. (2019) used three conditions with consistent predator speeds, yielding three distinct distances and timings of predator attacks. Thus, it remains unclear whether anxiety is more closely related to the timing of the attack or the actual distance of the threat.

C.3 Supplemental Tables

Table C.1: **Thirteen specific behavioral patterns in response to attack from a conspecific, as depicted in Homer's Iliad**. These examples illustrate the richness of informally observed or imagined human behavior; we do not claim they are exhaustive or empirically supported. Translation: Richard Lattimore, The Iliad of Homer, Chicago 1951, as retrieved from Kahane A & Mueller M, The Chicago Homer, https://homer.library.northwestern.edu.

Behavioral pattern	Example	Chapter/verse
Turn around to run	For as [Periphetes] whirled about to get back, he fell over the out-rim of the shield he carried.	15/645-646
Flee forward	He spoke, and Sokos turning from him was striding in flight, but in his back even as he was turning the spear fixed between the shoulders and was driven on through the chest beyond it.	11/446-448
Flee backward	As [Euphorobos] was drawing back, [Menelaos] caught him in the pit of the gullet.	17/47
Flee backward	[The haughty Trojans] thrust him away from them so that he gave ground backward staggering.	5/623-625
Draw back, staring/turning around	[Aias] stood stunned, [...] and drew back, throwing his eyes round the crowd of men, like a wild beast, turning on his way, shifting knee past knee only a little.	11/545-547
Drop to the knee to avoid thrown object	Glorious Hektor kept his eyes on him, and avoided [the spear], for he dropped, watchful, to his knee, and the bronze spear flew over his shoulder.	22/274-275
Bend forward to avoid thrown object	[Automedon], keeping his eyes straight on him, avoided the bronze spear. For he bent forward.	17/526-527
Tremble	But the Trojans were taken every man in the knees with trembling.	20/44
Stand still	But Aineias, free of the long spear, stood still, and around his eyes gathered the enormous emotion and fear, that the weapon had fixed so close to him.	20/281-283
Huddle inside an enclosure	Thestor, Enops' son, who huddled inside his chariot.	16/402
Call for help	[Odysseus] gave back a little way and called out for his companions.	11/461
Flung at opponent's knee and beg for mercy	Brilliant Achilleus held the long spear uplifted above him, straining to stab, but [Lykaon] under-ran the stroke and caught him by the knees [...]: "Achilleus, I am at your knees. Respect my position, have mercy upon me."	21/67-74
Group of people fleeing in many different directions	The high-hearted Epeians fled one way and another in terror when they saw the man fall.	11/744-745
Withdraw into a group of people	But Alexandros the godlike when he saw Menelaos, [...] to avoid death he shrank into the host of his own companions.	3/30-32

Table C.2: **Summary of threat classes and selected threats for each study**

Threat	Instantiation(s)	Threat class(es)	E1	E2	Reason for change
Human	Male, medium skin tone, medium build	Conspecific	x	x	
Bear	Brown bear	Predatory, defensive	x	x	
Crocodile	American alligator	Predatory, defensive	x		Comparison to other scenarios limited
Big cat	Panther (black furred variant of leopard)	Predatory, defensive	x	x	
Bovine	Spanish fighting bull	Defensive	x		No specific behaviors relevant to study purpose
Elephant	African bush elephant	Defensive	x	x	
Canine	Dobermann dog	Defensive	x	x	
Snake	Viper (Bitis arietans)	Defensive	x	x	
Scorpion	Indian red scorpion	Disgust-relevant, Defensive	x		Sometimes not detected by participants
Spider	Australian funnel-web spider	Disgust-relevant, Defensive	x	x	
Insect	Yellow-jacket wasp	Disgust-relevant, Defensive	x		Sometimes not detected by participants
Rat	Brown rat	Disgust-relevant	x		No specific behaviors relevant to study purpose
Collision with large object	Boulder rolling down a hill	Environmental	x	x	
Red box	Medium Red box	Control	x	x	
	Small Red box	Control	x		No specific behaviors relevant to study purpose
Time bomb	Dynamite sticks	Control	x		No specific behaviors relevant to study purpose

Table C.3: **Threat speeds and the resulting calculated distance from fruit picking position for the two time-to-impact conditions.** "No threat" conditions were included to assess baseline behavior; here the "speed" is a dummy variable that served to determine the distance of the visual grass elements from the player. The slow/fast column relates to analysis; for the distance calculation, the "slow" threat equations were used for the (non-chasing) rock.

Threat name	Speed (m/s)	Slow/fast for analysis	Threat distance (1.5s time-to-impact)	Threat distance (5.0 s time-to-impact)
Elephant	6.4	Fast	18.1	43
Bull	5.49	Not included	14.46	36.175
Human	5.2	Fast	13.3	34
Panther	4.07	Not included	8.78	25.525
Bear	3.84	Fast	7.86	23.8
Wasps	3.75	Not included	7.5	23.125
Doberman	3.53	Fast	6.62	21.475
Snake	1.92	Slow	2.88	9.56
Rat	1.36	Not included	2.04	6.8
Spider	0.32	Slow	0.48	1.6
Scorpion	0.16	Not included	0.24	0.8
Rolling Rock	6.283	Not included	9.425	31.416
Crocodile	0.75	Not included	1.125	3.75
Time bomb	N/A	N/A	1.125	3.75
RedBox Medium	3.375	N/A	6.0	20.3125
RedBox Small	0.375	N/A	0.5625	1.875
No threat (1)	4.07	N/A	8.78	25.525
No threat (2)	1	N/A	1.5	5

Table C.4: **Number of epochs in the second part of E2 according to block type.** The second line indicates the possible numbers of the specific type of epoch described above and the third line indicates the actual number of the specific epoch realized on average across participants in E2.

Block name	Characteristics of epochs / Possible number of epochs / Actual number of epochs							
Force shield	Panther, Shield, 1.5s	Panther, Shield, 5s	Panther, No shield, 1.5s	Panther, No shield, 5s	No threat, Shield, 1.5s	No threat, Shield, 5s	No threat, No shield, 1.5s	No threat, No shield, 5s
	1-2	1-2	1-2	1-2	0-1	0-1	0-1	0-1
	1.67	1.77	1.87	1.83	0.80	0.83	0.80	0.80
Hands up	Panther, Hands-up, 1.5s	Panther, Hands-up, 5s	Panther, No hands-up, 1.5s	Panther, No hands-up, 5s	No threat, Hands-up, 1.5s	No threat, Hands-up, 5s	No threat, No hands-up, 1.5s	No threat, No hands-up, 5s
	1-2	1-2	1-2	1-2	0-1	0-1	0-1	0-1
	1.77	1.67	1.83	1.77	0.80	0.70	0.50	0.73
Medusa	Panther, Lethal force, 1.5s	Panther, Lethal force, 5s	No threat, Lethal force, 1.5s	No threat, Lethal force, 5s				
	3-4	3-4	0-2	0-2				
	3.60	3.77	1.57	1.60				

Table C.5: **All behavioral variables and their description on how they were computed**

Name	Description
Escape to shelter	Participant went into the shelter (as logged in Unity)
Survived	Participant did not go into the shelter (as logged in Unity)
Virtual death	Participant entered in contact with the threat (as logged in Unity)
Minimum distance from shelter	Smallest distance between participant's head tracker and shelter, regardless of outcome
Minimum distance from threat	Smallest distance between participant's waist tracker and threat center during escape
Time of escape initiation	Begin of the movement away from the bush (velocity threshold 0.1 m/s) that brings the participant's head tracker at least 75 cm away from fruit-bearing bush for the first time (at this distance, they cannot reach fruit any more)
Initiated escape	Participant initiated escape as per above
Interrupted escape	Participant initiated escape as per above but did not reach the shelter as logged by Unity
Fruit picking rate	Number of collected fruit(s) per second (as logged by Unity)
Escape speed	Speed of the participant's waist tracker between time of escape initiation, and end of escape. To compute speed, position data are resampled at a rate of 10 Hz and median-smoothed over 3 data points (300 ms), in order to avoid an impact of momentary tracker mislocation.
Body orientation	Cosine of angle between a vector pointing forward from the participant's pelvis, and the line between the participant and the threat, while ignoring the upward axis. This results in values between -1 and 1, where 1 is towards the threat, -1 is away, and 0 is perpendicular.
Head orientation	Same as body orientation but for a vector pointing forward from the participant's forehead
Visual scanning	Cumulative angle of frame-by-frame movements of a vector pointing forward from the participant's forehead (i.e. ignoring movements around this axis)

Table C.6: **Retained correlations between questionnaires and behavioral variables discovered in E1 and used to compose GLM models to test in E1.** The full set of correlations are only tested in E1. The p-values are not corrected for multiple comparisons and are presented as a heuristic guide only. The retained GLMs are listed in Table 4.2.

Dependent variable	Questionnaires	E1 (r, r^2, p)
Minimum distance from threat during escape	Fear (FSS)	$r = .48$, $r^2 = .23$, $p < .01$
	Spider phobia (SPQ)	$r = .44$, $r^2 = .20$, $p < .05$
Escape initiation time	Fear (FSS)	$r = -.50$, $r^2 = .25$, $p < .01$
	Spider phobia (SPQ)	$r = -.34$, $r^2 = .12$, $p < .05$
Fruit picking during 0-1.5 s after threat appears	Fear (FSS)	$r = -.46$, $r^2 = .21$, $p < .01$
	Spider phobia (SPQ)	$r = -.39$, $r^2 = .16$, $p < .05$
Head orientation during 0-1.5 after threat appears	Sensation seeking (BSSS)	$r = .42$, $r^2 = .18$, $p < .05$
	Fear (FSS)	$r = -.43$, $r^2 = .19$, $p < .05$
	Spider phobia (SPQ)	$r = -.43$, $r^2 = .18$, $p < .05$

Table C.7: **Statistical results for the second part of E2 (block 2 − 4).** Each hypothesis tests a distinct a-priori question and hence p-values are not corrected for multiple comparisons.

Name	Dependent variable	Contrast or Interaction	E2 ($\beta \pm SE, t(df), p$)
Force shield block			
E2-H1	Escape to shelter	Shield vs no shield, panther, 1st epoch	$\beta = -5.97 \pm 1.29$, $z = -4.65$, $p < .0001$
Supporting test	Minimum distance from shelter	Shield vs no shield, panther, 1st epoch	$\beta = 2.13 \pm 0.10$, $t(278) = 21.01$, $p < .0001$
E2-H2	Minimum distance from shelter	Panther vs no threat, shield, 1st epoch	$\beta = -0.13 \pm 0.10$, $t(278) = -1.26$, $p = .21$
E2-H3	Fruit picking rate from threat appearance to the minimum duration of the epoch (12.5 s)	Panther vs no threat, shield, 1st epoch	$\beta = -0.45 \pm 0.10$, $t(268) = -4.22$, $p < .0001$
Supporting test	Fruit picking rate from threat appearance to the minimum duration of the epoch (12.5 s)	Shield vs no shield, no threat, 1st epoch	$\beta = -0.44 \pm 0.11$, $t(268) = 4.07$, $p < .0001$
E2-H4	Visual scanning during 0-1.5 s after threat appears	Panther vs no threat, shield, 1st epoch	$\beta = 43.1 \pm 18.1$, $t(216) = 2.38$, $p < .05$
Supporting test	Gaze elevation during 0-1.5 s after threat appears	Panther vs no threat, shield, 1st epoch	$\beta = 25.7 \pm 6.55$, $t(215) = 3.92$, $p < .0001$
Medusa block			
E2-H5	Virtual death by lethal force	Interaction of lethal force and epoch order	$\beta = 0.60 \pm 0.13$, $z = -4.57$, $p < .0001$
Hands-up block			
E2-H6	Escape to shelter	Hands-up (mean = 0) vs no hands-up, panther, 1st epoch	Binomial test against 0: $p < .0001$
Supporting test	Minimum distance from shelter	Hands-up vs no hands-up, panther, 1st epoch	$\beta = 2.23 \pm 0.10$, $t(278) = 21.99$, $p < .0001$
Supporting test	Minimum distance from shelter	Panther vs no threat, hands-up, 1st epoch	$\beta = -0.03 \pm 0.10$, $t(278) = -0.32$, $p = .75$

Table C.8: **Exploratory correlations between questionnaires and behavioral variables in combined sample** (N=58). Spider phobia was measured by SPQ-12, snake phobia by SNAQ-12, and Anxiety by STICSA-T

Behavioral variable	Spider phobia	Snake phobia	Trait Anxiety	Sex
Escape to shelter	$r = .41, r^2 = .17, p < .001$	$r = .26, r^2 = .07, p < .05$	ns	ns
Interrupted escape	$r = -.28, r^2 = .08, p < .05$	ns	ns	ns
Survival	$r = .37, r^2 = .14, p < .005$	ns	$r = .28, r^2 = .08, p < .05$	ns
Minimum Safe Distance	$r = -.40, r^2 = .16, p < .005$	$r = -.26, r^2 = .07, p < .05$	ns	$r = -.28, r^2 = .08, p < .05$
Minimum distance from threat during escape	$r = .39, r^2 = .15, p < .005$	ns	ns	$r = .38, r^2 = .14, p < .005$
Escape Initiation Time	ns	ns	$r = -.34, r^2 = .11, p < .01$	$r = -.32, r^2 = .10, p < .05$
Number of fruits collected per trial (Performance)	ns	ns	ns	$r = -.28, r^2 = .08, p < .05$
Total number of fruits	ns	ns	ns	$r = -.27, r^2 = .07, p < .05$
Head orientation during 0-1.5 s after threat appears	$r = -.42, r^2 = .18, p < .001$	$r = -.32, r^2 = .10, p < .05$	ns	ns
Body orientation during 0-1.5 s after threat appears	$r = -.40, r^2 = .16, p < .005$	ns	ns	ns
Fruit picking during 0-1.5 s after threat appears	$r = -.28, r^2 = .08, p < .05$	ns	ns	ns

C.4 Supplemental Figures

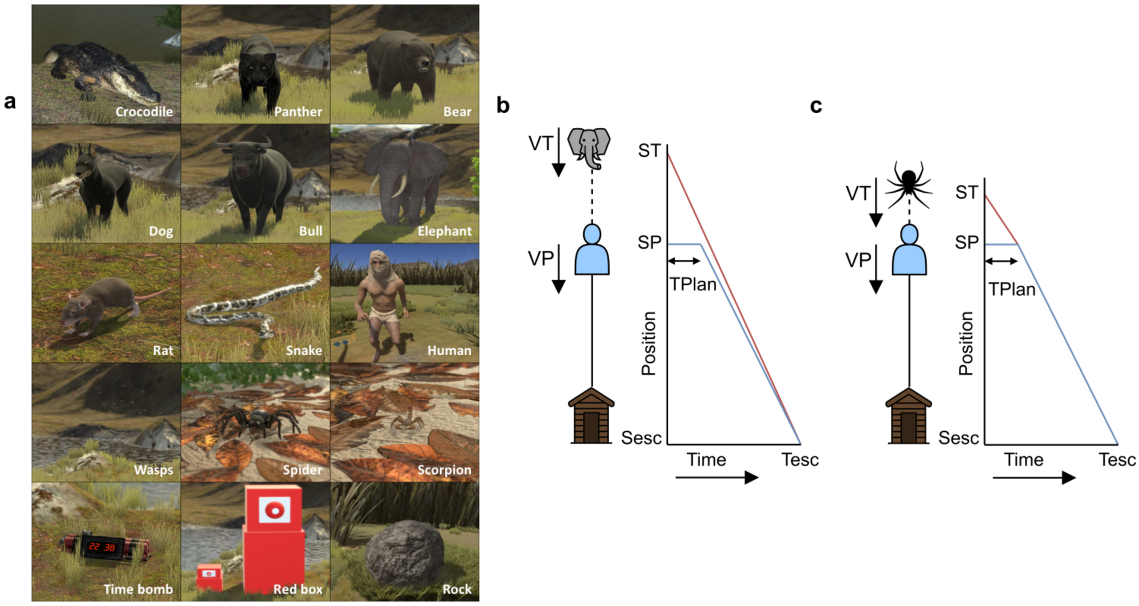

Figure C.1: **Experimental setup.** **(A)** Illustration of all 16 threats used across experiments 1-2. Note that the red box comes in two sizes (small and medium) with different speeds (slow/fast). **(B)** Fast threats were initially hidden behind grass (at ST), the position of which was calculated such that the participant (at SP) would just collide with the threat at the shelter if they took a certain time (TPlan) to plan and initiate their escape. **(C)** Slower threats could not outrun a moving participant, so instead were placed such that they would collide with the participant at the fruit bush. In both experiments, TPlan was realized as two time-to-impact conditions, 1.5 s and 5 s.

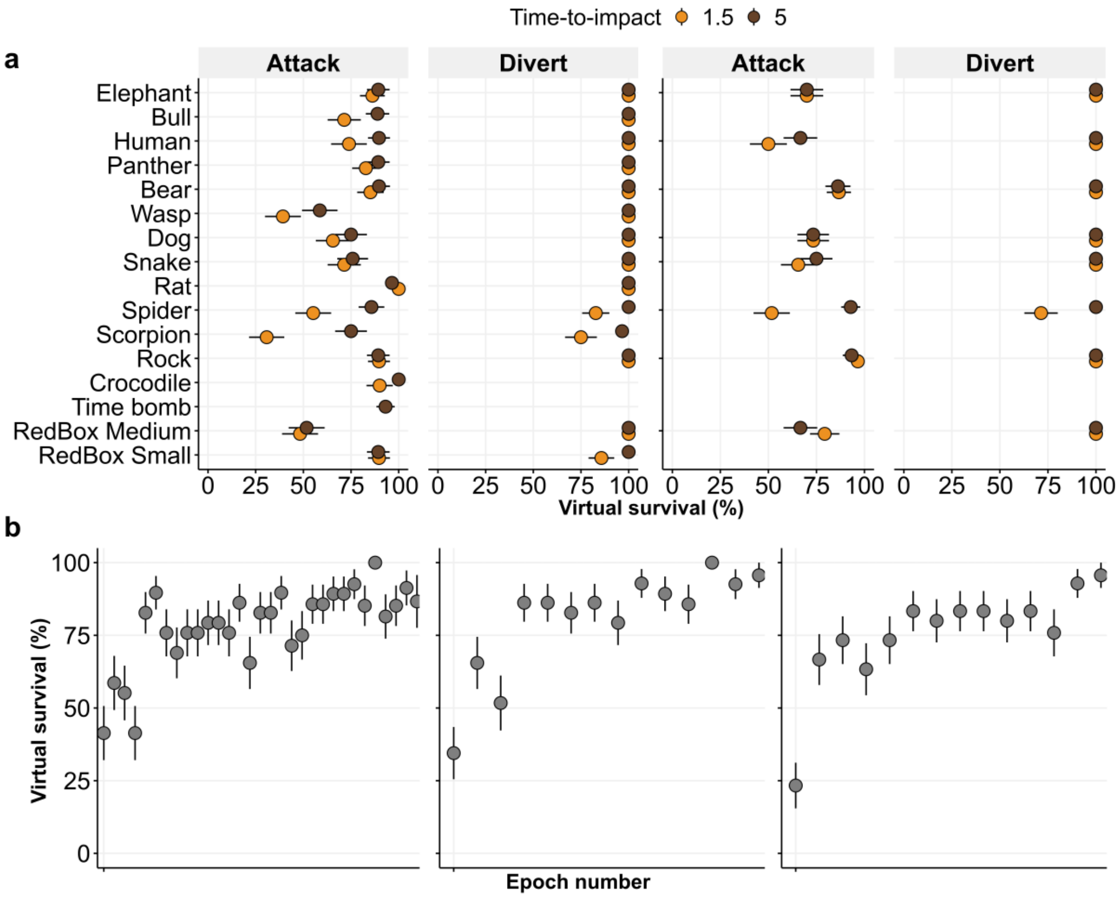

Figure C.2: **Virtual Survival.** **(A)** Percentage of virtual survival for all threats and conditions in E1 (left) and E2 (right). **(B)** Percentage of virtual survival during attack over epochs for all threats in E1 (left) and for the threats included across both experiments in E1 (center) and E2 (right). Points with error bars represent the mean and standard error across all participants and epochs.

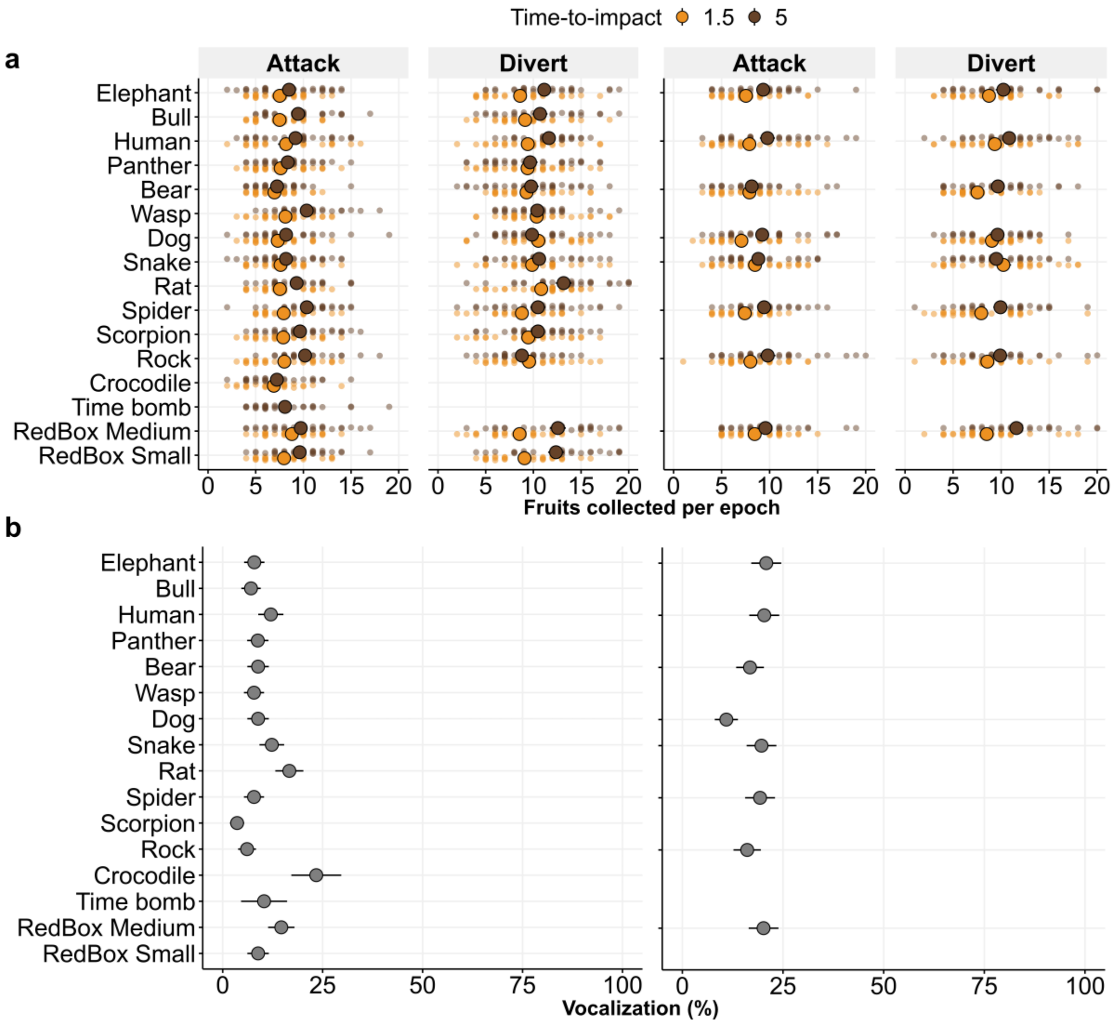

Figure C.3: **Participants are engaged as demonstrated by instructed and non-instructed behaviors. (A)** Fruits collected per epoch for all threats and conditions in E1 (left) and E2 (right). Large points with error bars represent the mean and standard error across all participants and epochs, and small points represent individual epochs. **(B)** Percentage of vocalization for all threats in E1 (left) and E2 (right). Points with error bars represent the mean and standard error across all participants and epochs.

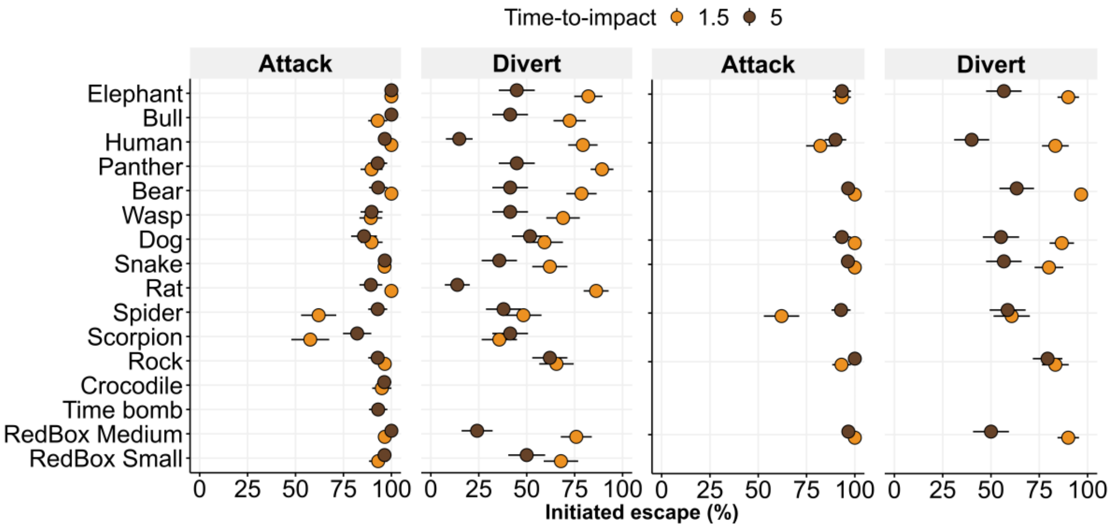

Figure C.4: **Initiated escapes.** Percentage of initiated escape of all epochs of the same type and for all threats and conditions in E1 (left) and E2 (right). Points with error bars represent the mean and standard error across all participants and epochs.

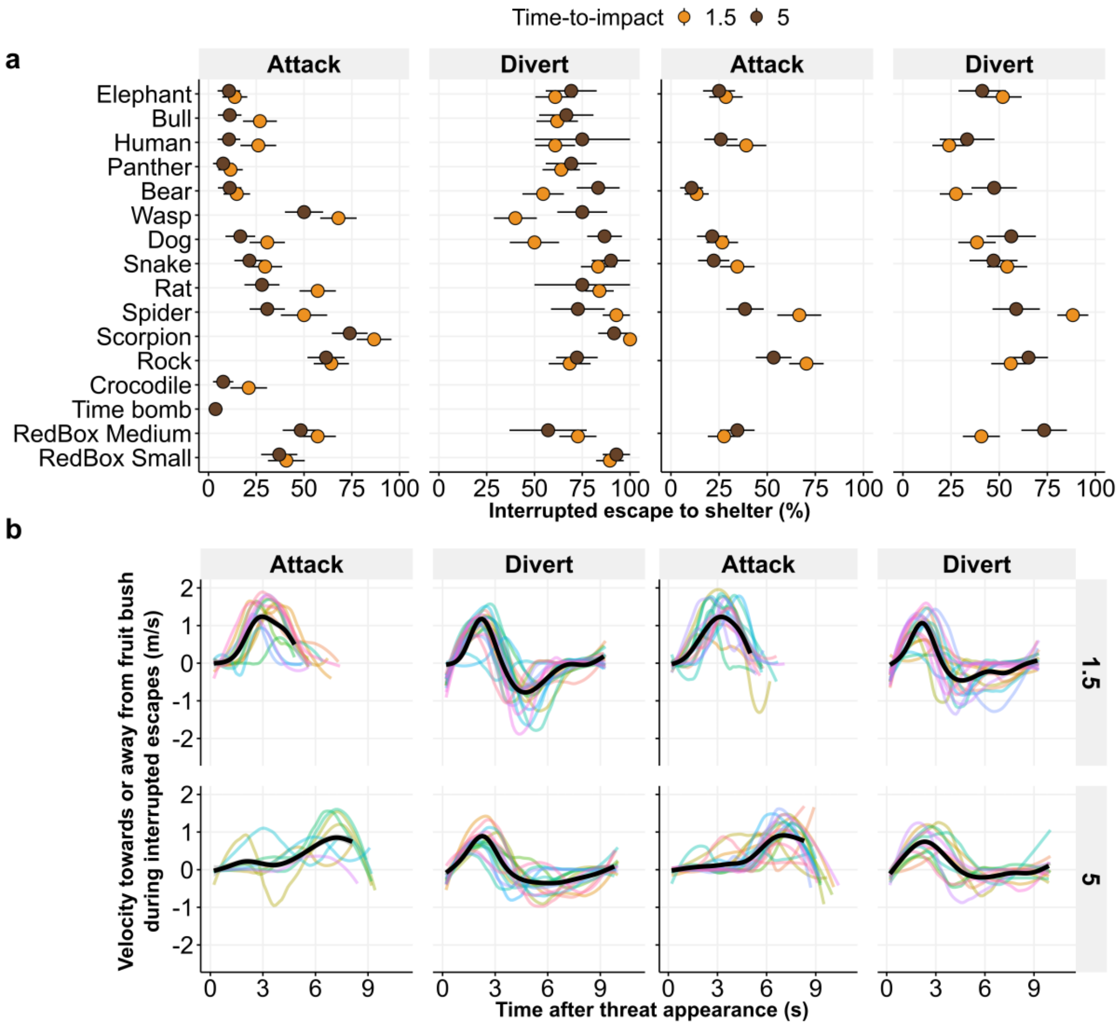

Figure C.5: **Interrupted escapes.** **(A)** Percentage of escape interruption of all initiated escapes of the same type for all threats and conditions in E1 (left) and E2 (right). Points with error bars represent the mean and standard error across all participants and epochs. **(B)** Velocity towards or away from the fruit bush during interrupted escapes for all conditions in E1 (left) and E2 (right). Each colored line represents a participant's mean across epochs of the same type, and the black line is their overall mean.

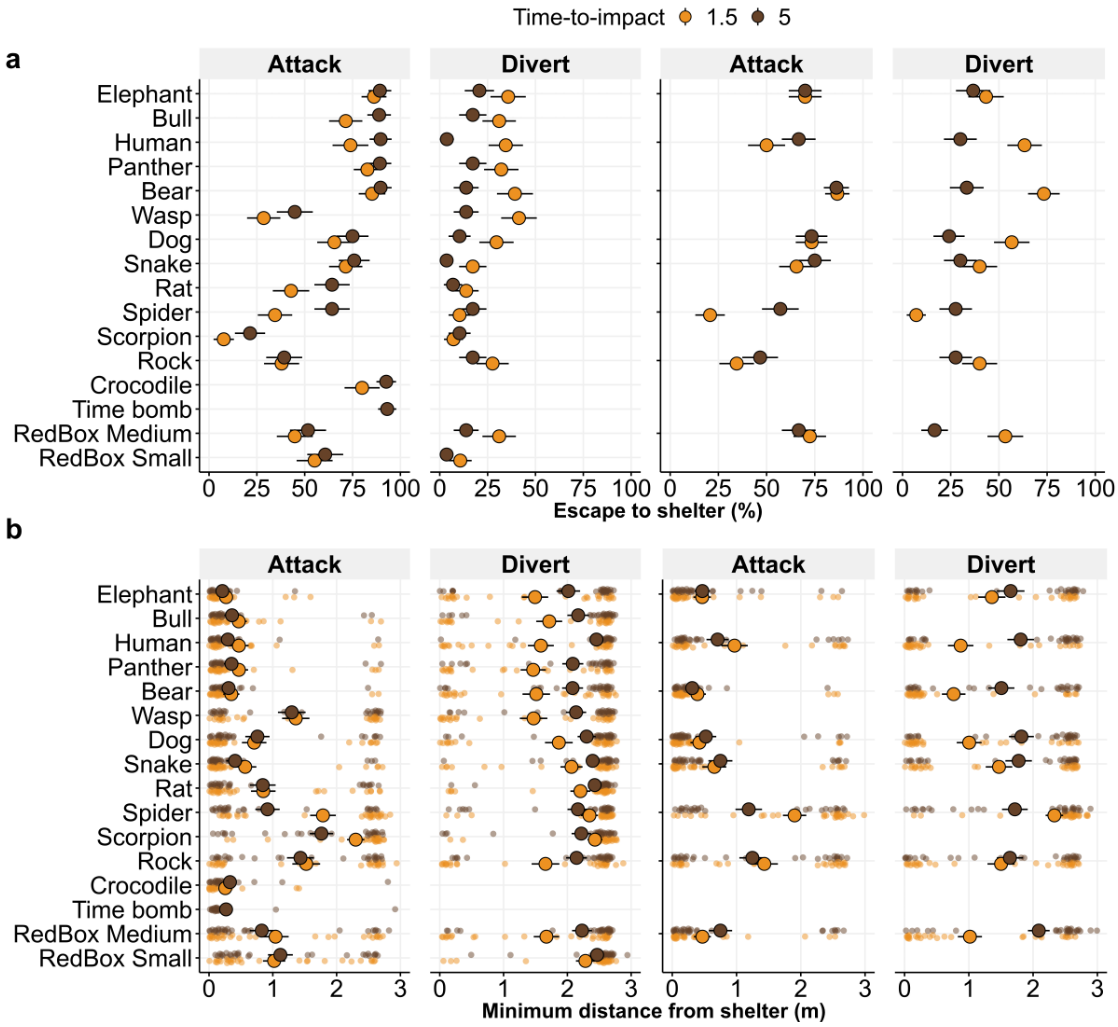

Figure C.6: **Use of shelter. (A)** Percentage of escape to shelter and **(B)** minimum distance from shelter for all threats and conditions in E1 (left) and E2 (right). Large points with error bars represent the mean and standard error across all participants and epochs, and small points represent individual epochs.

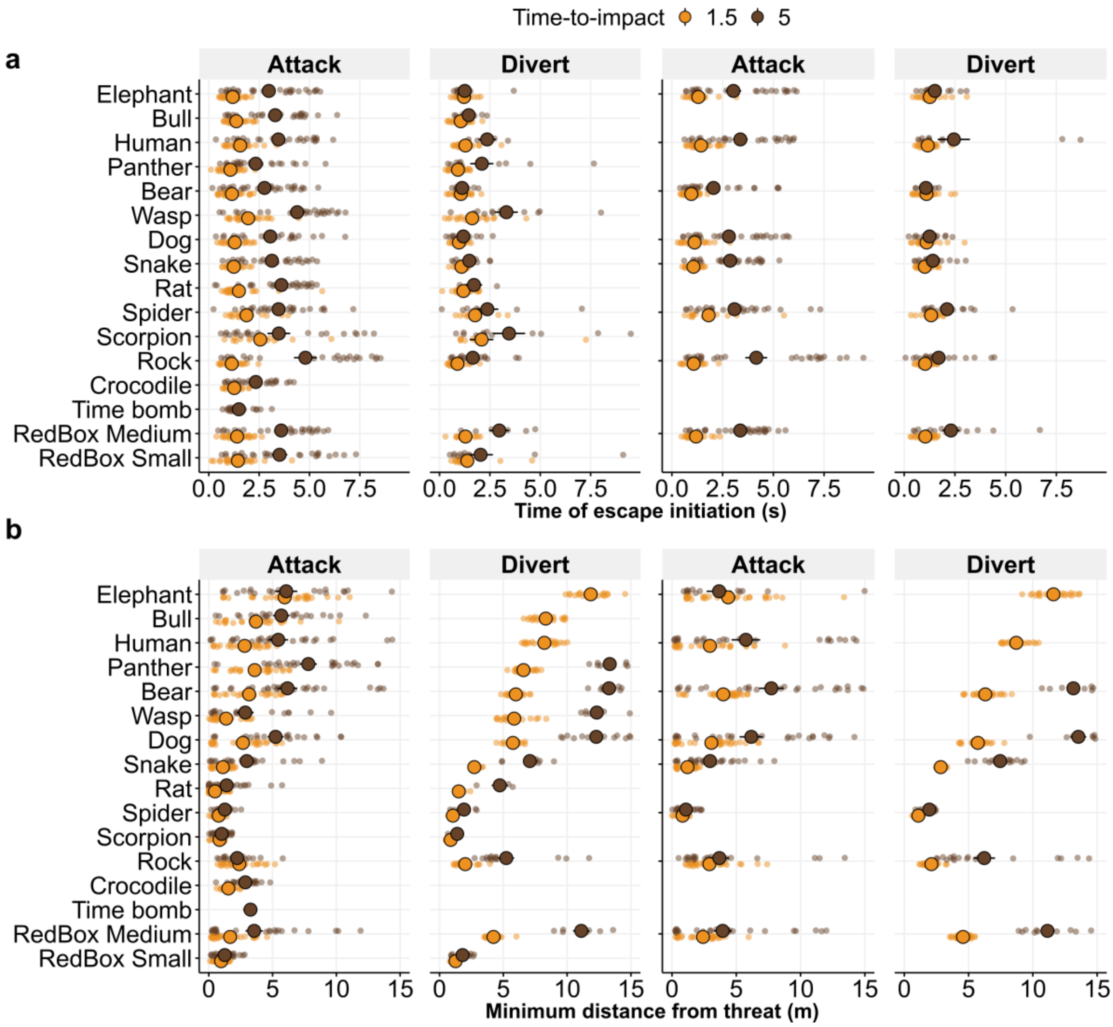

Figure C.7: **Time and distance from threat.** **(A)** Time of escape initiation relative to threat appearance and **(B)** minimum distance from threat during escape for all threats and conditions in E1 (left) and E2 (right). Large points with error bars represent the mean and standard error across all participants and epochs, and small points represent individual epochs.

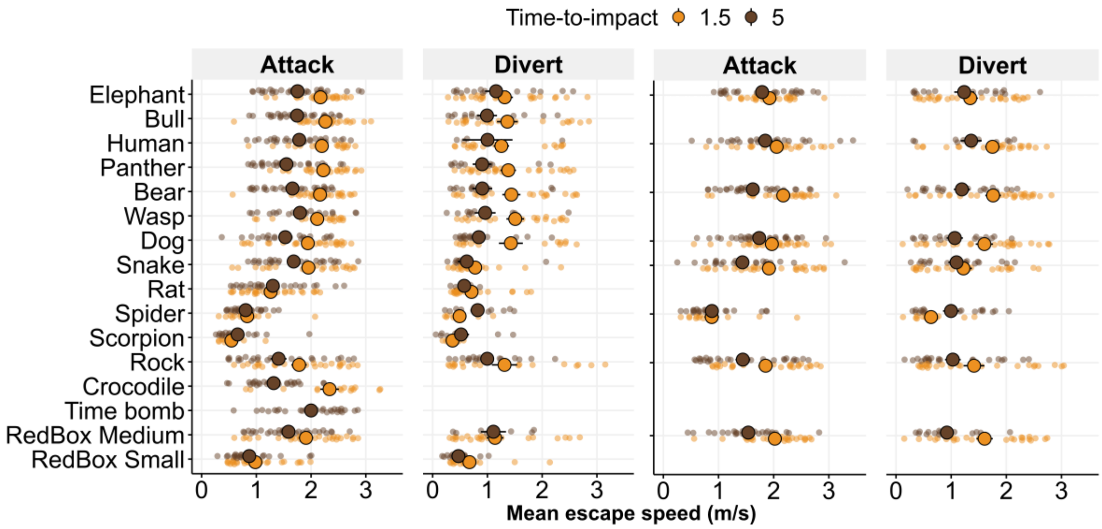

Figure C.8: **Mean Escape Speed.** Mean speed during escape for all threats and conditions in E1 (left) and E2 (right). Large points with error bars represent the mean and standard error across all participants and epochs, and small points represent individual epochs.

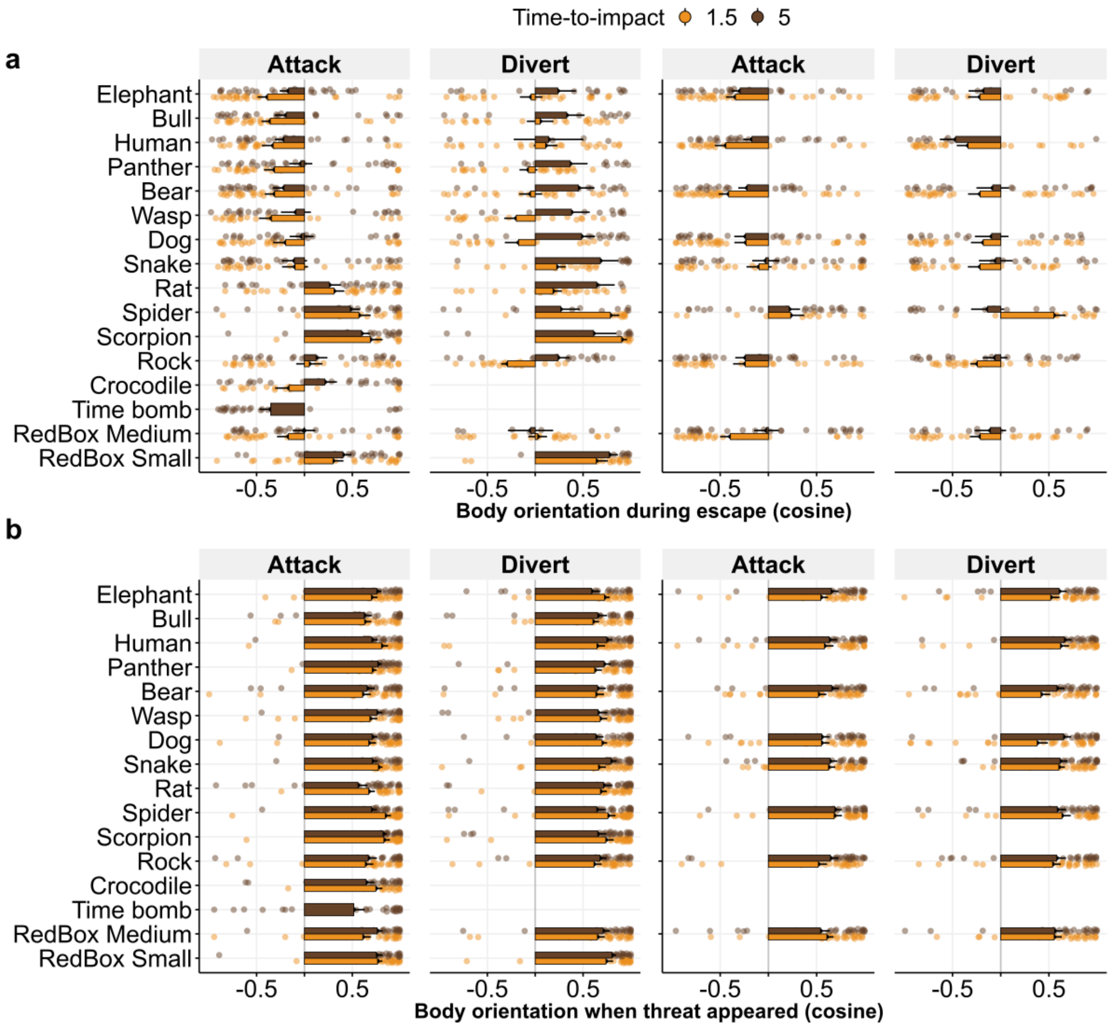

Figure C.9: **Body orientation.** **(A)** Body orientation (averaged cosine of orientation angle away from threat, ranging from -1: away from threat to 1: towards threat) during escape and **(B)** within the initial 1.5 s of threat appearance for all threats and conditions in E1 (left) and E2 (right). Bars with error bars represent the mean and standard error across all participants and epochs, and points represent individual epochs.

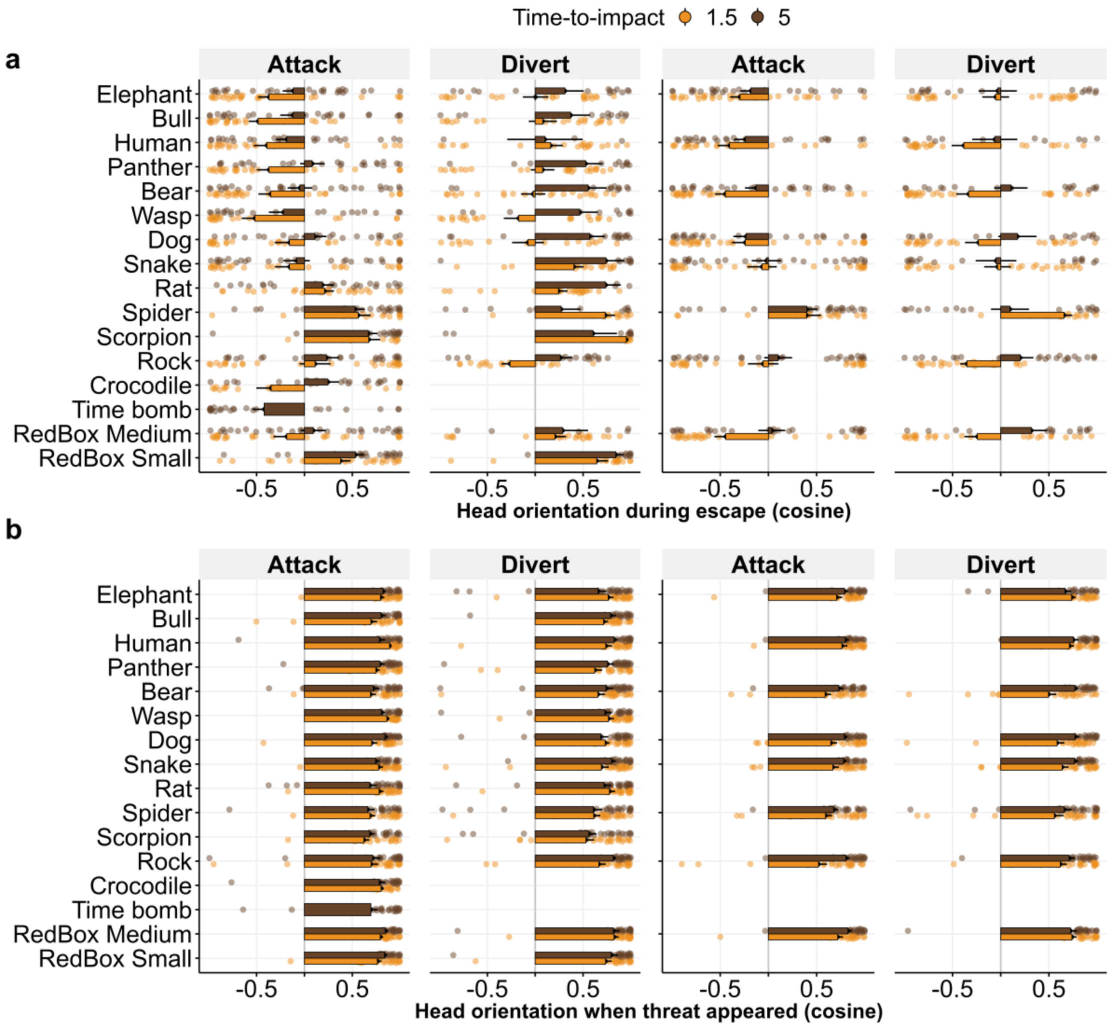

Figure C.10: **Head orientation.** **(A)** Head orientation (averaged cosine of orientation angle away from threat, ranging from -1: away from threat to 1: towards threat) during escape and **(B)** within the initial 1.5 s of threat appearance for all threats and conditions in E1 (left) and E2 (right). Bars with error bars represent the mean and standard error across all participants and epochs, and points represent individual epochs.

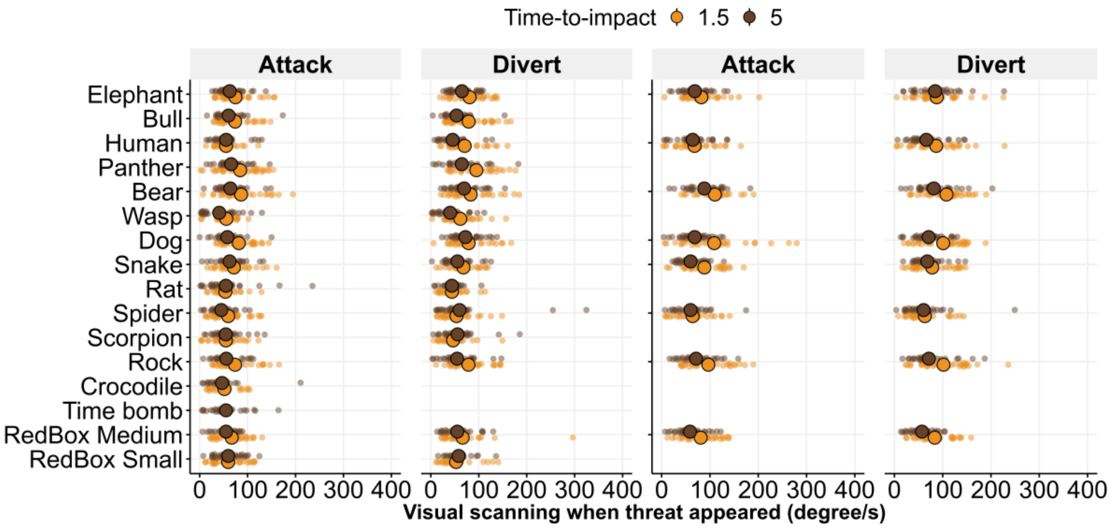

Figure C.11: **Visual scanning** within the initial 1.5 s of threat appearance for all threats and conditions in E1 (left) and E2 (right). Large points with error bars represent the mean and standard error across all participants and epochs, and small points represent individual epochs.

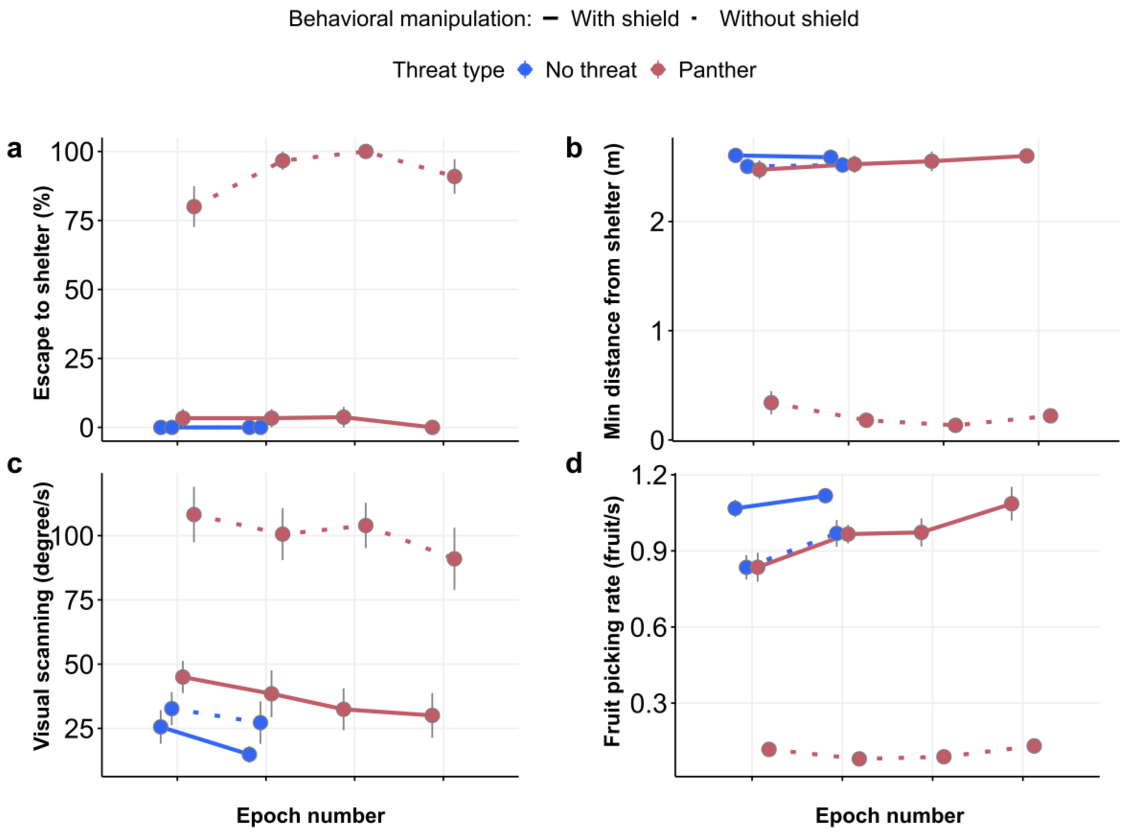

Figure C.12: **Force-shield block.** Behavioral variables over epoch in "force shield" block of E2 including **(A)** escape to shelter, **(B)** minimum distance from shelter during escape, **(C)** visual scanning within the initial 1.5 s of threat appearance, and **(D)** fruit picking rate over the entire epoch after threat appearance. Points with error bars represent the mean and standard error across all participants.

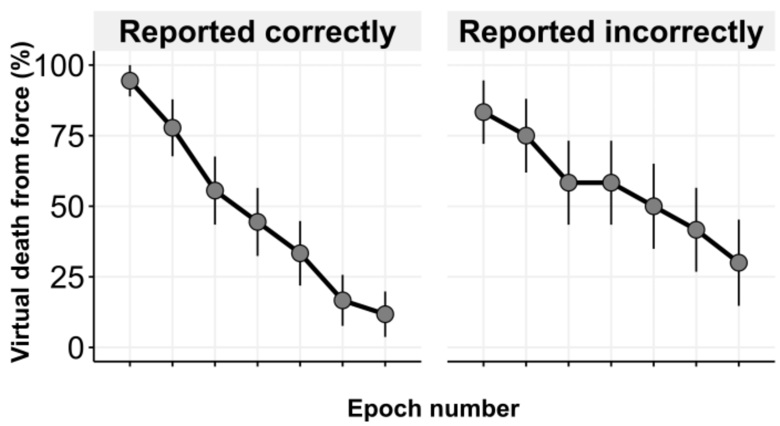

Figure C.13: **Medusa block.** Virtual death from magical force over epoch in "Medusa" block of E2. In a post-hoc interview, 60% of participants reported the correct force-activating movement, 40% did not. Points with error bars represent the mean and standard error across participants.

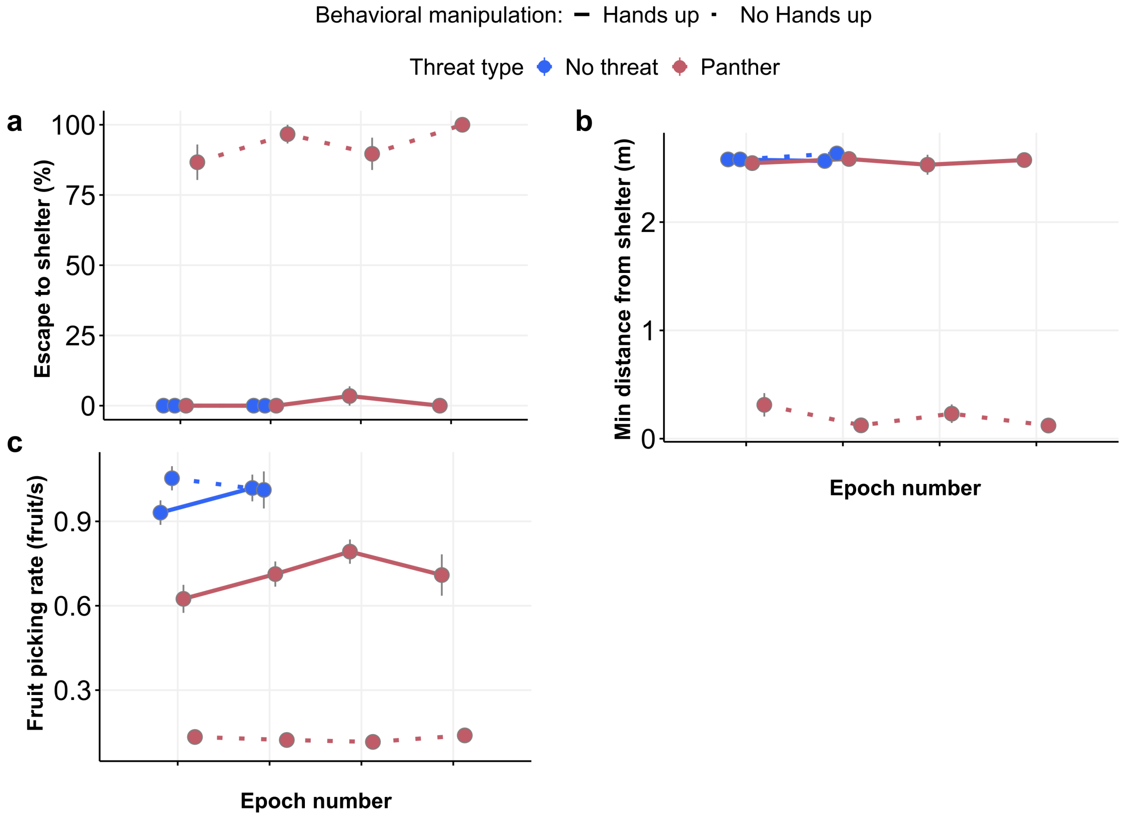

Figure C.14: **Hands-up block.** Behavioral variables over epoch in "hands-up" block of E2 including **(A)** escape to shelter, **(B)** minimum distance from shelter during escape, and **(C)** fruit picking rate over the entire epoch after threat appearance. Points with error bars represent the mean and standard error across all participants.

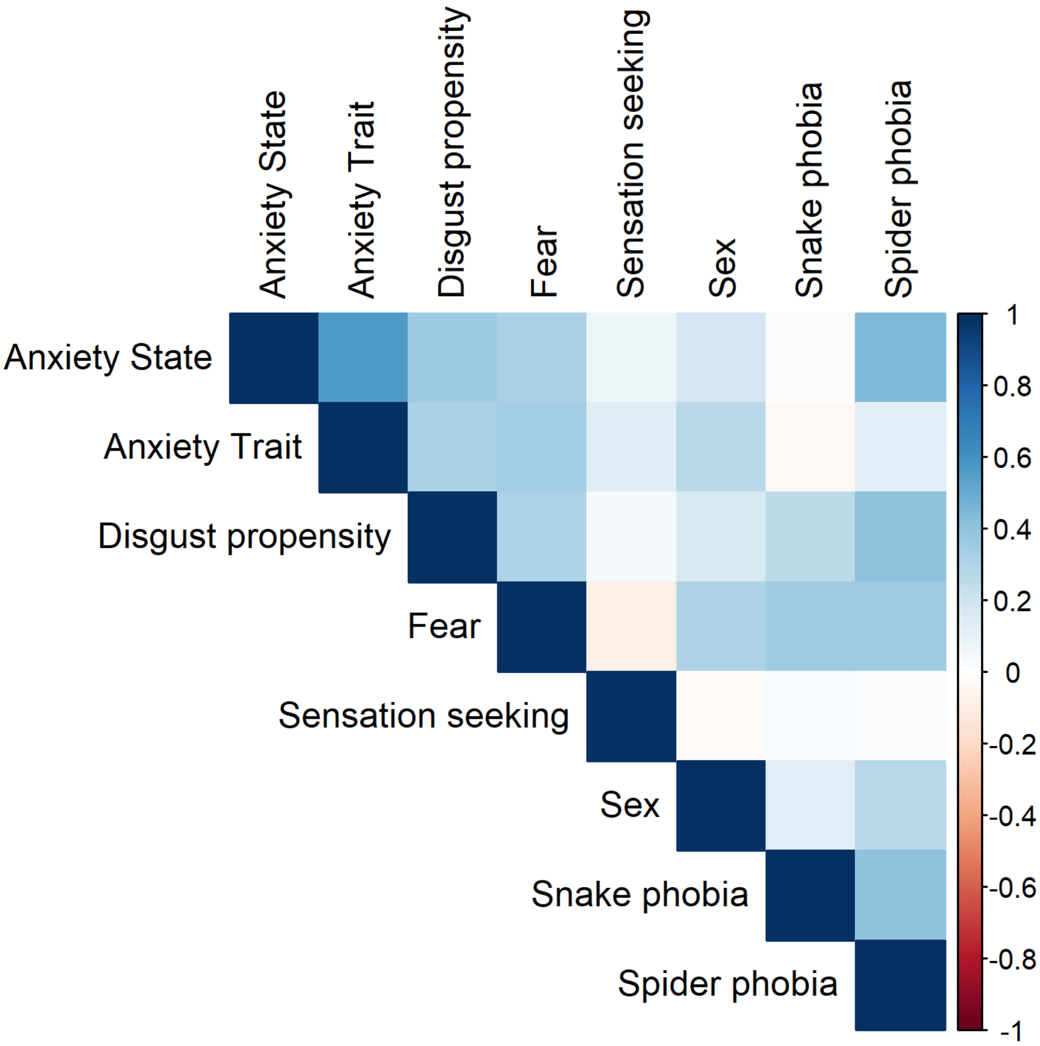

Figure C.15: **Correlation matrices between questionnaires in the combined sample**. The color scale indicates the correlation coefficient.