Decision Making under Threat: An Ecological Framework

Thesis by

Song Qi

In Partial Fulfillment of the Requirements for the

Degree of

Social and Decision Neuroscience

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2020

Defended 5/16/2019

Song Qi
ORCID: 0000-0002-5886-849X

ACKNOWLEDGEMENTS

First and foremost, I want to thank my primary supervisor, Professor Dean Mobbs, who provided me with abundant opportunities and constant support during my pursuit for a Ph.D. degree. He acted both as an advisor and a friend. There were silly mistakes I made, which he tolerated; and there were sparkling ideas which we subsequently pushed into fruition.

A similar gratitude belongs to my committee members. Professor John O'doherty has always been sharp in spotting flaws and alternative directions in my projects; Professor Ralph Adolphs provided me with numerous advice on potential paths to embark on (partially,with his highly educational lab meetings); Professor Colin Camerer offered indispensable help in the social influence branch of my work.

The four professors above are also role models for me, with their tenacious work ethic.

I would like to thank my lab mates, fellow graduate students and friends, both at Columbia University and Caltech. Together we sailed through scientific inquiries and created wondrous memories. I would not have been able to build up this thesis without their kind support.

I thank professor Keith Kendrick, who invited me to the entrance of cognitive neuroscience. He ignited my passion for research when I was still an undergrad lost in his ego, and carefully guided that passion to the direction that bears fruit.

I thank my parents for their unconditional support. Words carry little weight here.

Finally, I would like to express my gratitude to Joyce Ren, a dear friend who shared my emotional burdens and helped me endure hardships. Her aid has been indispensable during my darkest hours.

The matching of lived experiences with lessons learned is sheer serendipity. Some lessons come too early and we don't know their implications for our future. Some wisdoms simply come too late, when grave mistakes have been irrevocably made. I'm grateful for the inflection points that appear with perfect timing. It doesn't happen often enough.

ABSTRACT

Humans, like other animals, have evolved a set of neural circuits whose primary function is survival. In the case of predation, these circuits include "reactive fear" circuits involved in fast escape decisions, and "cognitive fear" circuits that are involved in more complex processing associated with slow strategic escape. In the context of flight initiation distance (FID), using neuroimaging combined with computational modeling, we support this differentiation of fear circuits by showing that fast escape decisions are elicited by the periaqueductal gray and midcingulate cortex, regions involved in reactive flight. Conversely, slower escape decisions rely on the hippocampus, posterior cingulate cortex, and prefrontal cortex, a circuit implicated in behavioral flexibility. We further tested whether individual differences in trait anxiety would impact escape behavior and neural responses to slow and fast attacking predators. Behaviorally, we found that trait anxiety was not related to escape decisions for fast threats, but individuals with higher trait anxiety escaped earlier during slow threats. Functional MRI showed that when subjects faced slow threats, trait anxiety positively correlated with activity in the vHPC, mPFC, amygdala and insula. Further, the strength of the functional coupling between the vHPC and mPFC was correlated with the degree of trait anxiety. A similar pattern of separation in survival circuits is also found in a follow up study utilizing the concept of margin of safety (MOS) with multivariate pattern analysis of fMRI data. In addition, we also discussed how decision making under threat was influenced by social factors such as reputation. Overall, these results provide new insights into decision making under threat and a separation of fear into reactive and cognitive circuits.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Bowen J Fung, Song Qi, Demis Hassabis, Nathaniel Daw, and Dean Mobbs. Slow escape decisions are swayed by trait anxiety. *Nature human behaviour*, 3 (7):702–708, 2019. doi: 10.1038/s41562-019-0595-5.
S.Q. designed the experiment, collected the behavioral and imaging data, participated in analyzing the data, and participated in writing the manuscript.

[2] Song Qi, Owen Footer, Colin F Camerer, and Dean Mobbs. A collaborator's reputation can bias decisions and anxiety under uncertainty. *Journal of Neuroscience*, 38(9):2262–2269, 2018. doi: 10.1523/JNEUROSCI.2337-17.2018.
S.Q. participated in designing the experiment, helped collecting the behavioral and imaging data, analyzed the data, and participated in writing the manuscript.

[3] Song Qi, Demis Hassabis, Jiayin Sun, Fangjian Guo, Nathaniel Daw, and Dean Mobbs. How cognitive and reactive fear circuits optimize escape decisions in humans. *Proceedings of the National Academy of Sciences*, 115(12):3186–3191, 2018. doi: 10.1073/pnas.1712314115.
S.Q. designed the experiment, collected the behavioral and imaging data, analyzed the data, and participated in writing the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

LIST OF TABLES

NOMENCLATURE

FID. Flight Initiation Distance. In ecology literature, it is the distance between the prey and the predator when the predatory chase starts.

MOS. Margin of Safety. In ecology literature, it is the distance between the prey and the safety refuge when the prey chooses his potential foraging location.

Chapter 1

INTRODUCTION

Fear, as one of the most primal and outstanding emotions, dominates our daily life in a wide range of aspects. This however, has its own merit. Contrary to the popular notion of fear being an emotion that hinders daily decision making and personal development, fear is also essential in modulating behaviors under various decision scenarios, and subsequently optimizes decision outcome and promotes survival.

This observation stimulates two important questions:

- On the behavioral level, how does fear modulates decisions under threat.
- On the neural level, what are the corresponding neural mechanisms.

The focus of the current thesis will tackle the two questions using a combination of behavioral experimental methods, brain imaging technique, and computational modelling. The line of research presented in this thesis is based on two basic ideas: One, that fear is not only a subjective feeling, but also a functional unit promoting survival; two, that fear is better studied and understood in an ecologically valid context.

1.1 Threat, fear and decision making

Fear in human is a combination of higher order subjective feeling and lower level defensive mechanisms, and neither side should be overlooked. LeDoux's "Survival Circuits Theory" argues that human fear is distinct from the "defensive circuits" studies in animals, because it also comprises a higher level feeling of horror or terror [51]. While this is a theory that's still under debate, it states the importance of approaching fear in humans - an opportunity to explore the interaction between the basic "defensive circuits" and higher level "cognitive circuits".

An important question arising from the differentiation between a basic "defensive circuit" and a higher level "cognitive circuit"is the boundary between fear and decision making. while we believe that fear, beyond simple reflexes, (especially fear in humans) intrinsically involves functional components that optimizes an organism's behavior [66] , when we get to higher level decision making which involves a lot

of conscious deliberations, both the behavioral and neural phenomenon we observe will be drifting away from fear per se. Thus, throughout the line of research, our investigation constitutes a combination of fear-induced behavior and the corresponding decision making behavior. As will be stated later, we admit and are well aware of an overlap between the fear circuit and the decision making related circuit. When we are investigating the basic defensive circuits, we are closer to the former; while we are approaching the higher level cognitive circuits, we are closer to the latter. Thus the title of the thesis: decision making under threat.

1.2 Ecologically inspired paradigm and real-life validity

Approaching the question from an ecological perspective is also advantageous. Indeed, in the field of ecology, a wealth of behavioral measures such as Flight Initiation Distance (FID) and Margin of Safety (MOS) has been developed to study organism's sensitivity to threat at an individual level. Theoretical models including This offers us the tool on the behavioral level.

However, this still leaves us the question of how much validity in real life we could obtain by using ecologically inspired paradigms. For example, in chapter 1, we used a computer game where participants play as a prey while an AI controlled predator aims to catch him/her. It makes sense for an organism to display fear and panic when a life-threatening scenario like prey-predator interaction happens in the natural world. However, for humans, such interactions no longer exists in modern society. And in our game, a real death scene upon getting caught by the predator is replaced by a mild shock to the wrist. How much validity and applicability could we achieve with somewhat "gimmicky" paradigms like this?

I would argue that real fear is still manifested in our paradigms, since real time calculation of utility happens throughout the process. A life threatening scene will have an extreme decrease in utility - on the other hand, an electrical shock, though mild, will also impose a less severe decrease in the utility function during decision making. The abstraction of prey-predators interactions function well as a stimulant to bring out the survival optimization process, of which fear is an indispensable part. Though not fully generalizable to life-threatening scenarios, our paradigms still capture the core process of threat avoidance in real life situations.

For the brain imaging Theoretical and neuroanatomical models support the existence of an interconnected defensive survival circuitry that is remarkably preserved across species [13, 70, 80, 81, 90]. Under the conditions of immediate danger, the "reactive

fear" circuitry is evoked. This circuitry includes the midbrain periaqueductal gray (PAG), central amygdala (CeA), hypothalamus, and the midcingulate cortex (MCC), which relay, update, and initiate essentially innate reactions including flight and freezing [35, 51, 67, 81, 94, 104]. Conversely, the ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC), hippocampus, and basolateral amygdala form a collective set of regions that constitutes the "cognitive fear" circuitry that promotes more complex information processing involved in behavioral flexibility, internal risk assessment, and cognitive avoidance strategies [26, 60, 61, 70].

It is important to note that we are not advocating an absolute, static dichotomy of neural networks associated with reactive and cognitive fear. Rather, we strive to identify a dynamic shifting betweenthe regions under various threat scenarios. Regions identified above are more like the extreme markers of a broader spectrum.

Next, I present an outline of the thesis. The basic rationale behind this line of research is consistent throughout all the projects: we identify a novel measurement of threat sensitivity in ecology research; the said measurement will then be converted and refitted for studies in humans; Finally, we obtain behavioral and neural markers of the novel measurement in humans.

1.3 Flight initiation distance and optimal strategy of escape

We first investigate how spatial and temporal distances influence people's decisions under threat. Animals go though various states including pre-encounter, post-encounter and circa-strike [29], and we believe humans do so too. As the threat switches between various attacking positions, we expect to observe differential behavioral and neural patterns from humans.

Flight initiation distance (FID), the distance at which an organism flees from an approaching threat, is an ecological metric of cost–benefit functions of escape decisions. We adapted the FID paradigm to investigate how fast- or slow-attacking "virtual predators" constrain escape decisions. We show that rapid escape decisions rely on "reactive fear" circuits in the periaqueductal gray and midcingulate cortex (MCC), while protracted escape decisions, defined by larger buffer zones, were associated with "cognitive fear" circuits, which include posterior cingulate cortex, hippocampus, and the ventromedial prefrontal cortex, circuits implicated in more complex information processing, cognitive avoidance strategies, and behavioral flexibility. Using a Bayesian decision-making model, we further show that optimization of escape decisions under rapid flight were localized to the MCC, a

region involved in adaptive motor control, while the hippocampus is implicated in optimizing decisions that update and control slower escape initiation. These results demonstrate an unexplored link between defensive survival circuits and their role in adaptive escape decisions.

It is still unclear, however, how anxiety and its neural substrates relate to these distinct defensive survival circuits. We tested whether individual differences in trait anxiety would impact escape behavior and neural responses to slow and fast attacking predators: conditions designed to evoke "cognitive" and "reactive" fear, respectively. Behaviorally, we found that trait anxiety was not related to escape decisions for fast threats, but individuals with higher trait anxiety escaped earlier during slow threats. Functional MRI showed that when subjects faced slow threats, trait anxiety positively correlated with activity in the vHPC, mPFC, amygdala and insula. Further, the strength of the functional coupling between the vHPC and mPFC was correlated with the degree of trait anxiety. These findings suggest that anxiety plays little or no role in escape under conditions of proximal threat. Instead, anxiety affects "cognitive" fear circuits that are involved in volitional strategic escape.

1.4 Margin of safety and preemptive decision under uncertainty

We next investigate how the level of uncertainty of the threat influence people's decisions under threat. Unlike the previous sections where threat emerges at different spatial/temporal distances, the mean value of threat appearance stays the same. However, their attacking behaviors come in with different levels of uncertainty. We expect a more conservative behavioral strategy when facing more uncertain threats.

Humans, like many other animals, preempt danger by moving to locations that maximize their success of escaping future threats. Ethologists have shown that prey select appropriate spatial margin of safety (MOS) between the predator and safety refuge, while human studies demonstrate that proximity to safety decreases subjective fear. Here, we test the idea that volitional spatial MOS decisions, a form of prospective avoidance, result in participants placing themselves closer to safer locations when facing more intense and unpredictable threats. Using multivariate pattern analysis on our fMRI data, we show that key parts of the cognitive fear circuitry are evoked when making safety decisions and may code for the intensity and volatility of the threat. Multivoxel parttern analysis revealed activations in regions including the mPFC, hippocampus, amygdala and insula. Specifically, while the insula and hippocampus non-selectively code decisions for both normal

and volatile attacks, the amygdala and mPFC codes volatile and normal attacks selectively. Our data suggest that when preempting impending danger, a distinct set of parallel cortical-subcortical regions code for the threat intensity, uncertainty, and the decision to shift closer to safety.

1.5 Decisions under threat biased by collaborators' reputation

Finally, we investigate how social factors such as reputation, a information source coming from a thir party, can influence people's decision making under threat.

Humans look to others for advice when making decisions under uncertainty. Rational agents, however, do not blindly seek information, but often consider the quality of its source before committing to a course of action.

Informational social influence theory posits that under conditions of uncertainty, we are inclined to others for advice. This leaves us remarkably vulnerable to being influenced by others' opinions or advice. Rational agents, however, do not blindly seek and act on arbitrary information, but often consider the quality of its source before committing to a course of action. Here, we ask the question of whether a collaborator's reputation can increase their social influence, and, in turn, bias perception and anxiety under changing levels of uncertainty. Human male and female participants were asked to provide estimations of dot directions using the random dot motion (RDM) perceptual discrimination task and were paired with transient collaborators of high or low reputation whom provided their own estimations. The RDM varied in degrees of uncertainty and joint performance accuracy was linked to risk of an electric shock. Despite providing identical information, we show that collaborating with a high reputation compared with a low reputation partner, led to significantly more conformity during the RDM task for uncertain perceptual decisions. Consequently, high reputation partners decreased the subjects' anxiety during the anticipatory shock periods. fMRI data showed that parametric changes in conformity resulted in increased activity in the ventromedial PFC, whereas dissent was associated with increased activity in the dorsal anterior cingulate cortex (dACC). Furthermore, the dACC and insula, regions involved in anticipatory pain, were significantly more active when collaborating with a low reputation partner. These results suggest that information about reputation can influence both cognitive and affective processes and, in turn, alter the neural circuits that underlie decision-making and emotion.

1.6 Thesis Overview

Inspired by the questions discussed before, and building upon previous literature, this thesis strives to understand the behavioral and neural mechanisms behind human decision making process under threat.

In Chapter 2, we explore the relationship between spatial/temporal distance of the threat, the corresponding fear response and the subsequent decisions. By employing flight initiation distance and the prey-predator interaction widely studied in ecology research, we simulate real life scenarios where people need to make a trade off between potential threat (specifically, an aversive yet tolerable electrical shock in this experiment). This will help us understand how spatial/temporal immediacy impacts decision making behaviors, together with the neural mechanisms behind it.

Chapter 3 further looks at anxiety's effect on decision making behavior implicated in the flight initiation distance paradigm. STAI trait anxiety inventory was used to measure the level of anxiety. We intend to test whether individual differences in anxiety level would influence behavioral and neural responses to the "reactive" and "cognitive" fear circuits respectively, as defined in Chapter 2.

In chapter 4, we borrow another useful measurement from the prey-predator interaction system: Margin of Safety, where an organism actively selects an active zone relative to its safety refuge. This affords us to investigate preemptive avoidance behavior, in addition to online fear response as discussed in Chapter 3. Moreover, instead of looking at the effect of temporal/spatial distance, we vary the fear stimulus such that they display different levels of uncertainty. Thus, we aim to uncover the neural mechanism behind preemptive avoidance behavior under uncertainty.

Finally, we discuss how social factors (here specifically, reputation from a task partner) could impact decision making under threat.

Chapter 2

HOW COGNITIVE AND REACTIVE FEAR CIRCUITS OPTIMIZE ESCAPE DECISIONS IN HUMANS

2.1 Introduction

Survival depends on the adaptive capacity to balance fitness-promoting behaviors, such as copulation and foraging, with the omnipresent risk of lethal predatory attack [70]. In the field of behavioral ecology, this balance between survival behaviors is depicted by economic models of flight initiation distance (FID), which capture risk functions by measuring the distance at which an organism flees from an approaching threat, while considering the cost of fleeing [29, 113]. A wealth of ethological literature demonstrates that prey are remarkably adept at escape and make decisions based on the predator's directionality, lethality, velocity, and previous experience with the predator [98]. In addition to its capacity to measure escape decisions, FID is a well-established index of threat sensitivity, resulting in large variability within and between species [98]. Despite FID measures being applied to a large variety of taxa, this reliable measure has not been used to identify heterogeneity in threat sensitivity or escape decisions in humans, and the neural circuits remain unexplored.

Theoretical and neuroanatomical models support the existence of an interconnected defensive survival circuitry that is remarkably preserved across species [13, 70, 80, 81, 90]. Under the conditions of immediate danger, the "reactive fear" circuitry is evoked. This circuitry includes the midbrain periaqueductal gray (PAG), central amygdala (CeA), hypothalamus, and the midcingulate cortex (MCC), which relay, update, and initiate essentially innate reactions including flight and freezing [35, 51, 67, 81, 94, 104]. Conversely, the ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC), hippocampus, and basolateral amygdala form a collective set of regions that constitutes the "cognitive fear" circuitry that promotes more complex information processing involved in behavioral flexibility, internal risk assessment, and cognitive avoidance strategies [26, 60, 61, 70]. Although few behavioral ecologists have considered the neurophysiology underlying escape decisions, some have proposed similar dichotomies suggesting that fast, but inaccurate, decisions are processed by subcortical regions, while slow, but accurate, decisions are processed by cortical system [72, 106]. Under natural conditions, both cognitive-

and reactive-fear circuits work in harmony by adaptively switching between survival circuits to engage the most optimal strategy to maximize escape [5, 26, 67, 70].

Excitation and inhibition between these circuits is determined by the spatiotemporal distance to the threat [13, 37, 60, 67]. For example, distant threat often results in freezing and threat assessment, yet when the threat is close, active flight will be observed [13]. Distance to the threat, therefore, is crucial in choosing the best escape strategy. Evidence suggests that this pattern is conserved across various species. In humans, active escape tasks have been used, where the goal of the subject is to escape from a virtual looming threat with the capacity to chase, capture, and shock the subject in a virtual maze. Functional MRI (fMRI) results show that when a threat is distant, there is increased activity in the vmPFC, PCC, and basolateral nucleus of the amygdala. Conversely, as the threat moves closer, there is a switch to increased activity in the CeA and PAG [71–73]. However, these, and related studies, have failed to investigate the neural basis of escape decisions (i.e., flight initiation) or to examine the computational mechanisms that underlie escape decisions to changing attack distances.

We developed a paradigm to investigate how the defensive survival circuitry facilitates escape decisions when subjects encounter fast- or slow-attacking threats (Figure 2.1). In this task, participants encountered virtual predators of three colors, each representing different attack distances (ADs). On each trial, the actual AD was drawn from a Gaussian distribution that was unique to the particular predator type. Fast-attacking predators (i.e., predators that attack from a larger distance) were characterized by the virtual predator quickly switching from slow approach to fast attack velocity, therefore requiring the subject to make quick escape decisions. On the other hand, slow-attacking predators (i.e., predators that attack from a smaller distance) slowly approached for longer time periods, resulting in larger buffer zones leading to more time to strategize escape. All types of predators loomed and sped up at the same rate, and only differed in their timing of attack.

The goal of the task was to escape from the predator while, at the same time, attempting to acquire as much money as possible by fleeing as late as possible (Figure 2.1). Using this task, we proposed several hypotheses: (i) for fast escape decisions, we expected to see activity in the reactive fear circuitry, while slow escape decisions would reveal more pronounced activity in the cognitive fear circuitry. In addition, (ii) using a Bayesian decision-making model where subjects' preference to reward and avoidance to punishment were considered, we predicted that the reactive

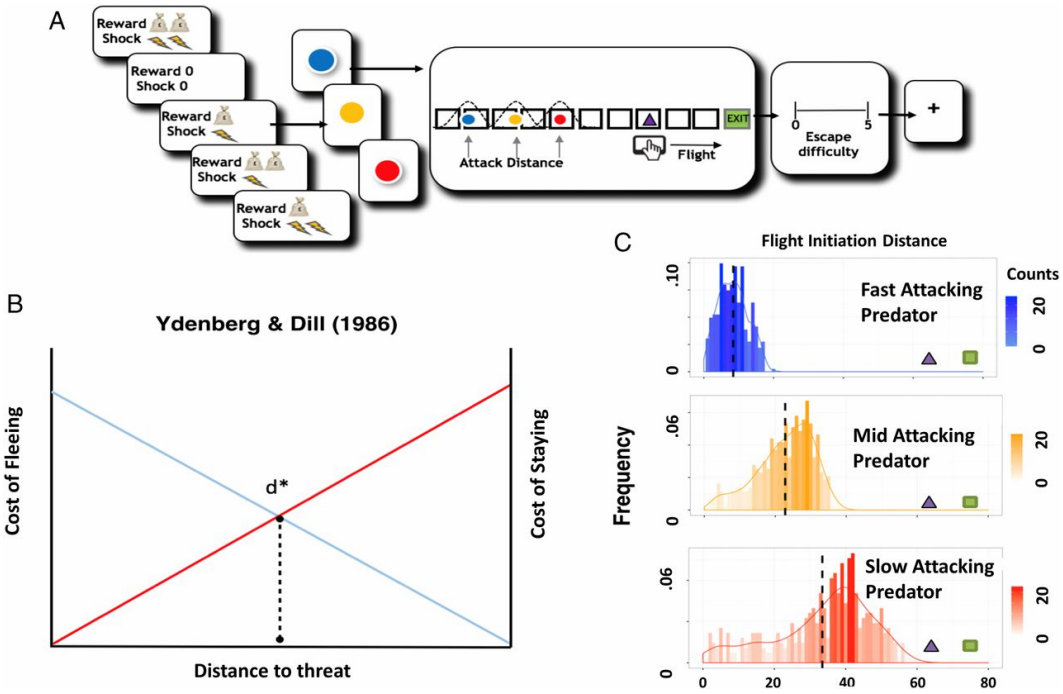

Figure 2.1: Experimental procedures, Ydenberg and Dill model (1), and distribution of escape decisions. (A) Subjects are told whether their decisions will result in high or low reward or shock. They are then presented with the image of the virtual predator where the color signals the attack distance (2 s) (e.g., blue, fast; red, slow). After a short interval, the virtual predator appears at the end of the runway and slowly moves toward the subject's triangle. After an unspecified amount of time (e.g., 4–10 s), the artificial predator will attack the subject's virtual triangle exit (i.e., attack distance). To escape, the subject must flee before the predator attacks. If the subject is caught, they will receive a tolerable, yet aversive, shock to the back of the hand. Trials end when the predator reaches the subject or the exit. To motivate longer fleeing time, the task will include an economic manipulation, where subjects will obtain more money the longer they stay in the starting position, and lose money the earlier they enter the safety exit. After each trial, the subject is asked to report how difficult they found it to escape the virtual predator (4 s). (B) Modified schematic representation from the model proposed by Ydenberg and Dill (3). As the distance between the prey and the predator decreases, the cost of fleeing decays, while the cost of not fleeing rises. D* represents an optimal point where the prey should flee. (C) Histograms showing the distribution of subjects' flight initiation decision (FID) choices for early-, mid-, and late-attacking predators, respectively. The x axis represents FID, while the y axis represents frequency of choice.

and cognitive fear circuits would play a corresponding role in facilitating fast and slow escape decisions, respectively.

2.2 Methods

Participants

A total number of 30 subjects completed informed consent in accordance with the guidelines of the Columbia University IRB and were remunerated for their participation. Data from one subject were lost due to computer error. One additional subject was excluded due to excessive movement during the scan. Our final sample consisted of 28 subjects (17 females; 11 males; age, 25.4 ± 7.3 y).

Experiment Design

Subjects were scanned while they viewed stimuli on a screen that displayed a 2D runway, with a virtual predator "attacking" from the left entrance. In the current paradigm, the goal of the subject was to escape the attack from a certain virtual predator, by pressing a button at the desired timing. Once the button was pressed, a triangle representing the subject started moving toward a "safety exit." Subjects gained reward if they escaped to safety before the predator caught them; on the other hand, they were given a mildly aversive electric shock if they were caught. The key was to choose the right timing to flee: acquire the maximum amount of reward while still escaping the virtual predator. Reward in each trial linearly scaled with time spent before pressing the button. The longer subjects stayed in the starting position (the smaller the FID), the more reward they got. However, if the subjects stayed for too long, they could have been caught, which would have resulted in both a loss of all reward for the current trial and the administration of an electric shock. However, they still maintain the cumulative reward they received from previous trials.

The runway has a total length of 90 units, where the subject's triangle is placed 10 units to the safety exit. While in the approaching mode, the predator oscillates toward the subject's triangle at a speed of 4 units/s; while in the chasing mode, the predator proceeds with a speed of 10 units/s. There are 96 trials, factorially divided to cover different predator attack distances, shock levels, and reward levels [3 × 3 × 3; three types of predators; three levels of shock (0, low, and high levels of shocks); and three levels of reward (0, low, and high levels of reward)]. In the high-shock condition, subjects received two shocks instead of one. In addition, in the high-reward condition, subjects received twice the original reward if they escaped. The control condition was the zero-reward and zero-shock condition. Subjects were first presented with a screen indicating which type of predator and shock/reward level will be presented in the next trial for 2 s. This shock/reward indicator informed the next four trials. Next, the trial began, where subjects observed an artificial predator

slowly looming toward the triangle representing themselves. After a designated time period, which was learned by the participant, the artificial predator will attack by speeding up when it reached the attacking position. To make sure reaction time played no role in FIDs, we manipulated the speed of the predator and the subject so that, once the threat speeded up to attack, it was impossible to escape after that time point. After the trial, subjects were required to rate the difficulty of escape using a visual analog (1–5) scale.

After the first 48 trials, the assignment of predator–color relationship was altered to introduce novelty and avoid the (habitual) fixation of subject strategies. Same colors were reassigned among the predators (e.g., the original fast-attacking predator changed color from blue to red).

Before the start of the actual experiment, subjects went through a brief practice session of eight trials to familiarize themselves with the paradigm. In the practice session, subjects played the same game, but the predators' attack distances were drawn from different distributions other than the ones used in the actual experiment. The level of the electric shock was calibrated according to subjects' level of tolerance (self-reported to be aversive, yet not painful). With a 1–10 scale, the average calibrated shock level was 6.4 (mean, 6.4; SD, 1.3).

Behavioral Analysis

Due to the relative simplicity of our task and exposure to a practice session, subjects' performance reached saturation very quickly after the beginning of the experiment. By "saturation," subjects quickly formed their own patterns of choice making and carrying out the rest of the experiment. Thus, instead of looking at trial-by-trial changes of the FID, we focused on the differences of FID between different predator conditions, and on the approaches subjects' learning behavior by a Bayesian decision-making model.

The subjects' choice of FID, reward from the trial, and escapability ratings were collected on each trial. We used repeated-measures three-way ANOVAs (of predator type by reward level by shock level) to assess differences in FID, reward, and escapability ratings between the various conditions.

Acquisition and analysis of fMRI data

All fMRI data were acquired using a GE Discovery MR750 3.0 T scanner with 32-channel headcoil. The imaging session consisted of two function scans, each twenty

minutes, as well as a high-resolution anatomical T1-weighted image (1mm isotropic resolution) collected at the beginning of each scan session. For functional imaging, interleaved T2*-weighted gradient-echo echo planar imaging (EPI) sequences were used to produce 45 3-mm-thick oblique axial slices (TR = 2 sec., TE = 25 ms, flip angle = 77°, FOV = 192 x 192 mm, matrix = 64 x 64). Each functional run began with five volumes (1000 msec) before the first stimulus onset. These volumes were discarded before entering analysis to allow for magnetic field equilibration. Stimulus were presented using Cogent (matlab-based package). Participants viewed the screen via a mirror mounted on the head coil, and a pillow and foam cushions were placed inside the coil to minimize head movement.

Analysis of fMRI data was carried out using scripted batches in SPM8 software (Wellcome Trust Centre for Neuroimaging, London, UK; http://www/l.ion.ucl.ac.uk/spm) implemented in Matlab 7 (The MathWorks Inc., Natick MA). Structural images were subjected to the unified segmentation algorithm implemented in SPM8, yielding discrete cosine transform spatial warping coefficients used to normalize each individual's data into MNI space. Functional data were first corrected for slice timing difference, and subsequently realigned to account for head movements. Normalized data were finally smoothed with a 6-mm FWHM Gaussian kernel.

Preprocessed images were subjected to a two-level general linear model using SPM8. The first level contained the following regressors of interest, each convolved with the canonical two-gamma hemodynamic response function: a 2-second box-car function for the onset of the trial (where the color of the incoming predator is shown); a 4-8 second (duration jittered) box-car function from the onset to 2s before when subjects make the flight decision; a 2-second boxcar (function for the phase before subjects make the flight decision; a 4-8 second (duration jittered) box-car function for the remainder of the trial. Mean-centered trait anxiety ratings, escapability ratings and parameters in the Bayesian decision model were included as orthogonal regressors. In addition, regressors of no interest consisted of motion parameters determined during preprocessing, their first temporal derivative and discrete cosine transform-based temporal low frequency drift regressors with a cutoff of 192-seconds.

Beta maps were used to create linear contrast maps, which were then subjected to second-level, random-effects one-sample t tests. In addition, a flexible factorial model was used to examine the main effects of predator type, reward level and shock level. Interaction effects between predator type, reward level and shock level were also examined using the factorial model. The resulting statistical maps were

thresholded at P < 0.05 corrected for multiple comparisons (false discovery rate [FDR] corrected). A flexible factorial model was used to examine the interaction effects between predator type, reward level and shock level. The threshold for those specific contrasts was set at p < 0.05 (FDR corrected).

A hypothesis driven regions of interest (ROI) analysis was performed after the whole brain analysis for regions with strong a priori spatial hypotheses. The ROI analysis was performed using regions associated with the processing of fear, threat and decision making. Independent ROIs were chosen from previous research showing similar effects. The threshold for these analyses was set at p < 0.05, small volume correction (SVC).

The functional connectivity analysis was performed for the response phase (escape decision) using a generalized psychophysiological interactions (PPI) approach. The connectivity analysis was carried out based on the [predator condition > control condition] contrast.

Bayesian decision making model

For details on the Bayesian decision making model, please refer to the supplementary material of this published paper within the published content [3].

2.3 Results

Behavior

We first examined the behavioral data by applying a repeated-measures, three-way ANOVA (predator type by reward level by shock level) for escape responses (e.g., FIDs). Results showed a main effect of predator type [$F(2,54) = 82.59$, $P < 0.001$]. Post hoc comparisons for the predator type by shock level interaction revealed that the difference in FID choices between high and low shock levels exists only in the slow-attacking predator condition ($P = 0.013$). This shows that subjects took the level of potential danger into consideration while choosing FID (more risk-averse when shock is higher), but only in the slow-attacking threat where there was time for strategic avoidance. The same repeated-measures three-way ANOVA was performed for escape difficulty ratings. A main effect of predator type was found [$F(2,54) = 49.77$, $P < 0.001$], showing that subjects estimate fast-attacking predator as the most difficult predator type to escape (all post hoc comparisons: $P < 0.001$). Significant interactions were found for predator type by shock level [$F(2,54) = 13.68$; $P < 0.001$] and predator type by reward level [$F(2,54) = 4.39$; $P = 0.017$]. For the predator type by shock level interaction, we found that rating was higher in the high

shock condition, but only in the slow-attacking predator (post hoc comparison: P < 0.001). This is intriguing because the predator's attack distance is identical at both shock levels, yet subjects perceived the threat to be more difficult to escape in the high-shock condition (please refer to the SI Appendix, Fig. S1 in the published content) [3].

Neural basis of fast and slow escape decisions

We next investigated the neural basis of the escape decisions for the fast- and slow-attacking threats. To control for timing differences between conditions, besides modeling the rest of the trial as a boxcar function, we specifically looked at the 2 s before the FID button press as a period where subjects form their final decisions. We chose to time lock 2 s before the flight initiation decision for several reasons: (i) it allowed us to examine the neural ramping up of the flight initiation, (ii) it controlled for the contamination of outcome, and (iii) it reduced the amount or trials that would be lost for the fast-attacking condition. Also, to control for any confounds of pain, we excluded the caught trials (number of caught trials: fast-attacking predator, mean, 8 ± 3; mid-attacking predator, mean, 5 ± 2; slow-attacking predator, mean = 4 ± 1), using these events as regressors of no interest. As the mid-attacking condition was a priori used as an anchor for the fast- and slow-attacking threats, we focused on activity for the fast and slow attacking predators. A whole-brain analysis was first performed to locate regions associated with decisions under reactive fear (fast-attacking predators) and cognitive fear (slow-attacking predators). Detailed regions of activation can be found in SI Appendix, Tables S1 and S2 in the published content [3]. As shown in Figure 2.2, data extracted from a priori and independent anatomical regions of interest (ROIs) of PAG, MCC, PCC, hippocampus, and vmPFC, were differentially activated for the different predator conditions.

To confirm the dissociation between the reactive and cognitive fear systems (represented by PAG and vmPFC, respectively), we computed a two-way ANOVA (region by predator type) using signal change drawn from independent ROIs from PAG and vmPFC. There was a main effect of region ($F = 5.77$, $P = 0.017$) and a significant interaction between region and predator type ($F = 11.50$, $P < 0.001$). For the [fast-attacking predator > control] contrast, we observed increased activity in PAG and MCC. A direct comparison between high and low shock levels for the fast-attacking predator revealed increased activity in the PAG, suggesting that PAG is evoked when the threat is high (please refer to SI Appendix, Table S7 in the published content) [3].

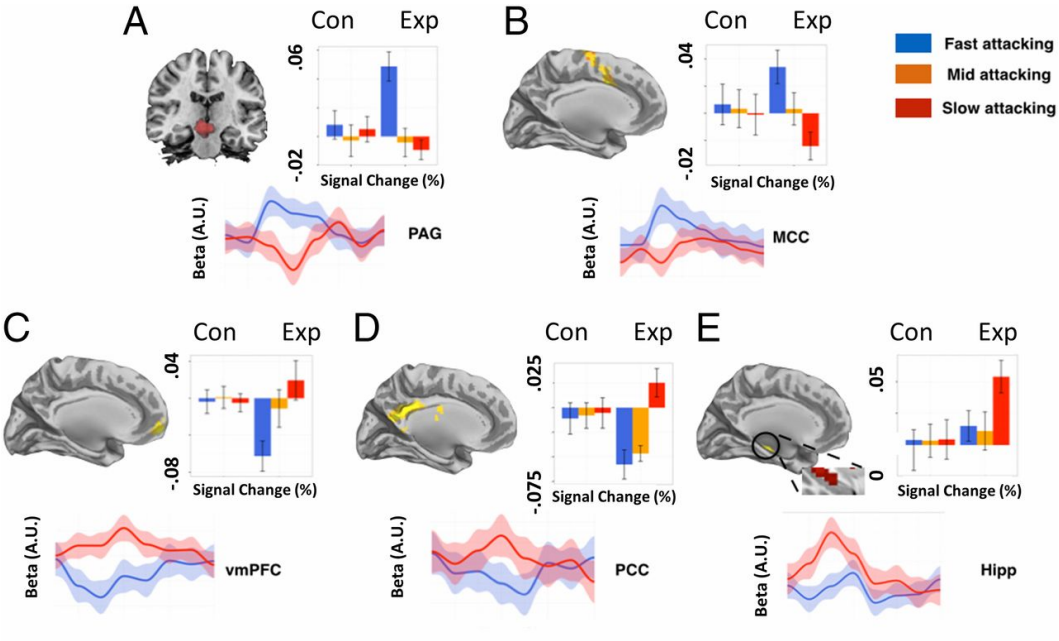

Figure 2.2: Main regions of interest and signal changes associated with fast- and slower-attacking threats. Parameter estimates and time series extracted from (A) midbrain, (B) MCC, (C) vmPFC, (D) PCC, and (E) the hippocampus. Activations shown in the graph show clusters from the whole-brain activation, while the signal change data were extracted from independent anatomical ROIs. The Upper graph displays parameter estimates. The y axis represents percent signal changes, and the x axis is the predator type. The Lower graph displays time series extracted in a course of 16 s. Blue line: fast predator; red line: slow predator. The beginning of the time series represents the time point when the FID event comes online.

On the other hand, the [slow-attacking predator > control] contrast revealed increased activity in the cognitive fear circuitry including the vmPFC, PCC, and the hippocampus. While no amygdala was observed for the main contrast, a direct comparison between high and low shock levels in the slow-attacking predator condition showed increased activity in the amygdala and hippocampus (please refer to SI Appendix, Table S8 in the published content) [3]. To further disentangle the effect and increase the sensitivity of the analysis, we extracted the signal changes and BOLD-signal time series from the predefined ROIs (i.e., PAG, MCC, vmPFC, PCC, and hippocampus), regions that have previously been associated with fear, anxiety, and decision making under stress [71]. A conjunction between fast- and slow-attacking threats showed that the medial dorsal thalamus (MDT) was commonly activated. Although this is an exploratory finding, it is intriguing because MDT is directly or indirectly connected to both fear circuits, since stimulation of

the MDT results in depression or potentiation of both circuits and it is thought to play a role in behavioral flexibility [49, 107].

Computations that support escape decisions

To explore how the observed FIDs might be understood in terms of rational decision making (i.e., the costs and benefits of flight), we developed a Bayesian decision-making model. The process by which subjects make escape choices under different predator ADs can be decomposed to two steps: (i) predicting predators' distribution of attack distances, by learning from experience; and (ii) choosing an FID by comparing the money that can be possibly obtained against the potential risk of shock for each possible FID, in expectation over the predicted attack distance distribution and informed by the individual's subjective preference levels for shock vs. money. We assume a Bayesian ideal observer model of subjects' learning to estimate the attack distances of different predators from trial-by-trial experience. FID choices are then determined (with softmax noise) by computing the expected utility for each possible escape distance. We then calculated the distance between utility resulted from subjects' actual FID and the predicted Bayes ideal FID, which is considered a measure of optimal performance. The modeling results are shown in Figure 2.3 and SI Appendix, Figs. S3 and S4 of the published content [3]. Details of the model are explained in SI Appendix, SI Text of the published content [3].

We next examined the neural circuits that correlated with each subjects' preference parameters in the Bayesian decision model. For a rational player, the preference for reward should be positive, while the preference for shock should be negative. Thus, greater reward or shock sensitivity here corresponds to larger (positive) $\beta 2$ and smaller (negative) $\beta 1$ amplitudes. The parametric modulation analysis over the [predator > control] contrasts revealed that, for the fast-attacking predator condition, higher reward sensitivity was associated with activations in bilateral putamen, while higher shock sensitivity was associated with engagement in PAG and bilateral insula. On the other hand, for the slow-attacking predator condition, right caudate was found to be associated with higher reward sensitivity, while PCC was found to be associated with higher shock sensitivity. A display of the activated regions can be found in SI Appendix, Fig. S3 of the published content [3]. A detailed layout of the activated regions can be found in SI Appendix, Tables S9 and S10 of the published content [3].

Next, to investigate what neural circuits are responsible for the optimization of

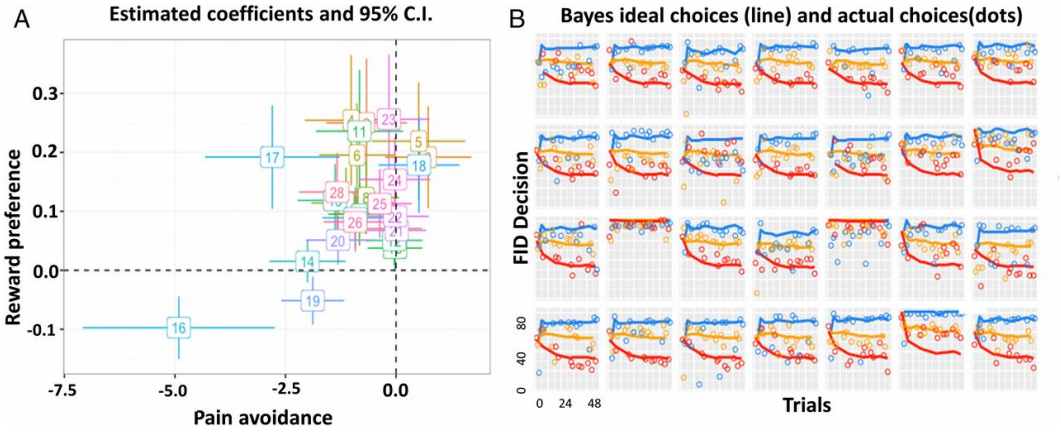

Figure 2.3: Visualization of Bayesian modeling results. (A) Estimated coefficients for each subject for the first scanning session, along with 95 percent confidence intervals. The x axis represents the pain coefficient $\beta 1$ in the utility function, and the y axis represents the monetary reward coefficient $\beta 2$. For a rational player, $\beta 2$ should be positive (seeking money), and $\beta 1$ should be negative (avoiding shock). (B) Model fit to observed FIDs for the first scanning session. The x axis represents trial numbers, and the y axis represents FID. Ideal FID choices predicted by the ideal Bayesian observer (lines), subjects' actual FID choice (dots). Average values of reward preference and shock avoidance of the two scanning sessions were used as parametric modulators for the fMRI analysis. Data for session 2 can be found in SI Appendix, Fig. S4 of the published content. [3]

.

escape decisions, we considered a measure of performance optimality related to the per-trial spread between subjects' actual and Bayesian ideal FIDs. In particular, we computed the difference between the actual trial-specific utility U(FID) and the maximum (Bayes optimal) utility the subject could possibly get on the trial [U(FID)max], given their estimated subjective utilities. A smaller difference (e.g., less regret relative to ideal) implies more consistent Bayesian decision making; variation around the ideal FID will increase the difference. The differences on every trial were entered as a subject-level parametric modulator separately under each [predator > control] conditions. For the fast-attacking predator condition, we found that better Bayesian decisions (smaller distance to ideal) was associated with activity in MCC, middle frontal gyrus, and superior motor cortex. On the other hand, better Bayesian decision making in the slow-attacking predator condition was found to be associated with activity in bilateral hippocampus, as shown in Figure 2.4.

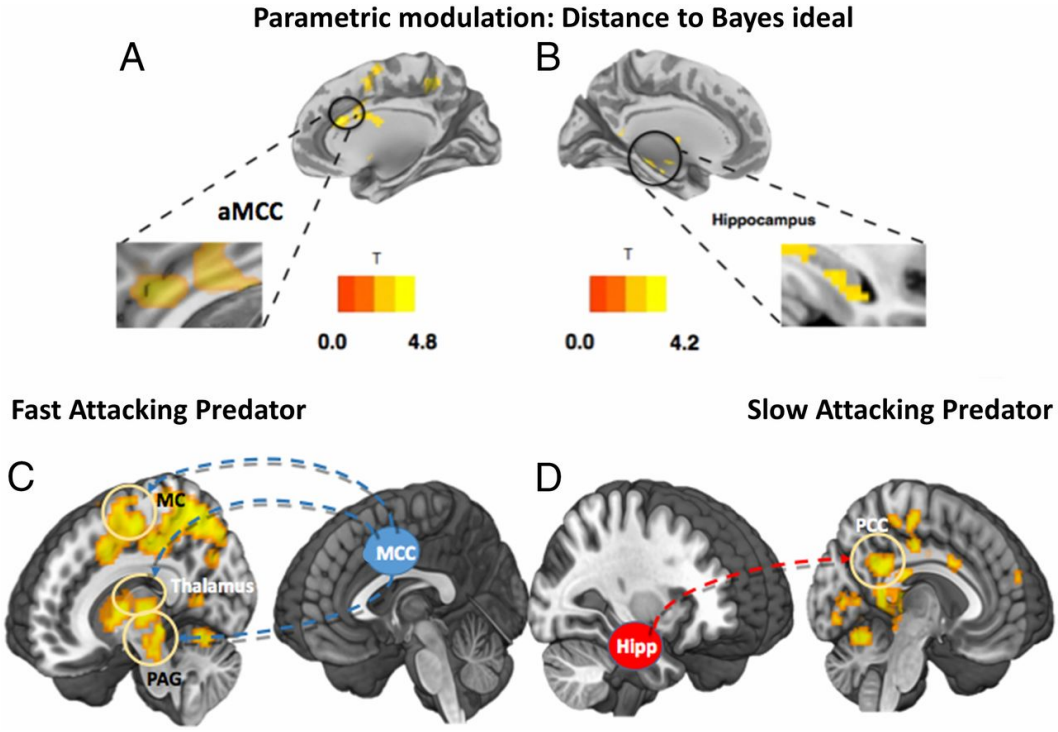

Figure 2.4: Regions, and their connectivity, associated with parametric modulation of "distance to ideal." (A) Brain regions associated with increased Bayesian decision optimality in the fast AD condition. Better decision making was associated with increased activity in MCC and superior motor cortex. (B) Brain regions associated with decreased distance (increased Bayesian decision optimality) in the slow AD condition activated regions include bilateral hippocampus and bilateral caudate. A display of the correlation results can be found in SI Appendix, Table S11 of the published content [3]. (C) Connectivity analysis using MCC as seed over the contrast [fast predator > control]. Positive connectivity was found between MCC, motor cortex (MC), thalamus, and the PAG. (D) Using the hippocampus as seed over the contrast [slow predator > control], positive connectivity was found between the hippocampus and PCC.

Functional connectivity between computationally defined regions

To investigate the interplay among the brain regions involved in escape decision optimization, a functional connectivity analysis was performed for the response phase (escape decision) using a generalized psychophysiological interactions approach [59]; to confirm the patterns observed in the whole-brain flexible model, we first adopted independent seed regions of MCC and hippocampus from previous research [72]. For the contrast of [fast-attacking predator > control], we showed a significant coupling between the MCC seed, the PAG, motor cortex, and bilateral thalamus. For the contrast of slow predator > control, we showed a significant coupling between

the hippocampus seed and PCC. This suggests that, when the subjects are provided time for decision flexibility, they use a search-and-employ approach that prepares them for action, as shown in Figure 2.4.

2.4 Discussion

We have demonstrated that subjects apply different nodes of the survival circuitry when escaping fast- and slow-attacking threats. Our analysis revealed increased activity in reactive-fear circuits, namely the PAG and the MCC for the fast attacking predators, regions that are implicated in motor response to fast and imminent threats. Supporting comparative work [6], connectivity analysis revealed a significant couple between the MCC and PAG. Recent animal work has also shown the optogenetic activation of glutamatergic neurons in the dorsal lateral PAG induced motor responses [e.g., flight [104]]. The MCC is also a critical component of the defensive survival circuitry and has afferent projections to the ventral striatum, receives efferent signals from the medial dorsal thalamus, and has bidirectional projections with the amygdala [94]. It has also been suggested that control signals in the MCC may resolve conflict between defensive strategies (e.g., freezing or fleeing). This has led to the theory that the cells in the MCC are involved in linking motor centers with defensive circuits [94].

Our analysis for the slow-attacking threat contrast revealed activation in three key areas of the cognitive fear circuitry involved in more complex information processing—the vmPFC, hippocampus, and PCC. Structural and function connectivity between these structures has been shown in humans and primates, supporting conserved pathways across species [90]. Primate research has found that the primate PCC responds to risky decision making and scales with the degree of risk [57]. The PCC is also correlated with a salience signal reflecting the deviation from the standard option, suggesting a role in the flexible allocation of neural resources [39]. A function of the PCC may be to harvest information for escape decisions under conditions of protracted threat. This fits with the proposal that, through its connections with the hippocampus, the PCC may integrate memory guided decisions with current decision processes that may involve a "preparation for action" by anticipating and altering behavioral policies [84].

The vmPFC is also a key player in the defensive survival circuitry. Single-cell recordings in rodents have shown that the mPFC contains "strategy-selective" cells, which are thought to be involved in the coordination of defensive responses [37].

This fits with the idea that the mPFC plays a role in selecting adaptive strategies that are mapped onto motor responses. Indeed, work in humans shows that larger buffer distances are associated with activity in the vmPFC, and decreased activity in these regions is associated with panic-related motor actions [71, 73, 86]. Our data build on these findings by showing that the vmPFC, hippocampus, and PCC form a strategic and flexible decision process [26, 74], when the agent has time to contemplate the best escape action. Our findings tentatively support the role of complex cortical information processing circuits (i.e., cognitive fear circuits) in "slower" escape decisions associated with flexible and strategic avoidance through internal risk assessment that involve model-based memory search [52, 61]. This fits with a model-based perspective where actions are deliberative and employ a cognitive-style representation, which is an internal map of events and stimuli from the external world, and take prospective assessment of the consequences of an action [27]. Thus, the cortical activity observed here could represent "reflective computations" associated with higher information processing and cognitive architecture [84].

Our Bayesian model also provides insights into how the distinct regions of the survival circuits associated with optimal escape. Two core regions were associated with optimal escape: the MCC for the fast-attacking threat and the hippocampus for the slow-attacking threat. While it is accepted that the PAG needs input to make optimal decisions, it is unclear where this input comes from. A few candidates exist; among them is the MCC. The MCC is highly connected to the lateral PAG and according to adaptive control theory is a "central hub" where information about reinforcers are passed to motor control areas to coordinate goal-directed behaviors [94]. Our connectivity results support this conclusion showing that the MCC was coupled with activity in the PAG and the motor cortex. This proposes that the MCC is one candidate region for the integration of current goals and implement aversively motivated instrumental motor behaviors [i.e., when to flee a threat [94]].

Theorists have proposed that the hippocampus computes comparators that assess multiple goals and in turn corrects actions [62] possibly through a flexible constructive process involved in problem solving [38] and predictive mapping. When there is time to gather information, the hippocampus may play a role in drawing on previous threat encounters to form a predictive map and optimize current actions [97]. The hippocampus also plays a role in spatial and temporal "where" and "when" memory and has theoretically been linked to escape decisions and may act to resolve conflict between fitness-promoting behaviors [8, 29, 32, 101, 105, 113]. Our

computational analysis also revealed a Bayesian role for the hippocampus, where it potentially gathers information to optimize directed escape during slow-attacking, but not fast-attacking, threat. Our connectivity analysis did not reveal connections between these regions but did show that the hippocampus was also coupled with activity in the PCC, a region thought to be involved in adaptive decisions [84].

Chapter 3

SLOW, BUT NOT FAST, ESCAPE DECISIONS ARE SWAYED BY TRAIT ANXIETY

3.1 Introduction

Anxiety is often described as an enduring, conscious state of apprehension. Theoretical work[14, 51, 70] proposes that anxiety is an emotional state independent from fear, which is instead evoked when a threat is increasingly proximal, and which ought to be minimally influenced by the anxiety state of the organism [29, 66]. While this is generally well recognized in the non-human animal literature, researchers in the field of human affective neuroscience have paid relatively little attention to the question of whether anxiety and fear have different associated neural circuitry, and under what conditions anxiety might influence defensive behaviors in ecological scenarios. Moreover, recent advances have distinguished different classes of defensive responses which rely on distinct neural circuits, and which may complicate the theoretical relationship between fear and anxiety [3, 72].

Non-human animal research has shown that anxiety states involve a well-defined set of neural circuits [17]. The vHPC and mPFC are of particular interest as they have repeatedly been shown to be recruited during the regulation and representation of anxiety provoking features of the environment [1, 2, 11, 79]. The vHPC has input into the mPFC and it appears to be the interaction between these regions that drives anxiety related behaviors [1]. More recently, CA1 cells in the vHPC have been shown to exhibit stable representations of anxiety provoking environments and these cells drive avoidance behaviors [44].

In humans, functional magnetic resonance imaging (fMRI) has been employed in conjunction with "active escape" paradigms, the goal of which is to evade an artificial predator with the capacity to chase, capture and shock the subject. Studies have shown that when an artificial predator is distant, increased activity is observed in the ventromedial prefrontal cortex (vmPFC) [72]. However, as the artificial predator moves closer, a switch to enhanced activation in the midbrain PAG is observed [72]. More recently, using a novel escape decision task, work from our lab has supported a similar "cognitive" and "reactive" fear differentiation of defensive survival circuits, by showing that fast escape decisions are associated with activity in the PAG [3], a

region shown previously to be involved in reactive flight [72], while slower escape decisions rely on the vHPC, posterior cingulate cortex and mPFC [3], a circuit implicated in behavioral flexibility and internal risk assessment [61].

The vHPC-mPFC anxiety circuit therefore overlaps with the "cognitive" fear circuit recruited during these slower escape decisions [66], but appears to be independent from "reactive" fear regions that are involved with threat under limited time constraints. In general, these "reactive" fear areas (e.g. PAG) have limited interaction with higher level cortical brain regions, thus are unlikely to be implicated in anxiety. Thus, it is possible that while anxiety plays no role during imminent threat (when "reactive" fear circuits are recruited), it may be important within "cognitive" fear circuits, and subsequently affect defensive behavior in the face of less imminent threats.

In order to provide evidence for this possibility, a critical question is whether individual differences in levels of trait anxiety will selectively affect "cognitive" fear circuits during defensive decision making, or whether "reactive" fear circuits are also influenced by the trait anxiety of the individual. Moreover, it is equally important to determine whether there are commensurate changes in survival behaviors and decision making as a result of differences in trait anxiety, as would be expected if anxiety has an ethological origin [70].

To address these questions, we reanalyzed behavioral and neural data collected in our previously published study [3], along with previously unanalyzed trait anxiety data (the Spielberger State-Trait Anxiety Inventory; STAI-Y). In each trial of the behavioral task, participants passively earned money while they encountered virtual predators of three colors, each representing different attack distances (Figure 3.1). These attack distances were drawn from Gaussian distributions that were unique to the particular predator type. Fast attack predators (i.e. far or early attacking) were characterized by the virtual predator quickly switching from slow approach to fast attack velocity, therefore requiring the subject to make quick escape decisions. On the other hand, slow attack predators (i.e. close or late attacking) slowly approached for longer time periods, resulting in larger buffer zones and more time to contemplate escape. (It is important to emphasize that "fast" and "slow" here describe the timing of the predator attack, not the speed of the predators.) The goal of the task was to try and successfully escape, while at the same time maximizing the amount of money earned by fleeing as late as possible (i.e. at the shortest distance from the predator, or flight initiation distance, FID).

Subjects performed this task while undergoing functional magnetic resonance imaging (fMRI) in order to assess the relative contributions of the "reactive fear" and "cognitive fear" networks to their escape decisions, and whether behavior or brain activity in these circuits varied as a function of trait anxiety. Given the theoretical and neural differentiation between "reactive fear" and "cognitive fear", we hypothesized that individuals with high trait anxiety would show preferential activity in the "cognitive fear" circuitry, but not the "reactive fear" circuitry. We also hypothesized that individuals scoring higher in trait anxiety would make earlier escape decisions, but only when there is sufficient time to assess threat.

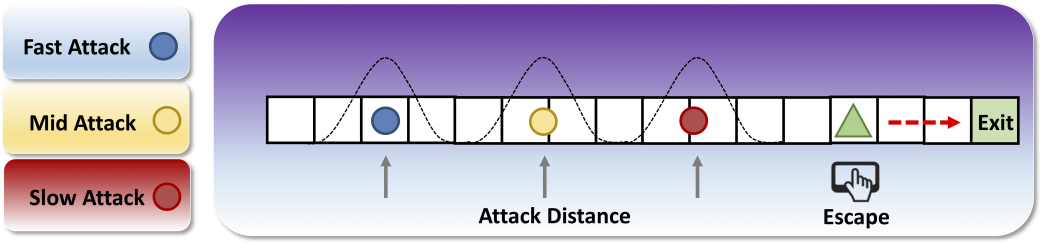

Figure 3.1: Predator escape paradigm. In each trial, participants were presented with a cue indicating the predator type. The predator would appear on the left side of the runway, and slowly move toward the participant (green triangle). Participants passively accrued money while they waited, but at any time could press a button to begin their escape toward the exit. The predator would speed up (attack) at a random distance drawn from the respective Gaussian distributions shown above. If participants were caught by the predator, they would receive a mild electric shock and lose any money accrued on that trial.

3.2 Methods

Participants

30 subjects were recruited according to the guidelines of the Columbia University Institutional Review Board after providing informed consent. This sample size was chosen consistent with previous studies using similar designs [71, 72]. Data from one subject was lost due to computer error. One additional subject was excluded due to excessive movement during the scan. Our final sample consisted of 28 subjects (17 female, age = 25.4 ± 7.3 years).

Stimuli, apparatus and procedure

This article constitutes an independent analysis of data from a previously published study [3], with detailed methods reported here for completeness. Participants completed a computer-based task while in an fMRI scanner. The goal of the task was to

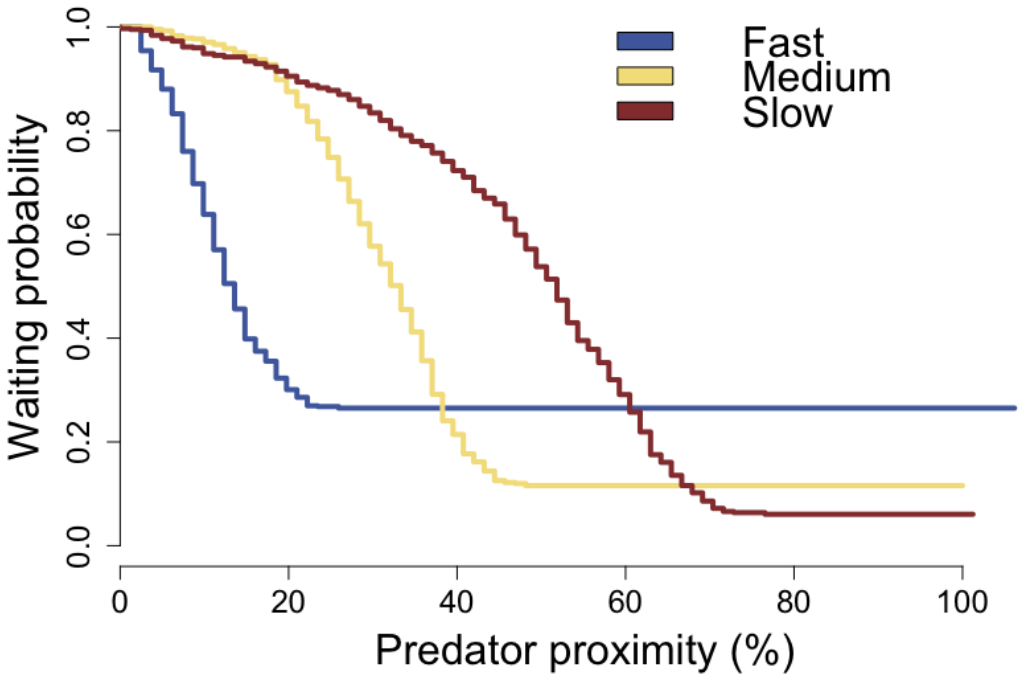

Figure 3.2: Kaplan-Meier survival curves for each predator type, as a function of predator proximity. Curves reflect pooled data from all subjects.

earn as much money as possible while avoiding being caught by a virtual predator. Prior to the beginning of each trial, participants were presented with a 2 second cue indicating one of three different predator types that would be present in the upcoming trial. The participants were then shown a two-dimensional runway (90 units distance), with an triangle icon representing the position of the participant toward the right of the runway (at 80 units distance), and a circle icon representing the position of a predator at the left side of the runway (at 1 unit distance). This predator had two distinct modes of movement. In "approach" mode, the predator would proceed rightward along the runway at 4 units per second. At a randomly chosen distance (i.e. the attack distance) the predator would switch to "chase" mode, at which point it would advance at 10 units per second. These attack distances were randomly sampled from one of three Gaussian distributions, with means of 25, 40, 50 (standard deviations: 20, 20, 20; for the "slow", "medium", and "fast" predator types, respectively [1]). Participants would passively gain money at a rate

[1]Note that these predator types differed only in their mean attack distance, and not actually the speed of their attack.

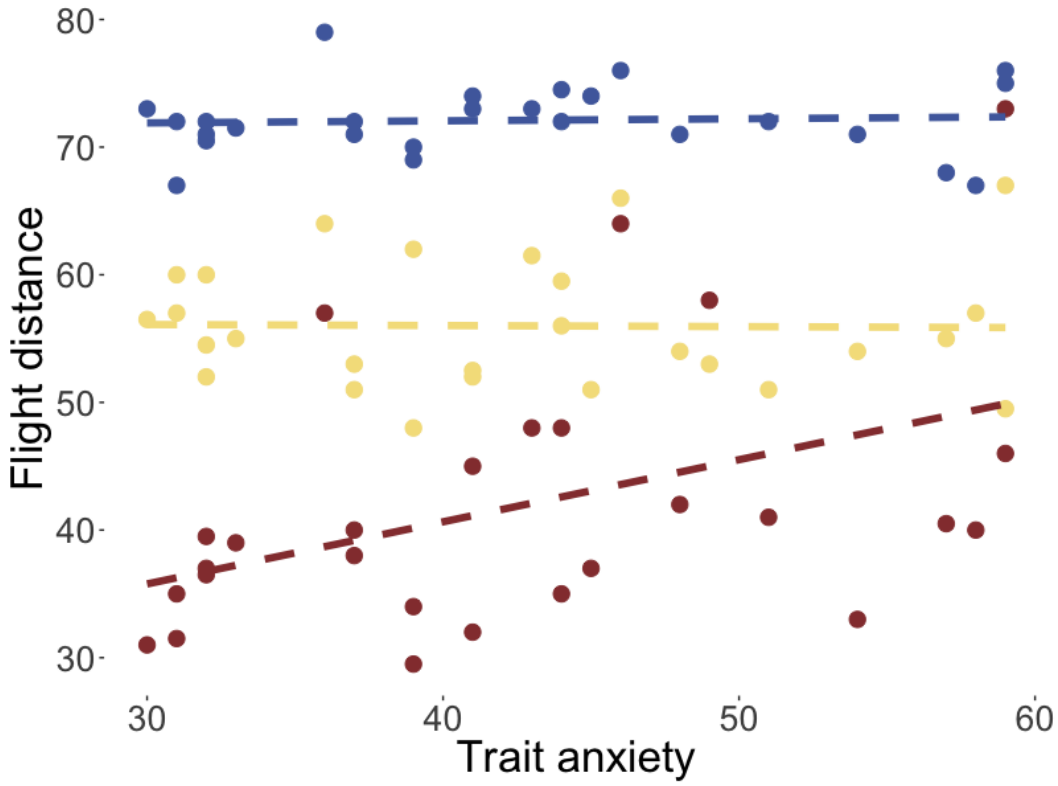

Figure 3.3: Flight initiation distance for each predator type, as a function of STAI-Y scores. Each dot corresponds to a single subject's median FID in one condition. Dashed lines show the linear fit to the data.

of 2 cents per second while they remained on the runway, and at any time could press a button to begin an escape toward the right side of the runway at 2 units per second. Notably, if participants did not respond prior to the predator reaching its attack distance, it was not possible for them to escape. This prevented participants from merely relying on their reaction time by responding after the predator switched modes. If participants escaped successfully, they would earn the monetary reward accumulated during that trial. If they failed to escape successfully (i.e. were caught by the predator) participants were given a mildly aversive electric shock (the shock magnitude was calibrated to each individual prior to testing), and the monetary reward earned in that trial would be forfeited. Thus, to perform this task optimally, participants had to learn the distributions of attack distances for each of the predator types, and respond as late as possible, provided the distance between them and the predator (i.e. the FID) was sufficient for a successful escape. Prior to the beginning of this main task, participants completed a brief, 8-trial practice session to familiarize themselves with the paradigm. (The attack distances of the predators were

drawn from different distributions to those used in the main task.) Participants then completed 96 trials of the main task. After 48 trials, the predator-color cue was re-assigned in order to maintain the attentional demands of the task. Participants also performed a matching control condition for each predator type, without the risk of shock or the incentive of monetary reward, but otherwise identical to the main task. After completion of the computer task, subjects were asked to complete a series of personality questionnaires that included the trait subscale of the Spielberger State-Trait Anxiety Inventory, Form Y and the behavioral inhibition/activation scale (BIS/BAS) [19](see Supplementary materials for an analysis of BIS scores). The computer task was programmed in Cogent with Matlab. Data collection and analysis were not performed blind to the conditions of the experiments.

All fMRI data were acquired using a GE Discovery MR750 3.0 T scanner with 32-channel headcoil. The imaging session consisted of two function scans, each twenty minutes, as well as a high-resolution anatomical T1-weighted image (1 mm isotropic resolution) collected at the beginning of each scan session. For functional imaging, interleaved T2*-weighted gradient-echo echo planar imaging (EPI) sequences were used to produce 45 3-mm-thick oblique axial slices (TR = 2 s., TE = 25 ms, flip angle = 77, FOV = 192 x 192 mm, matrix = 64 x 64). Each functional run began with five volumes (1000 msecs) before the first stimulus onset. These volumes were discarded before entering analysis to allow for magnetic field equilibration. Participants viewed the screen via a mirror mounted on the head coil, and a pillow and foam cushions were placed inside the coil to minimize head movement. Electric stimulation was delivered using a BIOPAC STM100C.

Data analysis

All statistical analyses for the behavioral data were carried out in R [91], using the packages 'ezANOVA' [50], 'coxme' [102], and 'lme4' [9]. Prior to analyses, data were tested for normality and equal variances using Shapiro-Wilk and Mauchly's sphericity test, respectively. Where appropriate, log transformations of data were performed to account for non-normality, and Greenhouse–Geisser corrections were performed to account for violations of sphericity, with the correction factor values (ϵ) and original degrees of freedom reported. Partial eta-squared effect sizes are reported only for significant analyses. Where appropriate, we corrected for multiple comparisons using Holm-Bonferroni. All tests were two-tailed unless otherwise specified. We used an alpha level of .05 for all statistical tests[2].

[2]Post-hoc power analyses are available from the authors by request.

Analysis of fMRI data was carried out using scripted batches in SPM8 software (Welcome Trust Centre for Neuroimaging, London, UK) implemented in Matlab 7 (The MathWorks Inc., Natick MA). Structural images were subjected to the unified segmentation algorithm implemented in SPM8, yielding discrete cosine transform spatial warping coefficients used to normalize each individual's data into MNI space. Functional data were first corrected for slice timing difference, and subsequently realigned to account for head movements. Normalized data were finally smoothed with a 6-mm FWHM Gaussian kernel.

Preprocessed images were subjected to a two-level general linear model using SPM8. The first level contained the following regressors of interest, each convolved with the canonical two-gamma hemodynamic response function: a 2 s box-car function for the onset of the trial (during predator type cue presentation), a 4-8 s (duration jittered) box-car function from the onset to 2 s prior to participants' flight decisions, a 2 s boxcar function for the time prior to participants' flight decisions, and a 4-8 s (duration jittered) box-car function for the remainder of the trial. Mean-centered STAI-Y scores ratings were included as orthogonal regressors. In addition, nuisance regressors consisted of motion parameters determined during preprocessing, their first temporal derivative and discrete cosine transform-based temporal low frequency drift regressors with a cutoff of 192 s. Beta maps were used to create linear contrast maps, which were then subjected to second-level, random-effects one-sample t-tests. In addition, a flexible factorial model was used to examine the main effects of predator type. The resulting statistical maps were thresholded at $p < 0.05$, and we corrected for multiple comparisons using false discovery rate correction (FDR whole brain corrected) [32].

After whole-brain analyses, a hypothesis-driven region of interest (ROI) analysis was performed. These regions were chosen based on results from a previous study using the same behavioral task (see [3]).

The functional connectivity analysis was performed for the response phase (escape decision) using a generalized psychophysiological interactions (gPPI) approach [59]. vHPC was chosen as the seed region for subsequent PPI analysis due to its functional role in fear, stress and emotion [60, 72] and its empirically demonstrated involvement in our previous study [3]. (See Supplementary Materials for a similar analysis that includes the dorsal hippocampus.) In the PPI model, regressors of interest included the 3 predator conditions (slow/medium/fast), their corresponding control conditions, and the PPI terms for the above mentioned 6 conditions. Using the gPPI

toolbox [59], a first level connectivity analysis was carried out based on the PPI term of the direct comparison between the two predator conditions (slow versus fast attacking predator). (A similar connectivity analysis based on the PPI term of the comparison between the slow predator and control condition can be found in the Supplementary Materials.) As a second level analysis, STAI-Y scores were then introduced as a co-variate to examine how trait anxiety alters the strength of the PPI with respect to the seed regions.

Data and code availability

Behavioral data and accompanying code for all behavioral analyses and figures can be found on the Open Science Framework (https://osf.io/c4qbr/). FMRI data and analysis code are available from the corresponding author on reasonable request.

3.3 Results

Trait anxiety and flight initiation distance

To test the hypothesis that trait anxiety would affect escape decisions, we estimated a mixed effects linear regression model, with subjects' median FIDs as the dependent variable, and predator type and STAI-Y scores as the independent variables (Table 3.1). Relative to the fast predator type, we observed the expected effects of the medium ($\beta = -17.88, SE = 2.23, p < 0.001$) and slow ($\beta = -52.22, SE = 2.26, p < 0.001$) predator types. Importantly, we observed a significant interaction effect between the slow predator type and STAI-Y scores ($\beta = 0.57, SE = 0.05, p < 0.001$), suggesting that trait anxiety and FID were related, but only for the slow predator condition (see Figure 3.3)

Survival analysis

Note that because participants were given electrical stimulation when they were caught by the virtual predator, in order to obviate interference it was necessary to exclude these trials from the imaging analysis reported below. For consistency, the behavioral analysis above also excluded unsuccessful escape trials. However, unsuccessful escape trials still contain information about subjects' tolerance to predator distance. To ensure that the analyses above were not biased by this possibility, we adopted a technique from survival analysis, which allowed us to take into account the unsuccessful trials as censored data. To appropriately prepare the data for this analysis (which is more commonly used to model time-based responses rather than distance-based responses) we transformed the dependent variable of FID by sub-

Table 3.1: Linear regression of predator type and STAI-Y scores on flight initiation distance.

	Dependent variable:
	Flight initiation distance
Medium predator	-17.879^{***}
	(−22.252, −13.507)
Slow predator	-52.219^{***}
	(−56.646, −47.792)
STAI-Y	−0.010
	(−0.212, 0.192)
Medium predator:STAI-Y	0.072
	(−0.029, 0.173)
Slow predator:STAI-Y	0.567^{***}
	(0.465, 0.668)
Constant	72.239^{***}
	(63.395, 81.083)
Observations	1,691
Log Likelihood	−5,892.115
Akaike Inf. Crit.	11,800.230
Bayesian Inf. Crit.	11,843.690
Note:	*p<0.1; **p<0.05; ***p<0.01

tracting FID from the maximum FID, then normalizing this by the maximum FID. This new dependent variable can be thought of as predator proximity, expressed as a percentage. The Kaplan-Meier estimated survival curves (i.e. probability of waiting as a function of predator proximity) for each predator are shown in Figure 3.2.

To control for the potential effect of data censoring, we repeated the analysis of behavioral data using a mixed effects Cox regression model on the probability of flight responses over time, which took into account predator type and participant heterogeneity. This model again revealed the expected effects of the medium ($\beta = -0.98, SE = 0.29, z = -3.34, p < 0.001$) and slow ($\beta = -3.09, SE = 0.3, z = -10.43, p < 0.001$) predator types. Importantly, it also again revealed a significant

interaction effect between the slow predator type and STAI-Y scores ($\beta = 0.05, SE = 0.01, z = 6.74, p < 0.001$). This effect had a hazard ratio of 1.05, equivalent to a 5% increase in chance of fleeing per unit increase of STAI-Y.

Trait anxiety and performance

The results above provide clear evidence that trait anxiety influences subjects' propensity to escape earlier when given enough time to prepare an escape. However, it is unclear whether this should negatively affect their economic performance in the task. To test this, we performed a two-way repeated measures ANOVA, with predator type and STAI-Y scores as independent variables, and subjects' cumulative total earnings as the dependent variable. Given that subjects could earn more money in the slow predator condition, we first standardized reward scores for each predator type. There was no significant effect of predator type on standardized earnings ($F(2, 52) = 0.34, p = 0.667, \epsilon = 0.81$), but we observed a significant main effect of STAI-Y scores on total earnings ($F(1, 26) = 4.32, p = 0.048, \eta_p^2 = .09$), suggesting that subjects with higher STAI-Y scores had poorer economic performance in the task, across all predator types. There was no interaction effect of STAI-Y scores and predator type ($F(2, 52) = 0.36, p = 0.656, \epsilon = 0.81$).

Although economic gain is an index of performance in this task, it could be argued that the more ecologically important performance measure is escape success. Notably, subjects' economic performance and proportion of escape trials were not significantly correlated across all predator types ($r(26) = .09, p = .643$)[3]. To test whether trait anxiety was related to how frequently subjects successfully escaped the predators, we again performed a two-way repeated measures ANOVA, with predator type and STAI-Y scores as independent variables, and the proportion of successful escape trials as the dependent variable. While there were no main effects of STAI-Y scores ($F(1, 26) = 0.23, p = 0.633$) or predator type ($F(2, 52) = 1.89, p = 0.175, \epsilon = 0.53$), the ANOVA revealed a significant interaction effect between STAI-Y scores and predator type ($F(2, 52) = 4.46, p = 0.031, \epsilon = 0.68, \eta_p^2 = .15$). Simple effects analyses (one-way repeated measures ANOVAs within each predator type) revealed a significant effect only for the slow predator type ($F(1, 26) = 5.49, p = 0.027, \eta_p^2 = .17$), but not for the fast ($F(1, 26) = 2.12, p = 0.158$) or medium predator ($F(1, 26) = 0.39, p = 0.536$). This suggested that, similar to the analysis of FID above, STAI-Y score was positively

[3]Economic performance and proportion of escape trials were not significantly correlated within the slow predator condition ($r(26) = -.31, p = .108$)

related to escape success only in the slow predator condition. Overall, these results show that subjects with higher trait anxiety tended to more successfully escape predators, but that this also negatively impacted how much money they earned in the task (a summary of performance measures can be found in the Supplementary materials, in Table 3.5).

Imaging data

We next tested our hypothesis that only during slow attack would we see a positive correlation between trait anxiety and activity in the "cognitive fear" circuitry. For this analysis we excluded unsuccessful escape trials due to the interference of the electric stimulation on BOLD response (mean trials excluded were 6.88, 3.71, and 3.37, per subject, out of 23, 24 and 25, for the fast, medium and slow predator types, respectively). We focused on the 2 seconds prior to participants' flight initiation responses, which allowed us to examine the neural activity in anticipation of the escape response (detailed methodology of the base fMRI analysis can be found in [3]). We first contrasted the slow attacking predator condition with the fast attacking predator condition. We then used participants' STAI-Y scores as 2nd level regressors for this contrast, such that any significant increase in activity would indicate positive modulation by trait anxiety for the slow predator condition.

After thresholding and correction, we observed significant BOLD responses in regions including the amygdala, hippocampus, vmPFC and midcingulate cortex (Figure 3.4A, Table 3.2). This was consistent with our hypothesis, and supported the behavioral findings whereby STAI-Y score exclusively influences escape decisions when the threat is distant (in the case of the slow attacking predator), but not when the threat is imminent (in the case of the fast attacking predators). A visualization of BOLD response as a function of trait anxiety for select regions is shown in Figure 3.6.

To assess the the interaction of brain regions involved in escape decisions, we performed a generalized psychophysiological interaction (gPPI) analysis [59]. Given the theoretical and empirically demonstrated involvement of the vHPC in cognitive fear and anxiety [58, 72], and because of its exhaustive bidirectional anatomical connections with the amygdala and its nuclei, as well as its functional role in fear, stress and emotion [28, 44, 75], we chose vHPC as an independent seed region. A corresponding structural ROI was obtained using the WFU Pickatalas. This first level gPPI analysis on the slow versus fast predator contrast is reported in [3], and

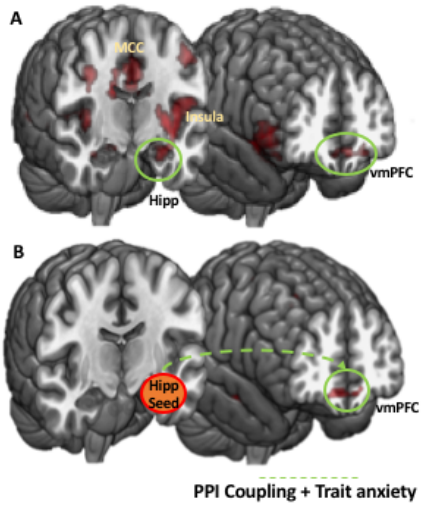

Figure 3.4: B

rain imaging results concerning STAI scores. (A) Neural activity associated with STAI-Y scores for the slow versus fast predator contrast. (B) PPI coupled brain areas modulated by STAI-Y score. vmPFC, ventromedial prefrontal cortex; Hipp, hippocampus; MCC, mid-cingulate cortex.

Table 3.2: Activation table for 2nd level STAI-Y score correlation for the slow versus fast predator contrast.

Brain Region	Left/Right	Cluster Size	t-score	MNI coordinates		
				x	*y*	*z*
Hippocampus	L	60	5.32	-15	-27	-6
Postcentral Gyrus	L	209	4.91	-45	-18	54
Medial Prefrontal Cortex	L	63	4.70	-3	51	-14
Insula	L	94	4.53	-40	8	-3
Insula	R	107	4.74	36	6	-6
Amygdala	R	15	4.93	22	0	-20

Note: p<0.05, FDR corrected

will not be reported here for brevity. We then added STAI-Y score as a regressor in a second level analysis. STAI-Y score significantly modulated the functional coupling between vHPC seed, bilateral mPFC, right IFG, the left insula (Figure 3.4, Table 3.6). Overall, this suggests that these macrocircuits are those that facilitated the impact of STAI-Y score on escape decisions in the slow predator condition.

3.4 Discussion

Our results provide evidence that trait anxiety can influence escape decisions, but only under conditions of relatively prolonged threat, compared to more imminent threats [67, 70]. This disassociation implies that trait anxiety selectively affects decisions of different ethnological classes, distinguished by the amount of time afforded for reflection and cognitive strategizing. The notion of a dichotomous mapping between temporally proximal threats and fear, and temporally distal threats and anxiety is not new. For example, rodents' defensive behavior differs when threat is distal versus when it is immediate [26], and anxiolytic drugs appear to only affect the former [14]. Likewise, previous models of threat evaluation have suggested that both anxious and non-anxious individuals will respond similarly to proximal threats, but individuals with high anxiety will exhibit differential behavior to more distal threats [56]. However, this is the first empirical study to show that trait anxiety selectively impacts escape decisions in humans under this specific class of threat.

The interpretation that trait anxiety affects only "cognitive" fear behavior was supported with our neuroimaging results. These results showed that brain areas previously indicated to be involved with behavioral flexibility and information processing aspects of fear responses (including hippocampus, amygdala, mPFC and insula [3, 61, 72]) covaried with trait anxiety. However, areas associated with "reactive" fear - the PAG, superior colliculus, mid-cingulate cortex and central nucleus of the amygdala [3, 33, 72] - were not significantly affected by variability in anxiety. Notably, these findings strongly support theories based on defensive distance [13], whereby defensive responses to immediate threats and dangers map onto low-level brain areas such as the PAG, whereas responses to physically or psychologically distal or anticipated threats map to higher-level areas such as the PFC [33, 60]. Our findings extend these theories by providing a clear disassociation of the effects of trait anxiety on one circuit over the other, with accompanying behavioral effects, in an ecologically relevant paradigm.

These seed-based functional coupling results are consistent with previous non-

human animal studies showing functional interactions between the ventral and dorsal hippocampus and vmPFC in anxiety provoking environments [1, 2]. For example, local field potential recordings in rodents have shown that there is synchrony in theta oscillatory activity between vHPC and mPFC, and that this synchrony is increased in anxiogenic environments [1]. In addition, single unit recordings have shown that cells in mPFC have stronger anxiety-related firing patterns when phased-locked with local field potentials in the vHPC [2]. Using magnetoencephalography, others have corroborated these non-human animal findings in humans [45]. Our results parallel both the human and non-human animal evidence for functional coupling between the vHPC and mPFC, and further consolidate the characterization of this interaction with a different brain imaging method.

The specific nature of the coupling between the vHPC and mPFC has garnered some previous discussion. For example, because vHPC-mPFC connections are unidirectional [82], it has been suggested that the vHPC primes mPFC to represent anxiety-related features of the environment, possibly using memories of threats to estimate threat probability [45]. MPFC has efferent projections to amygdala and PAG, and these connections have been suggested to be the downstream areas responsible for the initiation of defensive behavioral responses [33, 108], and the inhibition of exploratory behaviors [2]. To complement this, vHPC also has direct projections to BLA, BNST, and the lateral hypothalamic area (LHA), which can also facilitate anxiety responses [44].

In light of the results from our study, it is possible that vHPC may encode the previously learned threat context (i.e. the predator condition), and relay this information to the mPFC where it influences strategic decision making. Our results suggest that the observed increase in connectivity between vHPC and mPFC in trait anxious individuals may reflect a priming mechanism which lowered the threshold for escape responses, resulting in earlier escape decisions [17]. However, for the fast predator condition, this slow, deliberative priming is not sufficient, and thus the initiation of behavioral responses appears to bypass this connection. One compelling question is whether trait anxiety merely interacts with this vHPC-mPFC mechanism, or whether it can be fully identified with information processing between these subregions. While we speculatively provide this neural mechanism for trait anxiety - which is also supported by the non-human animal literature - we emphasize that this requires causal corroboration, perhaps in the form of pharmaceutical manipulations in humans. Another further piece of evidence that would provide compelling

support for such a mechanism would be trial-by-trial prediction of flight initiation distance using brain activity in vHPC-mPFC (an approach that our design lacked appropriate power for).

Notably, we did not observe modulation of BLA, BNST or LHA, by trait anxiety. One likely possibility is that these areas are involved in longer-term anxiety responses, requiring the recruitment of corticotropin-releasing hormone [26], and that our slow predator condition was not adequately protracted to cause these responses. Given that BLA and amygdala have strong inputs to vHPC [30, 31], another possibility is that these areas are more commonly recruited during fear learning (which we did not examine), and imbue the encoding of environmental stimuli with emotional salience (e.g. [53]). Indeed, most empirical evidence of the increased involvement of the amygdala in trait anxious individuals has come from learning paradigms and studies of fear conditioning (e.g. [40]). Thus, trait anxiety is likely to affect both the encoding of threats, as well as their retrieval from memory, potentially via different neural substrates. This latter point may be of critical import for many clinical anxiety disorders (such as post traumatic stress disorder), where threats have already been learned. One further possibility, as suggested previously [45], is that vHPC is specifically involved in threat memory retrieval only when there is approach-avoidance conflict [8, 41, 78], as in the case with the trade-off between reward and threat of shock in our task.

Previous research has also suggested the possibility that mPFC representation of the environment depends of the strength of vHPC input: moderate input appropriately signals the aversiveness of specific features, but strong input decreases discriminative capability, leading to generalized anxiety responses [2]. In our study we were not able to evaluate individuals' abilities to discriminate between different levels of threat, but this would be a promising avenue for future research. In particular, this might suggest that populations with clinical anxiety disorders may exhibit increased coupling between vHPC and mPFC across threat levels, and consequently faster escape decisions for all predator conditions.

The impact of trait anxiety on escape decisions could influence survival outcomes in at least two important ways [55, 87]. Firstly, if individuals with high trait anxiety escape predators earlier, they expedite other behaviors, like foraging, and thus may accrue less primary rewards. Our results support this idea by showing that those with higher trait anxiety earned less total reward in our task. On the other hand, it could be argued that a more survival-relevant performance metric is

successful escape - additional reward is irrelevant if caught by a predator. Our results also showed that individuals with higher trait anxiety made a higher proportion of successful escape decisions. However, unlike the reward, which was affected across all predator conditions, individuals with higher trait anxiety only made a higher proportion of successful escape decisions within the slow predator condition, in line with the idea that trait anxiety only affects flight decisions under these contexts. One possible explanation for this difference may have been that trait anxiety also affected escape responses in the medium and fast predator conditions, but to a lesser, non-significant degree. This is especially possible considering that there is some individual variability both in trait anxiety and in performance in general, and thus our specification of "cognitive" and "reactive" fear classes will not have perfectly divided performance in these individuals. A series of experiments spanning a large range of predator conditions and reward contingencies may be able to address this issue with more clarity, and perhaps reveal population level differences in how trait anxiety influences performance. Ultimately, both the accrual of reward and successful escape are important factors for survival, and differences in trait anxiety appear to arbitrate between these, depending on threat context.

Coexisting with a disassociation of anxiety and fear based on defensive distance is a disassociation based on defensive direction [60]. The "direction" of this construct refers to approach / avoidance, and theoretical work proposes that fear drives avoidance of danger, while anxiety drives approach toward danger [60]. In our experimental design, an approach avoidance conflict existed between reward and the threat of shock. Because the slow predator condition allowed individuals to earn greater rewards, this condition may have elicited greater relative anxiety. Under the defensive direction framework, we may have expected participants with higher trait anxiety to endure longer in this condition. However, we found that individuals scoring higher in trait anxiety escaped earlier, which speaks against defensive direction as a potential explanation for our behavioral results. It would be of interest however, for future experiments to more closely examine how defensive direction and trait anxiety relate to each other.

Previous studies have also found evidence that anxiety can affect decision making. For example, individuals with higher dispositional anxiety are more likely to be more risk-averse in tasks such as the balloon analogue risk task [54]. Our study makes an important contribution to this literature by situating individuals in an ecological setting, where the effect of anxiety can be seen as a plausible adaptive

role, rather than a straightforward deficit in decision making. As such, our findings support evolutionary accounts of anxiety disorders [63, 76]. While it is important to note that our current findings do not generalize to populations with clinical anxiety disorders, such as post-traumatic stress disorder, our hope is that future research will capitalize on the distinctions between threat contexts in order to better diagnose and treat these disorders. One potential avenue, for example, would be to tailor treatments and interventions based on individual differences in threat categorization.

Overall, this study provides strong empirical support for the notion that trait anxiety affects behavior only when there is sufficient time to appropriately cognize a threat, and not when threats require an immediate reactive response. These behavioral results were borne out in an ecologically relevant paradigm, and were complemented with neural data which suggest that previously learned threat contexts more heavily influence strategic decision making in trait anxious individuals. The present study provides a complement to previous work describing the contexts under which "reactive" fear defensive responses manifest [3, 66, 72], and the behavioral and neural signatures of these responses, and in combination, point to the importance of examining different ecological classes of threat in future work.

3.5 Supplementary materials

Variability in flight initiation distance

It is important to note that participants had a larger time window in which to respond in the slow predator condition, thus, while the variances of the empirical attack times were not significantly different (all subjects experienced the same empirical attack times; fast vs medium: $F(24, 23) = 1.36, p = 0.464$, medium vs slow: $F(23, 22) = 0.57, p = 0.186$, fast vs slow: $F(24, 22) = 0.78, p = 0.537$), the variance in escape distances was not equal across predator types, neither across subjects (variances across median FIDs, fast vs medium: $F(26, 27) = 0.25, p < 0.001$, medium vs slow: $F(27, 27) = 0.27, p = 0.001$, fast vs slow: $F(26, 27) = 0.07, p < 0.001$), nor within subjects (t-test of per subject FID variance, fast vs medium: $t(32.94) = -6.12, p < 0.001$, medium vs slow: $t(32.22) = -5.67, p < 0.001$, fast vs slow: $t(27.59) = -7.91, p < 0.001$). In particular, the slow predator condition had significantly larger variance in responses (mean per subject variances were 90.44, 32.51, and 12.82, for slow, medium, and fast predators, respectively).

Importantly, these differences in response variability were a direct consequence of the experimental manipulation, that is, the manipulation designed to elicit "reac-

tive" fear allowed a relatively shorter response window, and thus entailed increased urgency. For this reason, we do not consider the differences in response variability to be a confounding factor, but rather a necessary feature of the manipulation. However, to provide some evidence that the wider time window alone was not responsible for the relationship between FID and STAI-Y scores, we pooled the responses across the fast and medium predator types. The variance of median responses in this pooled data was not significantly different from that of the slow predator condition ($F(54, 27) = 0.73, p = 0.329$). We performed a similar linear regression analysis with this pooled data, which showed that the interaction effect between the slow predator condition and STAI-Y scores remained significant ($\beta = 0.56, SE = 0.06, p < 0.001$). Overall, this suggested that this relationship between STAI-Y scores and FID was not simply due to subjects having a larger variance of responses in the slow predator condition.

Behavioral inhibition and flight initiation distance

Another trait factor that may have played a role in escape decisions is sensitivity to punishment, or behavioral inhibition [60]. Originally, this was proposed as a neurobiological substrate for anxiety [34], and is still routinely believed to play an important role in anticipating and assessing threats. We wished to investigate whether behavioral inhibition could also explain some of the variance in flight distance, above or beyond that of trait anxiety as measured by the STAI-Y. Firstly, we tested whether STAI-Y scores and BIS scores were related within our sample. A Pearson correlation showed the relationship between STAI-Y scores and BIS scores was not significant across participants ($r(26) = .09, p = .660$). We then ran a mixed effects regression analysis similar to that used in the main text, with FID as the dependent variable, and predator type, STAI-Y scores, and BIS score as independent variables (Table 3.3). The results of this analysis recapitulated the effects observed in previous model, including the significant interaction between STAI-Y scores and the slow predator condition ($\beta = -7.4, SE = 0.2, p < 0.001$). The model additionally revealed a significant interaction effect of BIS score and predator type for the slow predator condition ($\beta = -2.98, SE = 0.47, p < 0.001$). It also revealed a significant three-way interaction between BIS score and STAI-Y scores in both the medium ($\beta = 0.02, SE = 0.01, p = 0.031$) and slow predator conditions ($\beta = 0.08, SE = 0.01, p < 0.001$). As can be seen in the median split visualization plotted in Figure 3.5, the relationship between STAI-Y scores and FID appears to be driven predominantly by those with higher BIS scores. However, it

is critical to note that this result should be interpreted with caution, as three-way interaction effects require substantially more experimental power to appropriately detect, and the sample size of this study was not chosen with this in mind. In general, this analysis suggests that the STAI-Y and BIS scores are separable, and that BIS similarly, but independently, influences FID.

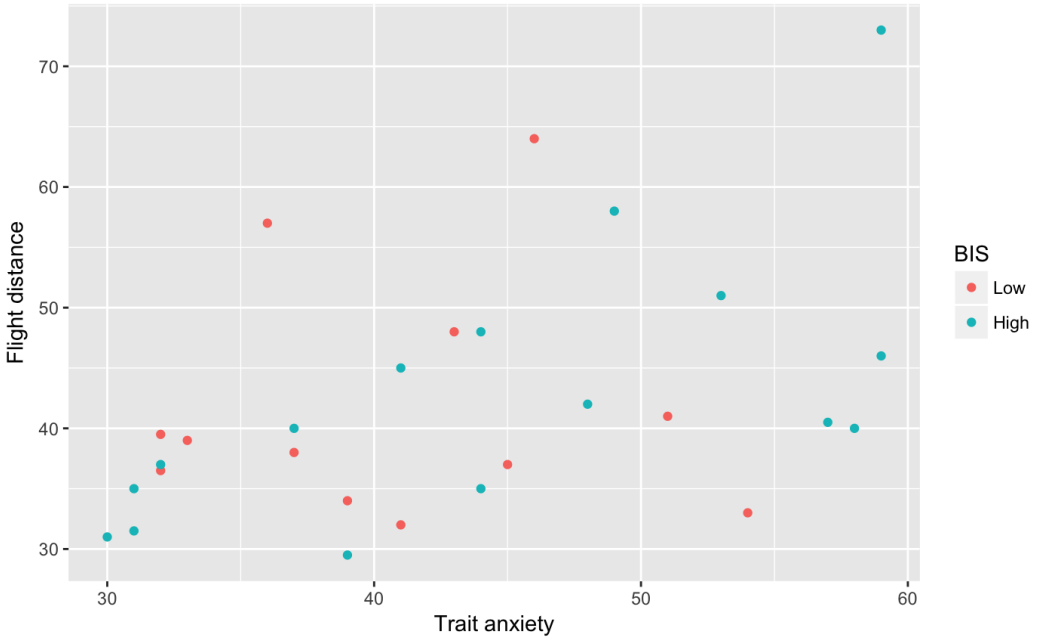

Figure 3.5: Visualization of the interaction of STAI-Y and BIS on flight initiation distance within the slow predator condition

We have also performed an exploratory fMRI analysis similar to the analysis of trait anxiety within the slow predator condition (see main text), but instead using BIS score. Here we find significant activity in the thalamus and right caudate (Table 3.4).

Performance data

Here we report the summary statistics for participants performance in the task, as a function of predator condition (Table 3.5).

Activation table for 2nd level STAI-Y score correlation for PPI (vHPC seed)

PPI with hippocampus seed

In the main text we reported the results of a PPI analysis showing modulation of brain areas by STAI-Y score from a dorsal hippocampus seed. Given that literature has also pointed to interactions between dorsal hippocampus and mPFC[114], here

Table 3.3: Linear regression of predator type, STAI-Y and BIS on flight initiation distance

	Dependent variable:
	Flight initiation distance
Medium predator	−7.048
	(−23.123, 9.027)
Slow predator	−0.273
	(−16.512, 15.966)
STAI-Y	0.014
	(−0.786, 0.813)
BIS	−0.032
	(−1.941, 1.876)
Medium predator:STAI-Y	−0.346*
	(−0.728, 0.035)
Slow predator:STAI-Y	−0.741***
	(−1.126, −0.357)
Medium predator:BIS	−0.562
	(−1.467, 0.344)
Slow predator:BIS	−2.978***
	(−3.892, −2.064)
STAI-Y:BIS	−0.001
	(−0.046, 0.043)
Medium predator:STAI-Y:BIS	0.024**
	(0.002, 0.045)
Slow predator:STAI-Y:BIS	0.075***
	(0.054, 0.097)
Constant	72.634***
	(38.897, 106.371)
Observations	1,691
Log Likelihood	−5,871.537
Akaike Inf. Crit.	11,771.080
Bayesian Inf. Crit.	11,847.140
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3.4: Activation table for 2nd level BIS score correlation for the slow versus fast predator contrast

Brain Region	Left/Right	Cluster Size	t-score	MNI coordinates		
				x	*y*	*z*
Thalamus	R	41	4.50	12	-12	0
Caudate	R	31	4.94	3	20	0

Note: p<0.05, FDR corrected

Table 3.5: Summary of performance measures

Predator Type	N	Mean earnings (SD)	Mean escape proportion (SD)
Slow	28	889.89 (174.6)	0.9 (0.09)
Medium	28	563.52 (75.54)	0.88 (0.11)
Fast	28	267.56 (83.97)	0.74 (0.2)

Table 3.6: Activation table for 2nd level STAI-Y score correlation for PPI (vHPC seed)

Brain Region	Left/Right	Cluster Size	t-score	MNI coordinates		
				x	*y*	*z*
Insula	L	49	5.13	-33	9	0
Medial Prefrontal Cortex	R	168	5.00	15	60	-6
Medial Prefrontal Cortex	L	124	5.18	-18	51	-6
Inferior Frontal Gyrus	R	38	5.11	42	15	9

Note: p<0.05, FDR corrected

we report the activation table for a similar analysis, using the entire hippocampus (Table 3.7).

Slow predator versus control contrast modulated by anxiety

Here we report an analysis for the effect of STAI-Y scores within the slow predator condition similar to that presented in the main text, but using a contrast based on the control condition (Table 3.8. Note that an identical analysis using the fast predator condition versus the control condition does not reveal any significant activation in any areas.

Visualization of BOLD signal change as a function of trait anxiety in four brain regions

Table 3.7: Activation table for 2nd level STAI-Y score correlation for PPI (entire hippocampus)

Brain Region	Left/Right	Cluster Size	t-score	MNI coordinates		
				x	*y*	*z*
Insula	L	77	5.37	-31	13	5
Medial Prefrontal Cortex	R	96	4.62	11	52	-14
Medial Prefrontal Cortex	L	83	4.77	-8	59	-12
Inferior Frontal Gyrus	R	63	4.91	48	15	-9
Parahippocampal Gyrus	R	57	5.56	26	-20	15
Amygdala	L	38	4.79	-24	-2	-15

Note: p<0.05, FDR corrected

Table 3.8: Activation table for 2nd level STAI-Y score correlation for the slow versus control predator contrast.

Brain Region	Left/Right	Cluster Size	t-score	MNI coordinates		
				x	*y*	*z*
Amygdala	L	42	7.70	-9	-24	-9
Hippocampus	L	25	6.26	-27	-39	-6
Medial Prefrontal Cortex	R	80	7.46	18	60	-6
Postcentral Gyrus	L	144	4.38	-57	-21	48
Insula	R	133	5.48	45	-24	24

Note: p<0.05, FDR corrected

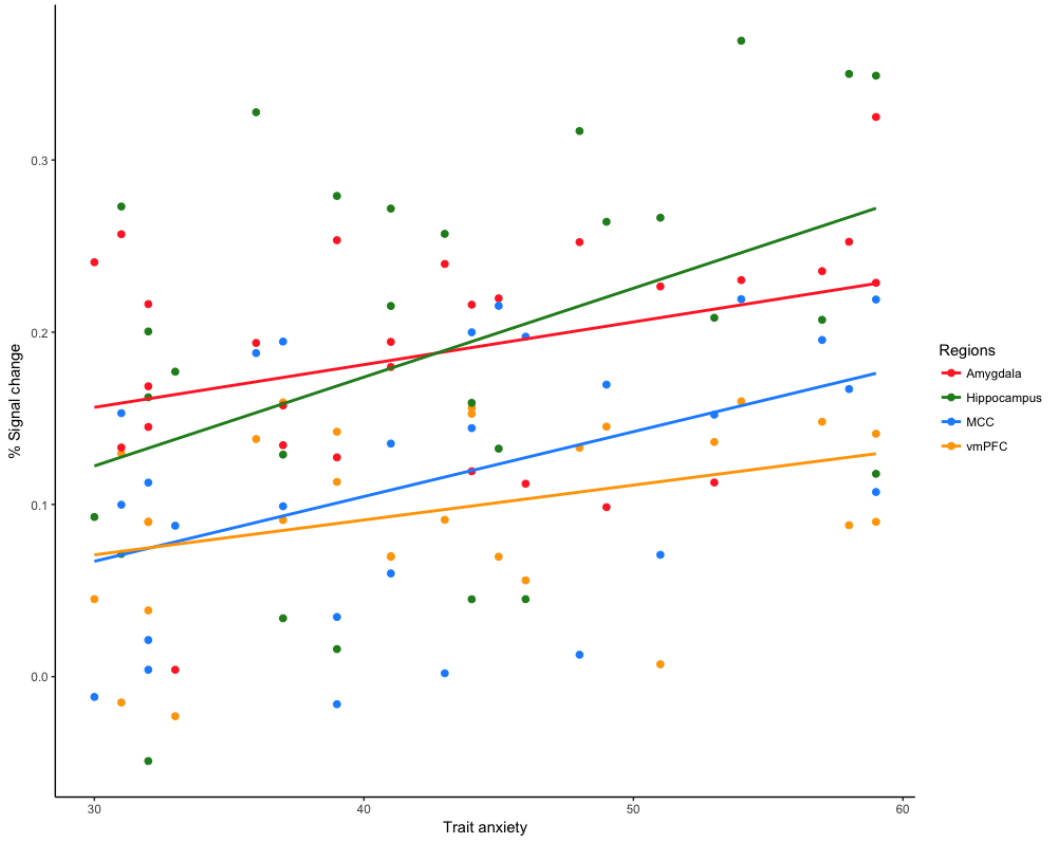

Figure 3.6: Visualization of BOLD signal change as a function of trait anxiety in four brain regions. vmPFC, ventromedial prefrontal cortex; MCC, mid-cingulate cortex.

Chapter 4

SPATIAL MARGIN OF SAFETY IN THE FACE OF VOLATILE ATTACK DISTANCES

4.1 Introduction

Natural observations hint that during threat assessment, prey scrutinize both the danger level of the environment and monitor the distance to locations of safety. One metric that is commonly used to preempt predation is spatial margin of safety (MOS), where prey adopt choices that prevent deadly outcomes from occurring by keeping close proximity to a safety refuge and increasing the success of escape. These pre-emptive safety decisions include prospective planning on how far to move away from a safe refuge or how close one should move to a threat during foraging. Humans may act in similar ways. For example, when human subjects are placed close to a safety exit, measures of fear decreases and when under threat, the sight of safety signals reduces fear and fear reinstatement. Further, safety seeking is also particularly prevalent in a number of affective disorders in humans. These observations point to the important role of safety cues in subjective fear and defense-related behaviors. No studies, however, have investigated the human brain during MOS decisions, nor have they tested how the statistical uncertainty of a threat's attack position influences such decisions.

In the natural world, prey encounter predators that attack with varying degrees of predictability. Predictabilty is often determined by the likelihood of attack and the distance at which the threat will attack. These two threat assessment variables determine defensive behaviors such as pre-emptive avoidance via proximity to a safety refuge (i.e. MOS decisions). Uncertainty, alerts the prey that information is missing concerning the predators attack and results in increased anxiety and risk reduction by pre-emptively avoiding the threat. For example, during central place foarging, some animals leave the safety of their refuge to reach food patches. Once the food patch has been reached, the animal needs to make the decision to either stay and eat, or move back to the safety of the refuge and eat. This latter decision takes more energy, yet results in successful avoidance of predator. By choosing to forage in locations spatially close to a safety refuge, animals can reduce the risk of predation at the cost of lowered foraging quality. Distance to safety, therefore is a

critical when pre-empting and surviving ecological threats.

The pre-emptive avoidance associated with spatial MOS decisions may elicit a set of neural circuits involved in anxiety. This set of circuits, which theoretically overlap, include the ventromedial prefrontal cortex (vmPFC), hippocampus, amydgala and anterior insula [3, 52, 66]. Research has shown that a safety stimulus during an aversive experience results in increased activity in the vmPFC while decreasing threat also results in increased activity in the same region. Research also shows that attention set to safety signals, extinction, and down-regulation of anxiety are associated with vmPFC activity. It is important to note that the vmPFC does not work in isolation and is robustly connected to others regions that play a role in safety and threat-related decisions, including the hippocampus, insula and amygdala, regions involved in prospective nature of safety decisions. These regions have been implicated in decision making under a variety of risks. For example, the insula is involved in harm avoidance [83], uncertainty and represents homeostatic states associated with risky decisions [112]. The hippocampus, on the other hand, reflects decisions that involve, spatial memory and prospective planning of escape and avoidance behaviors [3]. Further, the amygdala plays a role in tuning to different or new dangers [36], and the motivation significance and spatial allocation of stimuli [85]. What remains elusive, is the role of these circuits drive pre-emptive avoidance of danger.

To address this gap in knowledge, we created a task to investigate spatial margin of safety decision and elucidate: (i) How do changes in the threat's attack predictability impact the subjects' MOS decisions? And (ii) what is the neural bases of such decisions. This models the observed ecological phenomenon, where animals often need to venture further away from their safety refuge to acquire adequate supplies of food. How do changes in evolutionary novel leptokurtic attack distance distributions47. Leptokurtic distributions, which are volatile in nature, are generated as the composite of two normal distributions with similar means and contrasting variances. Leptokurtic distributions are probability density curves that have higher peaks at the mean and are fatter tailed where extreme outcomes (outliers) are expected more (Fig. C), we contrasted this with standard Gaussians (Fig. D and E)), which are more computationally familiar, impact subjects' rate of learning? We hypothesize that when subjects are facing virtual predators with higher frequency of outlier attack distributions, this will result in more uncertainty and therefore, increase the decisions to move closer to safety. This effect will be most pronounced in subjects

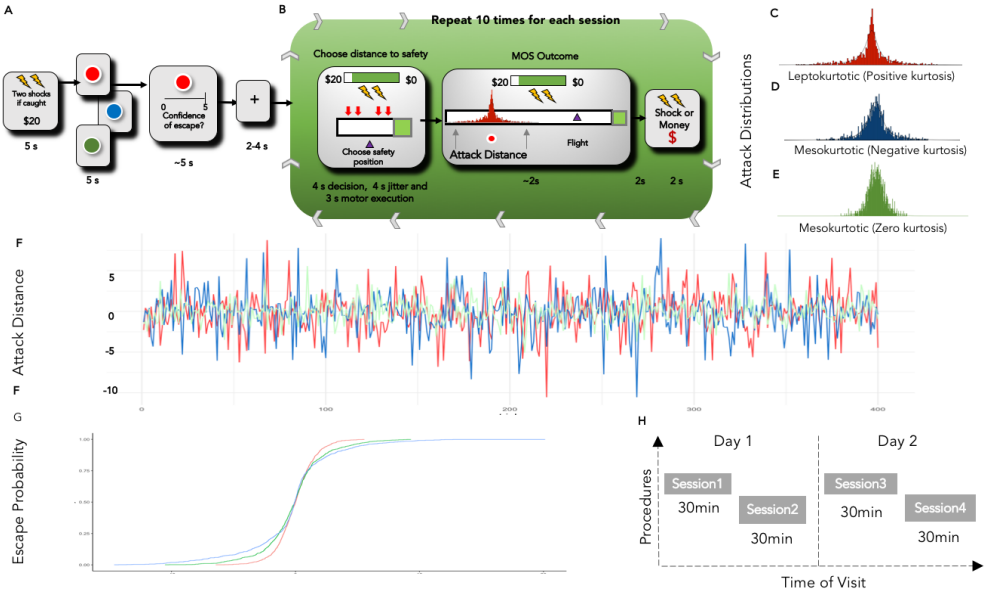

Figure 4.1: Paradigm structure. (A) 2x2 decision variables of high or low reward and punishment, the predator the subject will encounter and confidence of escape rating; This alternates every ten trials; (B) The pre-emptive avoidance decision and the outcome (C) Attack distributions for leptokurtic volatile, (D) gaussian distribution with matched variance and (E) half the variance gaussian; (F) Schematic representation of predators attacks distances through all trials. X axis stands for trials No., and Y axis stands for attack distance. While a "0" on Y axis marks the mean of the distribution, numbers represent how far away the drawn instance is away from the mean. (G) Escape probability. X axis represents possible margin of safety choices, while Y access represents the corresponding probability of escape. (H) Schematic representation of the experimental procedure. Participants undergo 4 session of scans scattered in 2 days.

who score high on the trait anxiety scale. We also predict that the amygdala and insula will be increased in activity when facing the more uncertain attacker.

4.2 Methods

We tested 24 subjects were recruited according to the guidelines of the California Institute of Technology Institutional Review Board after providing informed consent. Data from two subjects were lost due to incomplete scanning sessions. Our final sample consisted of 22 subjects (10 female, age = 24.3 +- 8.1 years).

Stimuli, apparatus and procedure

A complete pipeline of experimental procedures can be found in figure 1. Participants completed a computer-based task while in an fMRI scanner. The task

was set under the scenario where subjects place themselves at a desired location towards a safety exit in face of a potentially dangerous predator. The closer they place themselves to the safety exit, the more likely they'll be able to escape from the predator after the trial starts, but resulting in a smaller amount of potential reward they can earn. The goal of the task was to earn as much money as possible while avoiding being caught by the virtual predator. Prior to the beginning of the trial, the participants were presented with a 2 second cue indicating one of the three different predator types that would be presented in the upcoming trial. These predators differ in the location they speed up. These locations correspond to three distributions – a leptokurtic distribution, a normal distribution with matching variance, and a normal distribution with only half of the variance. The participants were then shown a two-dimensional runway (90 units distance, where a unit is the smallest increment in the program), with an triangle icon representing the position of the participant toward the right of the runway (at 80 units distance), and a circle icon representing the position of a predator at the left side of the runway (at 1 unit distance). This predator had two distinct modes of movement. In "approach" mode, the predator would proceed rightward along the runway at 4 units per second. At a randomly chosen distance (i.e. the attack distance) the predator would switch to "chase" mode, at which point it would advance at 10 units per second. The participants were then told to make a decision of where they want to start by pressing left or right arrows, to move from their randomly assigned initial location to a location they desire. The direction of the chase was counter balanced by adjusting the relative location of the predator, participant and the safety zone so that half of the chase was from the left to the right. After participants responded with their preferred margin of safety choice (MOS choice), they skip the actual animation of the chase (which was shown in full during the practice session), and was shown the final result of the trial: whether they got caught or not, and how much reward they earned.

The experiment starts with the subjects being shown that if captured, they will receive 1 or 2 shocks, and high or low reward if they escape (Fig. 1B). They will then be presented with one of three different colored spheres, each representing different attack distributions of the virtual predators. They will then be asked to rate how confident they are of escape. Next, the subject will be asked to make safety decisions by either staying or switching to a riskier position that is further away from the safety exit or stay or move closer to the safety exit. To motivate risky decisions, the subject will acquire more money if they are more risky (i.e. move further from safety). They will then be asked to move the cursor to the decided safety position. After a jittered

ITI, the subject will observe the outcome. If caught, they will receive a shock(s) and lose their money on this trial. This will repeat for another nine trials, before the subject is introduced to a new set of reward and shock contingences as well as a new virtual predator. The margin of safety task is different from the flight initiation distance as the subject has to make decision concerning their representation of the attack distributions of the virtual predators and make predictions of where it will attack next. The virtual predator attack distribution is either (i) mesokurtic (normal distribution), (ii) leptokurtic (positive kurtosis with fatter tails) or (iii) matched for variance platykurtic distributions (Fig. 7A). Leptokurtic distributions are rare in the natural environment, where distributions are often normally distributed and thus more difficult to learn.

A total number of 400 trials were administrated throughout 4 sessions (2 seesions per day).

After completion of the computer task, subjects were asked to complete a series of personality questionnaires that included the trait subscale of the Spielberger State-Trait Anxiety Inventory, and the behavioral inhibition/activation scale (BIS/BAS) (see Supplementary materials for an analysis of BIS scores). The computer task was programmed in Pygames with Python.

fMRI data acquisition

We will collect the fMRI images using a 3T Prisma scanner in the Caltech Brain Imaging Center (Pasadena, CA) with a 32-channel head receive array. BOLD contrast images will be acquired using a single-shot, multiband T2*-weighted echo planar imaging sequence with the following parameters: TR/TE = 1000/30 ms, Flip Angle = 60°, 72 slices, slice angulation = 20° to transverse, multiband acceleration = 6, no in-plane acceleration, 3/4 partial Fourier acquisition, slice thickness/gap = 2.0/0.0 mm, FOV = 192 mm × 192 mm, matrix = 96 × 96). Anatomical reference imaging will employ 0.9 mm isotropic resolution 3D T1w MEMP-RAGE (TR/TI/TE = 2550/1150/1.3, 3.1, 4.0, 6.9 ms, FOV = 230 m x 230 mm) and 3D T2w SPACE sequences (TR/TE = 3200/564 ms, FOV = 230 mm x 230 mm). Participants viewed the screen via a mirror mounted on the head coil, and a pillow and foam cushions were placed inside the coil to minimize head movement. Electric stimulation was delivered using a BIOPAC STM100C.

Data analysis

All statistical analyses for the behavioral data were carried out in R, using the packages 'ezANOVA' , 'coxme', and 'lme4' . Where appropriate, Greenhouse–Geisser corrections were performed to account for violations of sphericity, and the correction factor values (ϵ) and original degrees of freedom are reported. Partial eta-squared effect sizes are reported only for significant analyses. Where appropriate, we corrected for multiple comparisons using Holm-Bonferroni.

Analysis of fMRI data was carried out using scripted batches in SPM8 software (Welcome Trust Centre for Neuroimaging, London, UK) implemented in Matlab 7 (The MathWorks Inc., Natick MA). Structural images were subjected to the unified segmentation algorithm implemented in SPM8, yielding discrete cosine transform spatial warping coefficients used to normalize each individual's data into MNI space. Functional data were first corrected for slice timing difference, and subsequently realigned to account for head movements. Normalized data were finally smoothed with a 6-mm FWHM Gaussian kernel.

A multivariate pattern analysis was performed using PyMVPA (Hanke et al., 2009). We extracted the beta values associated with experimental conditions of all the voxels in each ROI, removing the mean intensity for each multi-voxel activity pattern. For each participant, the brain response pattern analyses of classification training and testing with linear support vector machines (SVMs) were conducted using a leave-one-run-out cross-validation procedure. Furthermore, to evaluate whether stimulus contrast modulates brain response patterns, cross-validations that use low-contrast condition data for training and high-contrast condition data for testing, and vice versa, were also conducted. ANOVAs were then conducted to compare classification accuracies.

Classification Accuracy

To explore the regions involved in the decision making process under threat within the Margin of Safety framework, we examined MVPA classification accuracies using both whole brain searchlight analysis and ROI analysis. We extracted voxel-wise fMRI responses to margin of safety trial (decision phase) as classification samples. For each participant and each run, we designed a general linear model (GLM). The GLM contained 3 regressors indicating the decision phases (duration=4s) of the 3 distribution types, as well as 4 regressors indicating the indication phase (duration=reaction time), motor phase (duration=4s), and feedback phase (duration=3s).

All the regressors were convolved with a canonical hemodynamic response function. In addition, six motion-correction parameters and the linear trend were included as regressors of no interest to account for motion-related artifacts. For each voxel, the parameter estimates of the 3regressors corresponded to the fMRI responses to each of the 3 distributions in each run. The fMRI responses to each distribution item were then entered into the classification analysis as classification samples.

Naturally, there are two main questions we took priority in. First, what brain regions are involved in determining which distribution type the participant is facing and second, what brain regions are involved in determining the MOS decision the participant is making. Thus, we used two sets of classification labels corresponding to the two questions: 1) Normal distribution with normal variance, normal distribution with half variance, and leptokurtic distribution 2) the 50 possible discrete MOS choice options.

We employed a linear support vector machine with a cost parameter C=1 as a classifier. Classification accuracy was estimated using a leave-one-run-out cross-validation: for each of the four runs, a classifier was trained on the other three runs and tested on the remaining focal run; and the procedure was repeated for the four runs (accuracy scores were averaged).

To validate whether the classification performance was significantly above chance, we further conducted Monte Carlo permutation-based statistical tests. This method entailed running a classification analysis 1000 times with randomly permuted experimental condition labels, allowing us to construct null distributions that were used to examine whether a classification accuracy was significantly above chance at an of $p < 0.05$.

4.3 Results

Behavioral Results

We first examined the behavioral data by applying a repeated-measures, One-way ANOVA (Distribution type) for the margin of safety (MOS) escape responses. The result showed a main effect of distribution type [$F(2,44) = 61.33$, $p < 0.001$]. A Tukey post hoc test revealed that participants' MOS choice was significantly more towards the safety zone in the leptokurtic distribution condition (0.74 +- 0.06) than in the normal distribution with matching variance (0.68 +- 0.03). The leptokurtic MOS choice (0.74 +- 0.06) was also significantly more conservative than the MOS choice within the normal distribution with only half of the matched variance (0.67

+- 0.01). This indicates that participants perceived leptokurtic distributions as more difficult to grasp and thus more dangerous, resulting in an overall safer MOS choice set. Interestingly, there's no significant difference in their mean of MOS choices between the two normal distributions. This might suggest that only a fundamental change in the statistical structure of the target distribution can impact participants' decision under threat.

We also examined participant's MOS choices within high/low shock conditions and high/low reward conditions. While there's no significant difference in their escape decisions facing different levels of rewards, their MOS choices were significantly more conservative in the high shock condition (0.75 +- 0.07), compared to the low shock condition (0.69+-0.05) : $t(21) = 21.21$, $p < 0.001$. This suggests that participants were senstivie to the level of danger and adjusted their margin of safety accordingly.

We collected participants' confidence ratings before every unique trial block. An Anova on the confidence ratings also revealed that participants were generally more confident on trials in the two normal distributions compared with trials in the leptokurtic distribution. A main effect of distribution type was found [$F(2,44) = 27.32$, $p < 0.001$], and a Tukey post hoc test showed that confidence rating in the leptokurtic condition (1.42 +- 0.42) was significantly lower than those in the normal distribution with matching mean (2.43 +- 0.68) and normal distribution with half mean (2.65 +- 0.62).

We were also interested in the question whether participants' level of anxiety would influence their MOS decisions. We then estimated a mixed effects linear regression model, with participants' MOSs as dependent variable, and distribution type and STAI-Y scores as the independent variables. Notably, there's only a significant interaction between the leptokurtic distribution type and STAI-Y scores (Beta = 0.43, SE = 0.07, $p < 0.001$), indicating that trait anxiety and MOS were correlated, but only for the leptokurtic distribution.

Whole brain searchlight

For the analysis of distribution types, activations were found in regions including right Insula and the PCC, with a decoding accuracy significantly higher than the monte-caro simulated chance level accuracy ($t(21) = 2.82$, $p = .010$). The whole brain decoding map was thresholded at $P<0.05$ (FWE); for the analysis of decision types, activations were found in regions including right hippocampus and amygdala,

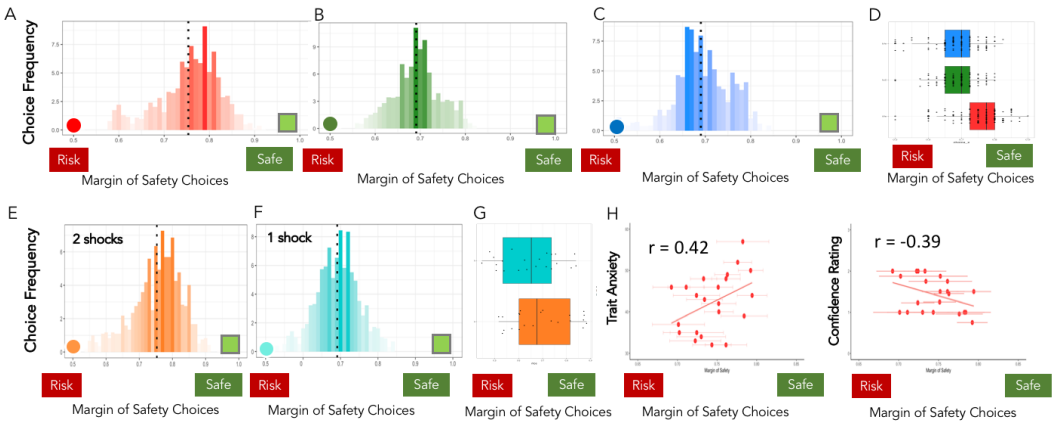

Figure 4.2: Choice frequencies for (A) leptokurtic, (B) matched variance and (C) half variance attacking threats. The avoidance decision phase and the outcome. (D) significant correlation between trait anxiety and pre-emptive avoidance for the leptokurtic condition. This task is run in four sessions over two days (total time 2 hrs).

with a decoding accuracy significantly higher than chance level too ($t(21) = 2.47$, $p = .022$). These results suggested that the MOS decision making process is robustly represented in the above mentioned regions.

At the same time, whole brain searchlights for reward level category and shock level category were also significantly above chance level (shock level: $t(21) = 3.15$,$p < 0.001$; reward level: $t(21) = 2.18$, $p = 0.048$).

ROI analysis

In order to better understand neural mechanisms behind the current decision scenario, and specifically, to isolate key regions implicated in danger (threat encounter) and safer (threat avoidance) conditions, we also performed MVPA classification analysis within a series of ROI identified from previous literature, shown to be critically involved in the process of decision making under threat. These ROIs include the vmPFC, the hippocampus, the amygdala and the insula. Within each specified ROIs, we investigated classification accuracy for the MOS decisions labels, separately for each distribution conditions. Thus, by comparing how well the process is decoded within each ROI, we can tell if one/both/neither of the processes are

ROI-specific. In this case, we could obtain separate and shared representations for the threat encounter and threat avoidance processes.

Within the vmPFC ROI, only classification for the leptokurtic conditions was significantly above chance level (leptokurtic distribution, $p < .001$; normal distribution with matching variance, $p = .365$; normal distribution with half variance, $p = .451$). An Anova showed a main effect of distribution type ($F = 3.475$, $p = .037$), where the accuracy for normal distribution with matching variance is significantly higher than the other two conditions, showing that vmPFC is more robustly decoded in a decision scenario where the main task is to avoid the threat. Within the hippocampus, classification for all 3 distribution types were also all significantly above chance level (leptokurtic distribution, $p < .001$; normal distribution with matching variance, $p = .011$; normal distribution with half variance, $p = .038$). A follow up ANOVA did not reveal a significant difference among the decoding accuracies.

Within the amygdala, decoding accuracy was only significantly above chance level for the leptokurtic distribution (leptokurtic distribution, $p < .001$; normal distribution with matching variance, $p = .213$; normal distribution with half variance, $p = .276$), An Anova showed a main effect of distribution type ($F = 3.875$, $p = .0259$), where the accuracy in leptokurtic distribution is significantly higher than the other two distributions. This indicates that in a condition that's more difficult to decipher (thus more dangerous), amygdala is more robustly decoded. Compared to vmPFC, it is implicated in the direct encounter of imminent threat.

Connectivity analysis

Based on the key regions obtained during MVPA searchlight analysis, we further performed connectivity analysis using gPPI (gPPI; http://www.nitrc.org/projects/gppi), which is configured to automatically accommodate more than two task conditions in the same PPI model by spanning the entire experimental space, compares to the standard implementation in SPM8.

From the MVPA analysis, we took amygdala and vmPFC as seed regions for the leptokurtic distribution contrast and normal distribution contrast respectively, since they were identified as regions representing the process where participants make risk decisions under the corresponding predator conditions. For the amygdala seed, with the contrast [leptokurtic - normal], we observed increased connectivity to regions including the hippocampus, insula and MCC. On the other hand, for the vmPFC seed, with the contrast [normal - leptokurtic], we found increased

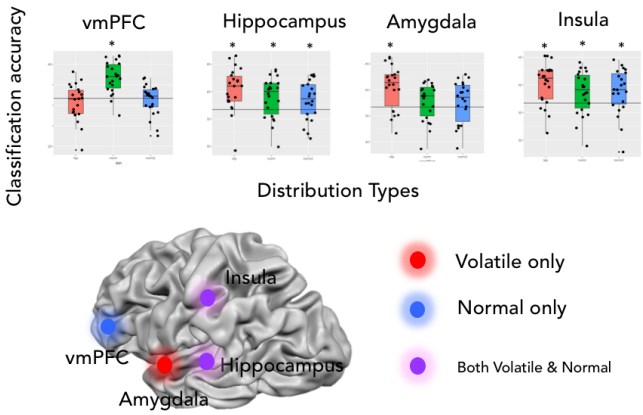

Figure 4.3: Neural representation of pre-emptive avoidance decisions. Avoidance decisions decoded in the vmPFC, Hippocampus, Amygdala and Insula. Pilot results show that the vmPFC was most significant for the predictable threat, while the amygdala was evoked only for the uncertain threat. The hippocampus and insula where activated for all conditions. Box and whisker plots represent the accuracy.

connectivity to regions including the insula and putamen. Insterestingly, negative connectivity with vmPFC was also found with the amygdala seed, suggesting a potential downregulatory process.

What's also interesting, is that when we did the same analysis with the contrast [leptokurtic - control] (20 trials with the same distribution, but without reward and shocks were used as the control condition), we observed increased connectivity to regions including the hippocampus, insula, MCC and PAG. While uncertainty drives different pathways in pre-emptive decisions, the flight circuit is still triggered.

Bayesian decision model: parametric modulation analysis

In order to further explore how participants optimize their decision under the threat, we obtained two measures of optimality from both the original behavioral data and our Bayesian decision making model. We considered a measure of performance optimality related to the per-trial spread between subjects' actual and Bayesian ideal MOS. In particular, we computed the difference between the actual trial-specific utility U(MOS) and the maximum (Bayes optimal) utility the subject could possibly get on the trial [U(MOS)max], given their estimated subjective utilities. A

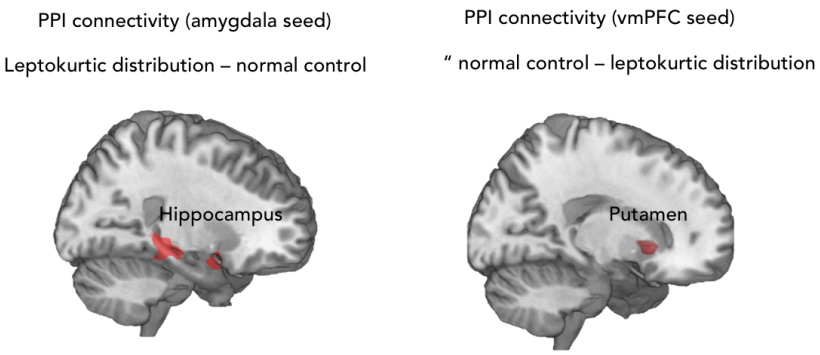

Figure 4.4: Neural activity associated with pre-emptive decisions. gPPI-coopled brain areas using the amygdala seed and the vmPFC seed respectively. Red areas represent significant activations thresholded at $p < 0.05$ (FDR corrected)

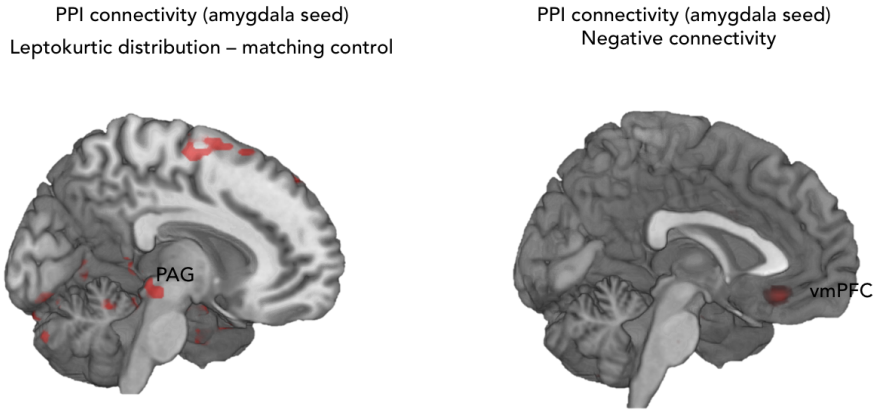

Figure 4.5: gPPI-coopled brain areas using the amygdala seed and the vmPFC seed respectively. Red areas represent significant activations thresholded at $p < 0.05$ (FDR corrected)

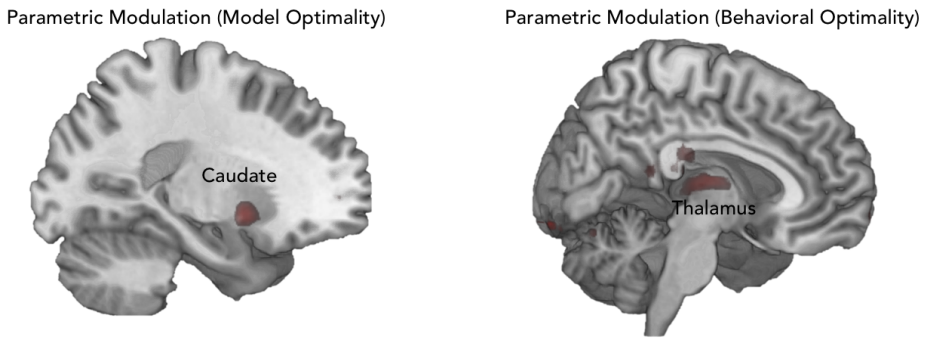

Figure 4.6: Parametric modulation analysis with optimality parameters from the model and actual behavioral measures.Red areas represent significant activations thresholded at $p < 0.05$ (FDR corrected)

smaller difference (e.g., less regret relative to ideal) implies more consistent Bayesian decision making; variation around the ideal FID will increase the difference. On the other hand, behaviorally, we isolated trials where participants nearly escaped the threat - the more "last-minutes" the escape is, the more optimal the actual decision is.

We then performed a first-level parametric analysis with our fMRI data, using these two measures as parametric modulators. We found bilateral caudate to be associated with increased model optimality. On the other hand, for the actual behavioral optimality, the modulation analysis revealed regions including caudate, thalamus, MCC and vmPFC.

4.4 Discussion

As predicted, our behavioral results showed a more conservative decision pattern when the threat avoidance scenario is more ambiguous and unpredicatable. The leptokurtic distribution, which is constructed through the combination of multiple normal distributions centered at the same mean, provides us a statistical tool to facilitate such a scenario. When facing leptokuctic predators, paticipants were presented with threats that had way more outlier situations than normal, rendering their typical min/maxing strategies ineffective. As a result, they choose to place

themselves closer to the safety refuge to avoid the volatile capture risk, compared to the normal distributions. Interestingly, there's no differences in MOS decisions between the two normal distributions, although one of them only had half variance of the other. This suggests that participant's level of uncertain decisions was not swayed by a simple change in distribution variance, but by a total structural change in the distribution, where a continuous random shift in contingency persists.

We are also interested in the neural circuits behind the decision making process. When dissecting the defensive circuitry, it is critical to understand which brain regions are involved in the avoidance of danger. Safety behaviors are actions that are performed to prevent dangerous outcomes from occurring and it is known that prey preempt danger by adopting choices that increase the success of escape. Such safety decisions include how far to move away from a safe refuge or how close one should be to a threat (i.e. margin of safety). Indeed, when human subjects are placed close to a safety exit, measures of fear decrease and when under threat, the sight of safety signals reduces fear20 and fear reinstatement. Safety cues have also been observed to abolish innate defense mechanisms in rodents, such as threat-related analgesia. One region involved in safety signal and learning is the vmPFC. Indeed, the presentation of a safety stimulus during an aversive experience results in increased activity in the vmPFC, while decreasing threat also results in increased activity in the same region. Despite its role in extinction25, anxiety and fear, little is known about the role of the vmPFC in these processes and how this region drives avoidance of danger and down regulates fear and anxiety.

We used MVPA analysis to decipher the neural basis of such decision making patterns. While regions including the hippocampus and insula contribute non-selectivly to both certain and uncertain threat encounters, the amygdala only significantly encoded the more uncertain, leptokurtic threats. The vmPFC, on the other hand, encoded the less uncertain, normal threats. This is interesting because it displays two potential keys regions under different threat scenarios. While amygdala seems to be the processing centre when the risk is more volatile and unpredictable, vmPFC, on the other hand, becomes online only when the threat is "milder" and the risk is more controllable. Theorists have proposed that the hippocampus computes comparators that assess multiple goals and in turn corrects actions possibly through a flexible constructive process involved in problem solving and predictive mapping. When there is time to gather information, the hippocampus may play a role in drawing on previous threat encounters to form a predictive map and optimize current

actions. This supports the hippocampus' role on the processing of uncertain stimuli in a universal, non-selective way.

Leptokurtic noises emerged as a results of complex, large-scale interactions of real life events and information. It provides an excellent source of unpredictable outliers, which adds to the level of uncertainty and difficulty of a given decision task. When the outliers occur due to a continuous random shifts in observational error variance, the distribution of the information source is far more peaked and displays much heavier tails than the Gaussian distribution, and is named "leptokurtic". This leptokurtic noise is rare in the natural environment, hence more difficult to learn about and adapt to.

Our Margin of Safety paradigm, on the other hand, offers an ecological context to assess how humans make decisions under the normal and leptokurtic risk factors. When approached by a potential predator, individuals decide the best timing to flee. We have begun to gather neuroimaging data on the critical neural circuits associated with online decision making of the escaping process. On the other hand, before the online chasing actually happens, individuals could assess risk by monitoring their distance to a safety exit (for animals, their nest) from an approaching threat. By adjusting the distance from the safety exit, individuals can trade off potential threats with potential reward, thus preemptively adapt to the risky environment before the threat actually approaches. We also performed functional connectivity analysis with the critical regions obtained during the MVPA analysis, together with parametric modulation analysis with the Bayesian model parameters. Regions including the Caudate were shown to encode the optimization of threat avoidance decisions. This is consistent with literature where dopamine signaling encoded avoidance learning during aversive events.

The medial and lateral subregions of OFC have been shown to be associated with threat processing, and more specifically, in mediating specific symptoms of anxiety disorders. Rodent studies have shown that macaque with orbitofrontal cortex lesions showed heightened defensive responses. After the lesion, they displayed a greater tendency to express defensive responses even when the threat is absent. This implies a role for the mPFC to downregulate threat avoidance responses, which is consistent with our fMRI data where vmPFC functions as the processing centre when the threat is relatively predictable in the normal distribution condition, compared to the amygdala centre when the more uncertain, leptokurtic threat is present. The vmPFC is also a key player in the defensive survival circuitry. Single-cell recordings in

rodents have shown that the mPFC contains "strategy-selective" cells, which are thought to be involved in the coordination of defensive responses. This fits with the idea that the mPFC plays a role in selecting adaptive strategies that are mapped onto motor responses. Work in humans also shows that larger buffer distances are associated with activity in the vmPFC, and decreased activity in these regions is associated with panic-related motor actions. Building upon this line of research, together with results shown in chapter 2 and 3, our results suggest a tentative role for vmPFC to be a center for modulating pre-emptive decisions under threat.

Our data suggest that when pre-empting impending danger, a distinct set of parallel cortical-subcortical regions code for the threat intensity, uncertainty, and the decision to shift closer to safety.

Chapter 5

DECISIONS AND ANXIETY UNDER UNCERTAINTY, BIASED BY REPUTATION

5.1 Introduction

According to psychological theory, when we are uncertain about what actions to take, we often look to others to guide our decisions [20, 111]. In turn, this leaves us remarkably vulnerable to being influenced by others' opinions or advice. This is common in daily life as important decisions are often based on limited information, and under such circumstances, we often look to the advice of others whom we have a high regard for, namely those with a good reputation. Supporting the conjecture that reputation can have a potent influence on social behavior is research showing that we are highly sensitive to the reputation and the competence of others when making decisions [15, 101, 103]. Further, classic social psychological experiments demonstrate subjects' susceptibility to conforming, that is changing one's behavior to match the opinions and actions of others [7, 21, 93, 96] is especially evident when the influencer is high in authority or reputation [64]. Despite this large body of research, no studies have yet attempted to understand how the reputation of others can modulate social influences on perceptual discrimination.

The information we garner from reputable others and how it influences our decisions become critical when it relates to decision outcomes that result in physical harm. For example, cage-reared monkeys who are not afraid of snakes instantly exhibit fear after they observe another conspecific exhibit fear of snakes [65]. It is known that, in general, the mere presence of a conspecific can ameliorate an individual's response to a stressor, a phenomenon called "social buffering" [46]. The identity of the present conspecific can also change the strength of the buffering response. For example, holding the hand of a romantic partner during the anticipation of a possible electric shock reduced the neural pain responses [22]. Others have shown that when the risk of a shock depends on the task performance of another person, the level of perceived competence extracted from facial features modulates both subjective and neural measures of anxiety [101]. Theories of such findings [10, 48] posit that conspecifics reduce threat responses if they offer protective resources that factor into predictive computations of threat severity. Little is known, however,

about what properties of conspecifics, such as wisdom, experience, competence, or reputation, modulate social buffering.

To explore the relationship between reputation, perceptual bias and emotion, we used functional MRI in conjunction with a standard random dot motion (RDM) discrimination task to test whether partners high or low in reputation would modulate perceptual decisions and anxiety of the subject. To create reputation ratings of the partners, subjects were shown the average ratings of each partner made by their friends [rounded off to either 1 (low), 3 (mid), and 5 (high) reputation]. Because perception and anxiety both rely heavily on expectations to inform interpretations of ambiguous stimuli [89, 100], we hypothesized that a partner's reputation would play a significant role determining the degree of social influence on perception and anxiety. In the task, healthy subjects were provided with RDM estimates by transient collaborators of high or low reputation. The subject's goal was to make RDM estimates based on their own judgements and where needed, incorporate the RDM estimates of the transient collaborator (Figure 5.1). The likelihood of an electric shock at the end of each block is based ostensibly on combined task performance. We show that RepHigh partners induce greater levels of social influence (i.e., conformity) during the RDM discrimination task, especially for the uncertain conditions and down-regulate subjective and neural markers of anxiety.

5.2 Methods

Participants

Twenty-five individuals recruited from the Columbia University community completed all parts of the experiment. Data from three participants were removed due to excessive head motion during scanning (1) or technical issues (shocks not working; 2), and all remaining analyses are based on the remaining 22 individuals (10 females; 12 males; mean age = 25.6 ± 4.6, range 19–35). All participants were right-handed, had normal or corrected-to-normal vision, had no history of neurological or psychiatric illness, and gave written informed consent for participation. The study was approved by the Institutional Review Board of Columbia University.

Experimental paradigm

The experiment consisted of a 42 min session consisting of 48 blocks. In each block (Figure 5.1), the participant was paired up with a new "partner", or someone who ostensibly played the same task on a previous visit to the psychology department. At the beginning of each block, the participant was presented with a picture of that

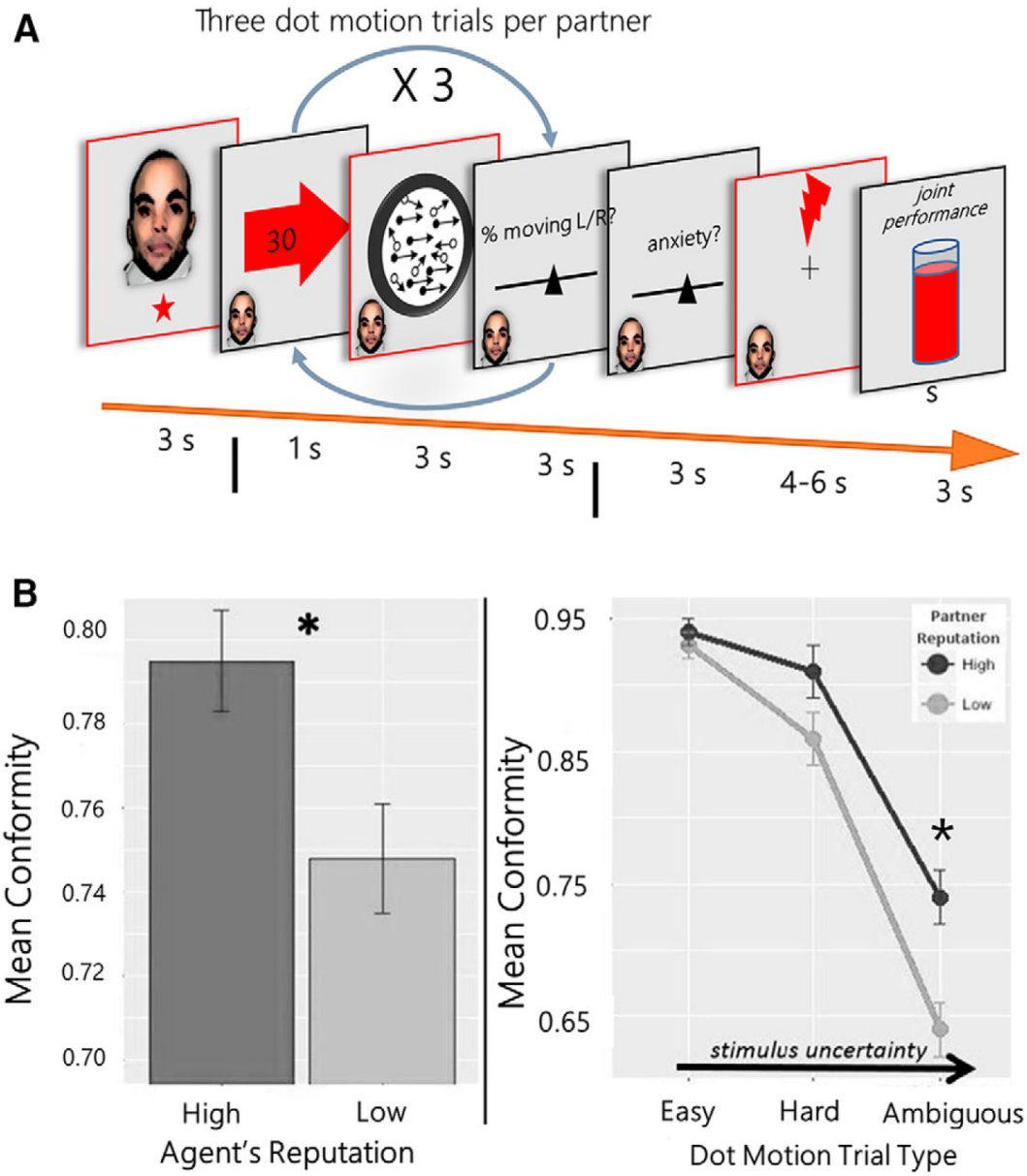

Figure 5.1: Task paradigm and behavioral results. (A) Experimental steps: the subject was first shown a picture of the partner's face and reputation (1 star = RepLow or 5 stars = RepHigh). Next, the subject was shown two screens: (1) an arrow indicating the partner's guess about the direction of the dots (arrow screen) and (2) their guess about the coherence percentage (30% in the example). The subject then saw a screen showing the dot movement and was asked to guess the coherence percentage (using a slider scale). The RDM discrimination estimation was repeated three times. Next, subjects reported how anxious they feel at the prospect of receiving a shock during the 4–6 s anticipation screen. The likelihood of receiving a shock was based on the joint performance accuracy between the subject and the partner. After the shock anticipation screen, they either received a shock or not. Then they saw a screen displaying information about the joint performance of themselves and the partner. The red-bordered boxes are the analyzed events. (B) Left: Conformity was higher for high-reputation partners. Right: Mean conformity differed across hard and ambiguous (uncertain), but not easy conditions. $*p < .001$.

partner along with a star rating below. The pictured partners had neutral facial expressions, were edited to an identical size and brightness, and were unfamiliar to subjects (as confirmed in a debriefing questionnaire). Confounds such as attractiveness, trustworthiness, gender, race, and the like were minimized by counterbalancing the faces across subjects.

Implicit competence was controlled for by using ratings from independent samples of raters from the same subject population (from multiple samples, total N = 91). Raters judged each partner's competence based solely on the partner's picture [101]. Ratings were trifurcated into 1, 3, or 5 stars (with more stars indicating better judged competence). Participants were instructed that the ratings were an average from the partners' friends' judgments of their general skill at perceptual, cognitive, or "mind training" games. This was made more realistic by asking the subject to give the names of friends who could rate them on their perceptual and cognitive abilities and by taking a photograph for future participants in the experiment. Before starting the task each participant was randomly assigned to one of two counterbalanced sets of pairings between partner pictures and reputation ratings. After viewing the partner's picture and star rating for 3 s at the start of the block, the star rating disappeared and a small thumbnail picture of the partner remained in a screen corner. There were also four blocks (2 in each session) without a partner; in those blocks the partner's picture and rating were replaced with "no partner" text.

After the 3 s partner and rating presentation, a 2–4 s jittered ISI occurred before the perception task began. The goal for participants was to make accurate perceptual judgments (3 per block) of the direction and level of motion coherence in the RDM discrimination task. The RDM discrimination task is a standard psychophysical stimulus used to study motion perception [77, 92]. In the RDM, most dots on the screen move in random directions but a portion of the total dots move coherently together either left or right. Coherence levels ranged from 0 to 40%, indicating the percentage of all dots on the screen that are moving together. Participants must judge whether the coherent dots are moving left or right (binary judgment), as well as the numerical level of coherence from 0 to 50% in steps of 5%. As with the rest of the task, dot stimuli were presented using PsychoPy software based in Python, using the built-in Dots Component function with the following parameters: 100 total dots on the screen, 10 pixels dot size, 0.01 U/frame speed, 1000 frame dot lifetime, and noise dots follow a constant direction. Each dot stimulus was presented for 3 s, after which the participant had 3 s to input their judgment by toggling left or right on a

scale that started in the middle at 0% coherence. A 2–4 s jittered ISI followed each rating.

In the blocks that included a partner, participants saw their partner's judgment for 1 s before each dot stimulus in the form of an arrow pointing left or right (indicating the partner's binary judgment) and a number inside the arrow (indicating the partner's continuous coherence judgment). In blocks without a partner, a blank screen replaced the arrow for the 1 s before the dot stimulus. Each block consisted of three dot motion trials, including one "easy" trial (30–40% coherence), one "hard" trial (10–20% coherence), and one "ambiguous" trial (0% coherence). There was no difference in task difficulty or the quality of partner judgments by partner reputation. By definition, however, there were differences in the quality of partner judgments by partner performance levels, such that the total error of the judgments provided by low performing partners (70–80% total error) was higher than that of high performing partners (10–20% total error). Subjects could select "no direction" by selecting 0% coherence on the visual analog scale. Participants were not explicitly instructed about the different levels of task difficulty, or the possibility that some trials would not have a true direction left/right. They were also not instructed on how they should use the information provided by the partner. Accuracy was calculated from the absolute distance between the subject's indicated coherence and the actual coherence of the dots.

Near the end of each block was an anticipation period (4–6 s, jittered) during which participants were instructed that they may receive an electric shock to the left wrist at any point. Before each anticipation period, participants were given 3 s to rate their level of anxiety on a 7-point Likert scale. Participants were told the shocks were probabilistically based on a lower combined performance of both themselves and their partner, such that worse combined performance was associated with a larger probability of shock. For blocks without a partner, participants were instructed that the probability of getting a shock was based solely on their own performance. In reality, all participants received the same number of shocks (5) and those shocks were paired with the same partners within each of the counterbalanced sets of partners. The shock stimuli were delivered using a Biopac MP150 with an STM100C module (Biopac Systems). Attached to the STM100C was a 200 V maximum stimulus isolation unit (STMISOC, Biopac Systems). Shocks were administered via pre-gelled radio translucent electrodes on the underside of the participant's left wrist and attached to the STMISOC with shielded leads. The shocks were calibrated for

each participant before the scan with a procedure that allowed the participant to select a shock level that was uncomfortable but not too painful, with an upper limit of 100 V.

After the anticipation period and a 2–4 s jittered ISI there was a 3 s feedback period. During this time, participants were shown a visual metric (a partially full cylinder) ostensibly indicating the combined performance of the partner and participant in that block. In reality this feedback bar only reflected the partner's performance indexed by the amount of total error in his/her three judgments. Importantly, the partner's performance was orthogonal to the reputation rating, and both were orthogonal to the partner's implicit competence, gender, and race.

fMRI data acquisition and analysis

All fMRI data were acquired using a GE Discovery MR750 3.0 T scanner with 32-channel head coil. The imaging session consisted of two function scans, each 20 min, as well as a high-resolution anatomical T1-weighted image (1 mm isotropic resolution) collected at the beginning of each scan session. For functional imaging, interleaved T2*-weighted gradient-echo echoplanar imaging (EPI) sequences were used to produce 45 3-mm-thick oblique axial slices (TR = 2 s, TE = 25 ms, flip angle = 77°, FOV = 192 × 192 mm, matrix = 64 × 64).

Structural images were subjected to the Unified Segmentation algorithm implemented in SPM8, yielding discrete cosine transform spatial warping coefficients used to normalize (warp) each individual's data (structural and functional) into MNI space. After discarding the first five volumes of each functional run to account for equilibrium effects, the functional data were preprocessed using the following SPM8 functions: slice-time correction, two-pass realignment to correct for head motion (rigid body registration of all frames to the averaged image after first pass), coregistration of each participant's functional mean image to the corresponding structural image, followed by applying the normalization parameters determined during segmentation to the functional images, and then using a 6 mm FWHM Gaussian smoothing kernel.

Statistical analysis of fMRI data

Preprocessed images were subjected to a two-level general linear model using SPM8. The first (individual participant) level contained the following regressors of interest, each convolved with the canonical two-gamma hemodynamic response function: a 3 s boxcar function for the partner and rating presentation period, a 3 s boxcar

function for the dot motion perception period, a 4–6 s (duration-jittered) boxcar function for the anticipation period, and a 3 s boxcar function for the feedback period. In addition, an orthogonal regressor using the mean-centered anxiety ratings parametrically modulating the anticipation period was used, as well as orthogonal regressors using mean-centered dot motion coherence and computed conformity levels parametrically modulating the motion perception period. Regressors of no interest consisted of motion parameters determined during preprocessing, their first temporal derivative, and discrete cosine transform-based temporal low-frequency drift regressors with a cutoff of 192 s.

Beta (regression weight) maps were used to create linear contrast maps (weighted sums of betas), which were then subjected to several second-level, random-effects (summary statistics) one-sample t tests, with the null hypothesis being that the mean over all participants is zero (0). Our model included the partner and rating presentation period (3 s), the motion perception period (3 trials each of 3 s), the anticipation period (4–6 s), and the feedback period (3 s). The analyses only included anticipation periods when a shock did not occur. The resulting statistical maps (one-sample t tests) were thresholded at $p < 0.05$, corrected for multiple comparisons, and false discovery rate (FDR)-corrected [32]. Additionally, for results with strong a priori spatial hypotheses, a small volume correction (SVC) was applied.

Statistical analysis of behavioral data

All behavioral data were analyzed in R. The task was a 2 × 2 factorial design with two main independent variables, partner reputation and partner performance. Each of these two variables had three levels, but there were fewer instances of the midlevel star ratings and performance bars since they were used mainly for psychological validity. The low and high levels of both variables were balanced and were the main conditions of interest in the analyses, therefore the task was treated as a 2 × 2 factorial design for analysis of the anxiety measures. Further, the "lmer" package in R was used to test hypotheses on the repeated-measures data with linear mixed-effect models. With a binary-dependent variable, we used a generalized linear mixed model with a logit link fit by maximum likelihood ("glmerMod"), which provides z-statistics for hypothesis testing of the fixed effects. With a continuous dependent variable, we used a linear mixed model fit by REML ("lmerMod"), which provides *t* statistics using a noninteger Satterthwaite approximation to degrees of freedom for the fixed effects.

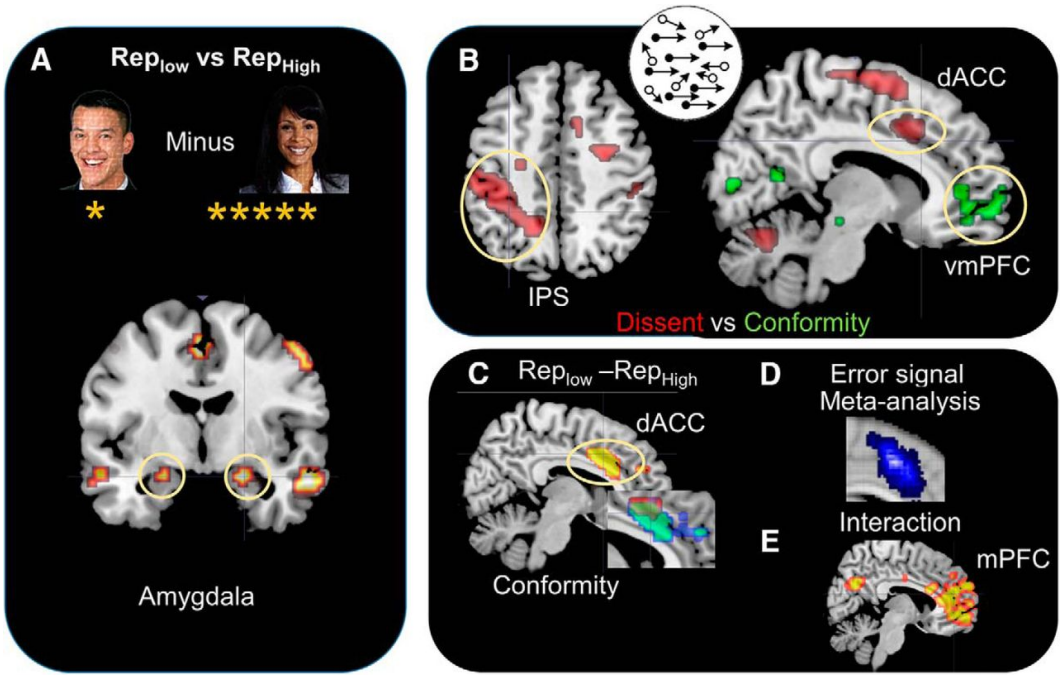

Figure 5.2: (A) Neural activity associated with the presentation of the RepLow compared with RepHigh transient collaborators. (B) Parametric analysis showing brain regions associated with increased dissent and conformity during the RDM task. (C) dACC activity associated with increased conformity with the RepLow compared with RepHigh and inset showing overlap between regions associated with dissent and RepLow conformity. (D) Neurosynth meta-analysis of 357 studies using the search term "Error" (cluster represents a forward inference); (E) medial PFC activity for the 2 × 2 interaction between RepHigh/RepLow × Easy/Uncertain RDM conditions.

5.3 Results

Partner reputation

We first analyzed the fMRI signal in response to the reputation of the partner at the time of the initial partner screen display. The amygdala (MNI-coordinates 18, 0, 24; p = 0.020 SVC), showed increased activity for the RepLow (1 star reputation) when compared directly to the RepHigh (5 star reputation; Figure 5.2A). Other activated regions including the precuneus, hippocampus and mPFC were also activated [$p < 0.05$ whole-brain corrected (WBC); 5.1]. No differential activity was observed for the RepHigh condition compared with RepLow.

Partner reputation and RDM coherence estimates

Next we show that subjects were no more accurate in their judgments of the RDM direction when paired with RepLow compared with RepHigh partner (88.5% vs

Table 5.1: Brain activation for contrast [low reputation > high reputation] (partner period)

Brain region	Left/right	Cluster Size	t score	x	y	z
low reputation >hight reputation (partner)						
Cerebellum	R	1657	8.50	39	-54	-24
Precuneus	L	556	8.32	-3	-54	15
mPFC	R	253	7.77	3	57	-9
Hippocampus	L	130	6.37	-18	-6	-18
Middle temporal gyrus	R	61	6.26	54	0	-21
Superior motor area	L	82	5.44	-3	-6	51
Amygdala*	L	14	3.56	-18	0	-24
All p<0.05, corrected for multiple comparisons (FDR)						
*SVC corrected						

86% correct; t(21) = 1.14, p = 0.13; one-sample t test). The partner's influence on participant judgments, however, was significantly higher for RepHigh compared with RepLow, based on both binary measures of social influence (whether partner and participant judgments of motion direction match; z = 2.48, p = 0.01; 5.1B) as well as a continuous measure (the absolute difference in judgments of motion coherence; t(21) = 2.80, p = 0.01; one-sample t test). There was also an interaction between task difficulty and partner reputation on the amount of social influence on perception. Task difficulty is equated to stimulus uncertainty (lower coherence corresponds to higher stimulus uncertainty and makes the task more difficult). As difficulty increased, the difference in the amount of social influence from RepHigh partner judgments, compared with RepLow, also increased (t(21) = 1.79, p = 0.04; one-sample t test; 5.1C). When the task was easy, there was no significant difference in measures of social influence between RepHigh compared with RepLow partners (apparently due to a ceiling effect in performance). When the task was hard or ambiguous (uncertain), there was significantly more conformity to RepHigh compared with RepLow partners (t(21) = 3.54, p < 0.001; one-sample t test).

To investigate the neural systems underlying these differing levels of social influence between partner reputation conditions, we looked at the fMRI data during RDM task. We first examined the parametric changes associated with increasing conformity and increasing dissent from the partner's judgment (5.2). Increasing conformity

Table 5.2: Brain activation for parametric modulation of conformity

Brain region	Left/right	Cluster Size	t score	x	y	z
Conformity +						
Fusiform	R	34	5.60	24	-69	-12
mPFC	L	182	5.53	-6	57	0
Postcentral gyrus	R	112	5.23	60	-9	30
Insula	R	115	4.97	45	3	3
All p<0.05, corrected for multiple comparisons (FDR)						

was associated with increased activity in the vmPFC (5.2B; 6, 57, 0; p = 0.05; WBC; 5.3). Increasing dissent was associated with increased activity in classic error monitoring areas, namely the dACC (1, 8, 51; p = 0.012 SVC and p < 0.05 WBC) and the intraparietal sulcus (p < 0.05 WBC), a region frequently activated during the RDM task (5.2B; 5.3). The latter finding suggest that the decision to choose an estimate different than the partner's estimate, results in an error or conflict signal (5.2D). Furthermore, the dACC was more active when conforming with the RepLow compared with RepHigh partner (p < 0.05 WBC; 5.3). This region overlapped with the conflict/error signal observed during dissent and with the peak regions found in a meta-analysis of neural activity using the term "error" conducted on 357 studies in Neurosynth (5.2D). Finally, a 2 (RepHigh/RepLow) × 2 (Easy RDM/uncertain RDM conditions) ANOVA was used to examine the neural interaction between easy and uncertain RDM conditions and high and low reputation. This analysis showed activity in the social cognition network including the mPFC, temporoparietal junction (TPJ), and temporal pole in socially influenced perceptual decision-making (p < 0.05 WBC; 5.4).

Finally, we did not observe any behavioral difference between low reputation partner trials and no partner trials. Similarly, when we compared the neural activations between low reputation and no partner conditions, we found no neural differences (at p < 0.5 FDR correction).

Partner reputation and threat-evoked anxiety

Being paired with RepLow compared with RepHigh resulted in suggestively higher levels of subjective anxiety, assessed immediately before the 4–6 s shock anticipation period ($t(21) = 3.65$, $p = 0.002$; one-sample t test; 5.3A). The RepLow and RepHigh

Table 5.3: Brain activation for parametric modulation of conformity (low reputation > high reputation)

Brain region	Left/right	Cluster Size	t score	x	y	z
Conformity + (low - high reputation)						
Caudate	R	57	5.36	13	7	-9
Precentral gyrus	L	48	5.22	-45	-6	48
Superior temporal gyrus	R	94	5.11	58	-29	15
dACC	R	203	4.97	9	23	27
All p<0.05, corrected for multiple comparisons (FDR)						

Table 5.4: Brain activation for interaction between reputation level and task difficulty (task period)

Brain region	Left/right	Cluster Size	t score	x	y	z
Interaction between reputation level and task difficulty (task period)						
Medial frontal cortex	R	1011	6.63	2	58	-2
Precuneus	L	371	5.51	-3	-62	27
TPJ	L	235	5.46	-47	-60	28
ACC	R	184	4.96	3	37	18
Insula	R	150	4.85	42	36	33
All p<0.05, corrected for multiple comparisons (FDR)						

partners performed equally well on average between levels of partner reputation. By design, partner performance was manipulated orthogonally such that some partners performed well at the task and others performed relatively poorly. Actual partner performance was defined as the sum total of error (difference between actual numerical motion coherence and partner's judgment) in the three perceptual judgments looped through for each stimulus. Actual performance modulated anxiety as much as reputation did: mean anxiety was significantly higher when the partner's performance was low compared with high ($t(21) = 3.17$, $p = 0.006$; one-sample t test). There was no interaction between partner reputation and partner performance on anxiety.

To further investigate what caused participants' anxiety, we ran a multiple linear regression model with subjective anxiety ratings as the dependent variable. The

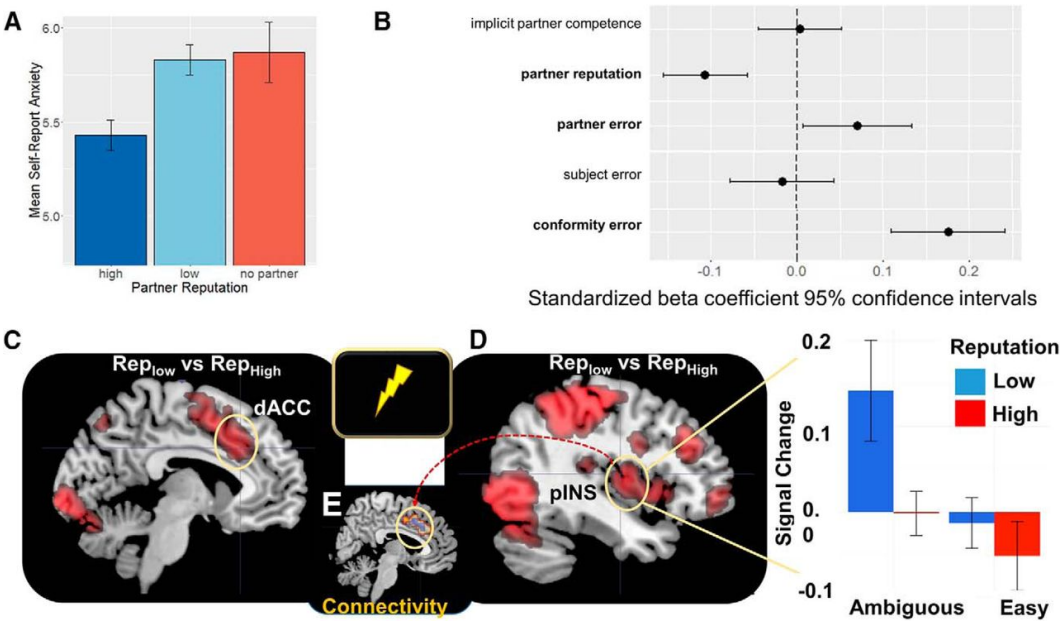

Figure 5.3: (A) Effect of partner reputation on self-report anxiety. (B) Regression coefficients comparing partner reputation to other variables that potentially affect anxiety ratings. (C) fMRI activity during anxious anticipation of potential shock for RepLow compared with RepHigh. Signal change reflects activity in the (C) dACC and (D) pINS. Betas show the differences in activity for RepLow and RepHigh for both the ambiguous (uncertain) and easy RDM conditions, using an independent ROI taken from [101]. (E) connectivity between the pINS (seed) and dACC.

variables included as predictors were the partner's implicit competence, reputation rating (i.e., 1 and 5 stars), partner's performance for the block, participant's performance for the block, and the overall level of social influence in the block (indexed by the total absolute difference between partner and participant numerical coherence judgments). Results of the model indicate that partner reputation, partner performance, and amount of social influence are each significantly associated with anxiety, whereas implicit partner competence and the participant's performance were not significantly associated with anxiety (5.3B).

Next, we analyzed the neural correlates of how partner reputation affected subjective anxiety during the shock anticipation period (trials with actual shocks were removed from the analysis; 5.5). Being paired with a RepLow compared with a RepHigh partner increased activity in anticipatory pain pathways, namely the dACC and bilateral posterior insula (pINS; $p < 0.05$ WBC; 5.3C,D). We also tested whether there were regions selectively more or less active during anticipation as a function of partner actual performance, but no regions survived multiple-comparisons cor-

Table 5.5: Brain activation for contrast [low reputation–high reputation (shock anticipation period)]

Brain region	Left/right	Cluster Size	t score	x	y	z
low reputation–high reputation (shock anticipation period)						
Middle occipital gyrus	L	822	8.88	-21	-96	6
Calcarine	R	761	8.38	12	-99	0
Insula	R	235	7.46	44	14	3
Precentral gyrus	L	625	6.84	-42	-18	57
Middle frontal cortex	R	146	6.25	42	36	33
All p<0.05, corrected for multiple comparisons (FDR)						

Table 5.6: Brain activation for PPI analysis (pINS seed)

Brain region	Left/right	Cluster Size	t score	x	y	z
PPI analysis (pINS seed)						
Insula	L	154	3.90	-39	12	0
Middle frontal gyrus	L	70	3.86	-45	21	33
Precentral gyrus	R	317	3.71	54	0	18
dACC	L	227	3.68	-3	15	39
All p<0.05, corrected for multiple comparisons (FDR)						

rection. Finally, we conducted a psychophysiological (PPI) analysis which showed the connectivity from the pINS to dACC ($p < 0.05$ WBC; 5.3E; 5.6), consistent with earlier findings about pain pathways [109].

5.4 Discussion

How others influence our behavior has been the target of social science research for decades. Based on previous research and theory, we hypothesized that RepHigh partners would create greater levels of social influence on perceptual judgments and lower anxiety toward a potential threat. We found that a partner's reputation did have an impact on the perception of RDM perceptual estimates, especially as stimulus uncertainty increased. We also found that a partner's reputation affected subjective and neural measures of anxiety, and levels of a partner's reputation were more predictive of anxiety than any other measure. Our observations build on social influence frameworks by linking the neural representations of perceived reputation

with decision-making and emotion.

The bilateral amygdala was significantly more active when the subjects were shown the face and reputation (i.e., number of stars) of a RepLow compared with a RepHigh player (5.2A). One explanation for this result is that pairing with subject with a RepLow partner resulted in a threat or saliency signal, potentially associated with the increasing likelihood of diminished performance, and hence, the increased chance of an electric shock. Research shows that the amygdala is evoked during explicit judgements of untrustworthiness [4, 110] , yet these studies do not link such judgments to behavioral outcomes. An alternative explanation might be that the amygdala plays a role in invigorating the subject to work harder at the task, although if this estimation is true, one might expect to see activity in the other dopamine enriched areas, such as the ventral striatum [88] which was not observed, making this conclusion less viable. Thus, the amygdala activity might be a signal for threat and behavioral avoidance [3, 69] associated with the increased potential for harm during performance outcome.

Parametric neural activity in the dACC increased as partner–participant perceptual conflict increased, whereas activity in a vmPFC region increased as the amount of social influence, or perceptual similarity, increased. In line with our findings, this dACC region associated with dissent or social conflict, has previously been linked to conflict or error monitoring, and more ventral mPFC region here associated with conformity may contribute to integrating socially relevant reward signals [68]. Indeed, the most consistent findings in fMRI studies of conformity show the dACC to be involved in anticonformity or dissent [12, 18, 42]. Recent controversy has, however, occluded the role of the dACC, including its role in motivating effortful behavior or expected value of control [95]. Without directly comparing these theories empirically, it is difficult to state the psychological role of the dACC, yet each of these theories provides viable explanations for our data. It is also important to note that we also found increased activity in the IPS, a region involved in perceptual judgments similar to those used in this study [16, 77, 92], where it is possible that this reflects greater recruitment of motion processing regions as the participant's judgment deviated from the partner's, or equivalently, less recruitment of motion processing resources when conforming more closely.

Another key finding of this study was that the participants' perceptions of uncertain stimuli were influenced more by RepHigh than RepLow partners. This reputation effect interacted with task difficulty such that participants were not influenced at

significantly different rates when the task was easy, yet as the task difficulty (i.e., stimulus uncertainty) increased, so did the difference in levels of influence between RepHigh and RepLow partners. This finding supports informational social influence theory, or in our case, the tendency to trust a RepHigh partner's judgment when the RDM task was difficult to estimate [111]. The neural data showed that when the subjects conformed with the RepLow partners, there was increased activity in the dACC, which overlapped with overall increased dissent and a meta-analysis of studies examining the neural basis of error. Although this suggests an alternative view of the dACC to the parametric analysis discussed above, it does fit with the notion that this region is involved in error monitoring or increased perceptual effort. The later theory seems unlikely, however, as activity in the dACC did not correlate with performance ($p = 0.22$, $r = 0.19$). Finally, we conducted an interaction between RepHigh/RepLow × Easy/Uncertain RDM conditions, which revealed a role for the social cognition network, including the TPJ, mPFC, temporal pole, and dACC [99], suggesting that these regions are involved in integrating social information with perceptual judgments.

Chief among our hypotheses is the notion that a partner's reputation can shape subjective ratings of anxiety. Previous work has demonstrated that when one's risk for a shock depends on the task performance of an unfamiliar person, the level of perceived competence in their appearance alone, modulates subjective, and neural measures of anxiety [101]. In the present task, having a RepHigh partner alone decreased anxiety ratings toward a possible shock compared with having a RepLow partner or no partner at all. We also see this decrease in brain regions that respond to the anticipation of shocks including the pINS, somatosensory cortex, and dACC where there was less activity during the anticipation period for the RepHigh compared with the RepLow partner. The pINS region appears to largely overlap with the regions observed in previous pain studies [101], perhaps indicating this region, known to be involved in interoception and pain processing [23, 24], is sensitive to these two different cues of another's reputation and, accordingly, one's risk of harm.

There has been a growing body of literature demonstrating a series of social contexts that influence decision-making and cognition. Studies have shown that information from experts can enhance subsequent memory effects [47] and decision-making processes [15]. Further, the opinions of others and group membership can also modulate choices and value judgments [18, 43]. These contextual effects were associated with activities in the ventral striatum, STS, and prefrontal regions, which

process value during social interaction [68]. Reputation, however, is significant because it shows that without experience, one can be led to trust others opinions. Reputation is likely one of many cues that drives us to conform to others (e.g., expectations, confidence, metacognitive abilities, and trust). Therefore, a future goal is to understand the variety of social variables that lead people to conform to other opinions.

These findings have implications across all fields in which people make judgments in a collective way (e.g., in teamwork), or are influenced by what others think and say. In politics, peer judgments can influence voting behavior based on limited data. In organizations, committees often make decisions about hiring and risky investments aggregating judgments of people who know a lot or a little. In economics and finance "herding" can be rational; helping less-informed investors to rely on the superior knowledge of others. In medicine, patient–practitioner interactions require patients to weigh how much their doctor knows, or to compare initial and second opinions to avoid harmful results. In law enforcement and combat, opinions are shared during life-and-death decisions. Game theory goes even further, by considering the incentives of experts to misrepresent how much they know, and how well consumers of expert opinion adjust for these incentives [25]. Future research, therefore, may be able to elucidate how the effects of reputation described here, reflect more general ways humans integrate cues about peer knowledge in the social environment and how they bias cognition and emotion.

Chapter 6

GENERAL DISCUSSION

6.1 Summary of findings

Summing up, this line of research offered tentative evidence for:

- A differentiation in behavioral portfolio under threat scenarios with divergent characteristics
- A separation of fear into cognitive and reactive circuits
- A unique neural network representing pre-emptive avoidance decisions under uncertainty
- A selective role of anxiety in modulating decisions under threat
- A role of partner social reputation for decisions under threat

We have demonstrated that subjects apply different nodes of the survival circuitry when escaping fast- and slow-attacking threats. Our analysis revealed increased activity in reactive-fear circuits, namely the PAG and the MCC for the fast attacking predator, regions that are implicated in motor response to fast and imminent threats. Our analysis for the slow-attacking threat contrast revealed activation in three key areas of the cognitive fear circuitry involved in more complex information processing—the vmPFC, hippocampus, and PCC.

Our Bayesian model also provides insights into how the distinct regions of the survival circuits are associated with optimal escape. Two core regions were associated with optimal escape: the MCC for the fast-attacking threat and the hippocampus for the slow-attacking threat. The MCC is highly connected to the lateral PAG and according to adaptive control theory is a "central hub" where information about reinforcers are passed to motor control areas to coordinate goal-directed behaviors [94]. Our connectivity results support this conclusion showing that the MCC was coupled with activity in the PAG and the motor cortex. This proposes that the MCC is one candidate region for the integration of current goals and implement aversively motivated instrumental motor behaviors.

Our results also provide evidence that trait anxiety can influence escape decisions, but only under conditions of relatively prolonged threat, compared with more imminent threats. This disassociation implies that trait anxiety selectively affects decisions of different ethnological classes, as distinguished by the amount of time afforded for reflection and cognitive strategizing. Our study provides empirical evidence that trait anxiety selectively impacts escape decisions in humans under this specific class of threat.

The interpretation that trait anxiety affects only cognitive fear behaviour was supported by our neuroimaging results. These results showed that brain areas previously indicated to be involved with behavioural flexibility and information-processing aspects of fear responses (including the hippocampus, amygdala, mPFC and insula) covaried with trait anxiety. However, areas associated with reactive fear—the PAG, superior colliculus, midcingulate cortex and central nucleus of the amygdala—were not significantly affected by variability in anxiety. Notably, these findings strongly support theories based on defensive distance, whereby defensive responses to immediate threats and dangers map onto low-level brain areas such as the PAG, whereas responses to physically or psychologically distal or anticipated threats map to higher-level areas such as the PFC. Our findings extend these theories by providing a clear disassociation of the effects of trait anxiety on one circuit over the other, with accompanying behavioural effects, in an ecologically relevant paradigm.

When uncertainty is involved, our behavioral results showed a more conservative decision pattern when the threat avoidance scenario is more ambiguous and unpredictable. The leptokurtic distribution, which is constructed through the combination of multiple normal distributions centered at the same mean, provides us a statistical tool to facilitate such a scenario. When facing leptokuctic predators, participants were presented with threats that had way more outlier situations than normal, rendering their typical min/maxing strategies ineffective. As a result, they choose to place themselves closer to the safety refuge to avoid the volatile capture risk, compared to the normal distributions. Interestingly, there's no differences in MOS decisions between the two normal distributions, although one of them only had half variance of the other. This suggests that participant's level of uncertain decisions was not swayed by a simple change in distribution variance, but by a total structural change in the distribution, where a continuous random shift in contingency persists.

We used MVPA analysis to decipher the neural basis of such decision making patterns. While regions including the hippocampus and insula contribute non-selectivly

to both certain and uncertain threat encounters, the amygdala only significantly encoded the more uncertain, leptokurtic threats. The vmPFC, on the other hand, encoded the less uncertain, normal threats. This is interesting because it displays two potential keys regions under different threat scenarios. While amygdala seems to be the processing centre when the risk is more volatile and unpredictable, vmPFC, on the other hand, becomes online only when the threat is "milder" and the risk is more controllable. Theorists have proposed that the hippocampus computes comparators that assess multiple goals and in turn correct actions possibly through a flexible constructive process involved in problem solving and predictive mapping. When there is time to gather information, the hippocampus may play a role in drawing on previous threat encounters to form a predictive map and optimize current actions. This supports the hippocampus' role on the processing of uncertain stimuli in a universal, non-selective way.

When social factors (in this case, partner reputation) is involved, we hypothesized that partners with a higher reputation level would create greater levels of social influence on perceptual judgments and lower anxiety toward a potential threat. We found that a partner's reputation did have an impact on the perception of RDM perceptual estimates, especially as stimulus uncertainty increased. We also found that a partner's reputation affected subjective and neural measures of anxiety, and levels of a partner's reputation were more predictive of anxiety than any other measure. Our observations build on social influence frameworks by linking the neural representations of perceived reputation with decision-making and emotion.

6.2 Limitations

In this line of research, we endeavour to design methodologically clean and ecologically valid paradigms. Still, I'm aware that some aspects of the experimental designs were regrettably confounded.

In the first study, where we examined flight initiation distance as a behavioral measure for threat sensitivity, the spatial and temporal components are somewhat mixed together. We define a predator that's accelerating late as "slow", and a predator that's accelerating early as "fast". The time point represents the threat level of the predator. This is thus intended to be a temporal measure of threat imminence. However, if we look at each individual predator, there is also a spatial parameter involved. No matter the predator is "fast" or "slow", psychologically, participants will feel less pressured when it's far away, and more pressured when it's nearby.

Ideally, we'll be able to fine-tune a trial into different components, or add a spatial parameter to control this effect. But this was not possible due to the sensitivity of our measurement.

Also in the FID study, we included a "mid" predator condition, which is intended as an "anchor" condition between the "fast" and "slow" predator conditions. Ideally, we'll be able to observe an intermediate state between the cognitive fear and reactive fear circuits. It is true, behaviorally, that participants responded to the "mid" predator conditions in such a way (their flight initiation distances for the mid predator condition fall between the fast and slow predators). However, we did not observe an intermediate state of neural networks between the cognitive and reactive fear circuits. A more thoughtful experimental design in the future could potentially solve the issue.

Similarly, our manipulation of reward levels seems to bring non-significant behavioral changes. This could be due to the fact that the monetary reward offered in the experiments were trivial compared to the aversion to the possibility of an electrical shock (which is reflected in the reward/punishment parameters in the Bayesian decision model). This issue could potentially be solved by presenting and emphasizing the reward component in a more salient way, or separating the reward component to another independent experiment run.

6.3 Future Directions

As mentioned in the result summary, we've gained some insight into the mechanism of human avoidance decision making under threat, both on a behavioral and neural level. However, a suitable computational model that can adequately address and even predicts individual behaviors is yet to be developed. Our Bayesian decision model achieved moderate prediction accuracy overall, but it "favors" the slow attacking threat more than the fast attacking threat. This may suggest that a simpler, more direct computation might be involved and a pure Bayesian model is not one hundred percent adequate. Due to the nature of the task design, a standard reinforcement learning model seems unlikely, but we are currently working on comparing performances from multiple models against a baseline model to find possible solutions.

Another aspect of follow-up would be utilizing other useful behavioral assets in the animal world. For example, in face of an incoming predatory threat, sometimes it's optimal for an organism to stay still, instead of fleeing, for the predator might have better motion detection than still objects. This "adaptive freezing" behavior can be

modeled into a controlled experiment which offers us the opportunity to study more dynamics within avoidance decisions.

BIBLIOGRAPHY

[1] Adhikari, A., Topiwala, M. A., and Gordon, J. A. (2010). Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety. *Neuron*, 65(2):257–269.

[2] Adhikari, A., Topiwala, M. A., and Gordon, J. A. (2011). Single units in the medial prefrontal cortex with anxiety-related firing patterns are preferentially influenced by ventral hippocampal activity. *Neuron*, 71(5):898–910.

[3] Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, 1191(1):42.

[4] Adolphs, R., Tranel, D., and Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, 393(6684):470–474.

[5] Amat, J., Baratta, M. V., Paul, E., Bland, S. T., Watkins, L. R., and Maier, S. F. (2005). Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature neuroscience*, 8(3):365–371.

[6] An, X., Bandler, R., Öngür, D., and Price, J. (1998). Prefrontal cortical projections to longitudinal columns in the midbrain periaqueductal gray in macaque monkeys. *Journal of Comparative Neurology*, 401(4):455–479.

[7] Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological monographs: General and applied*, 70(9):1.

[8] 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(5):541–547.

[9] Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

[10] Beckes, L. and Coan, J. A. (2011). Social baseline theory: The role of social proximity in emotion and economy of action. *Social and Personality Psychology Compass*, 5(12):976–988.

[11] Benoit, R. G., Davies, D. J., and Anderson, M. C. (2016). Reducing future fears by suppressing the brain mechanisms underlying episodic simulation. *Proceedings of the National Academy of Sciences*, 113(52):E8492–E8501.

[12] Berns, G. S., Chappelow, J., Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., and Richards, J. (2005). Neurobiological correlates of social conformity and independence during mental rotation. *Biological psychiatry*, 58(3):245–253.

[13] Blanchard, R. J. and Blanchard, D. C. (1990). An ethoexperimental analysis of defense, fear, and anxiety.

[14] 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.

[15] Boorman, E. D., O'Doherty, J. P., Adolphs, R., and Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron*, 80(6):1558–1571.

[16] Braddick, O. J., O'Brien, J. M., Wattam-Bell, J., Atkinson, J., Hartley, T., and Turner, R. (2001). Brain areas sensitive to coherent visual motion. *Perception*, 30(1):61–72.

[17] Calhoon, G. G. and Tye, K. M. (2015). Resolving the neural circuits of anxiety. *Nature neuroscience*, 18(10):1394.

[18] Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., and Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170.

[19] Carver, C. S. and White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the bis/bas scales. *Journal of personality and social psychology*, 67(2):319.

[20] Cialdini, R. B. and Cialdini, R. B. (1993). Influence: The psychology of persuasion.

[21] Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., and Winter, P. L. (2006). Managing social norms for persuasive impact. *Social influence*, 1(1):3–15.

[22] Coan, J. A., Schaefer, H. S., and Davidson, R. J. (2006). Lending a hand: Social regulation of the neural response to threat. *Psychological science*, 17(12):1032–1039.

[23] Craig, A. (2003a). Interoception: the sense of the physiological condition of the body. *Current opinion in neurobiology*, 13(4):500–505.

[24] Craig, A. (2003b). A new view of pain as a homeostatic emotion. *Trends in neurosciences*, 26(6):303–307.

[25] Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451.

[26] 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.

[27] Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8(12):1704–1711.

[28] Fanselow, M. S. and Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron*, 65(1):7–19.

[29] 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.

[30] Felix-Ortiz, A. C., Beyeler, A., Seo, C., Leppla, C. A., Wildes, C. P., and Tye, K. M. (2013). Bla to vhpc inputs modulate anxiety-related behaviors. *Neuron*, 79(4):658–664.

[31] Felix-Ortiz, A. C. and Tye, K. M. (2014). Amygdala inputs to the ventral hippocampus bidirectionally modulate social behavior. *Journal of Neuroscience*, 34(2):586–595.

[32] Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878.

[33] Graeff, F. G. (1994). Neuroanatomy and neurotransmitter regulation of defensive behaviors and related emotions in mammals. *Brazilian journal of medical and biological research= Revista brasileira de pesquisas medicas e biologicas*, 27(4):811–829.

[34] Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour research and therapy*, 8(3):249–266.

[35] Gross, C. T. and Canteras, N. S. (2012). The many paths to fear. *Nature Reviews Neuroscience*, 13(9):651–658.

[36] Grosso, A., Santoni, G., Manassero, E., Renna, A., and Sacchetti, B. (2018). A neuronal basis for fear discrimination in the lateral amygdala. *Nature communications*, 9(1):1–12.

[37] Halladay, L. R. and Blair, H. T. (2015). Distinct ensembles of medial prefrontal cortex neurons are activated by threatening stimuli that elicit excitation vs. inhibition of movement. *Journal of neurophysiology*, 114(2):793–807.

[38] Hassabis, D. and Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, 11(7):299–306.

[39] Heilbronner, S., Hayden, B. Y., and Platt, M. (2011). Decision salience signals in posterior cingulate cortex. *Frontiers in neuroscience*, 5:55.

[40] Indovina, I., Robbins, T. W., Núñez-Elizalde, A. O., Dunn, B. D., and Bishop, S. J. (2011). Fear-conditioning mechanisms associated with trait vulnerability to anxiety in humans. *Neuron*, 69(3):563–571.

[41] Ito, R. and Lee, A. C. (2016). The role of the hippocampus in approach-avoidance conflict decision-making: evidence from rodent and human studies. *Behavioural brain research*, 313:345–357.

[42] Izuma, K. and Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron*, 78(3):563–573.

[43] Izuma, K., Saito, D. N., and Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, 58(2):284–294.

[44] Jimenez, J. C., Su, K., Goldberg, A. R., Luna, V. M., Biane, J. S., Ordek, G., Zhou, P., Ong, S. K., Wright, M. A., Zweifel, L., et al. (2018). Anxiety cells in a hippocampal-hypothalamic circuit. *Neuron*, 97(3):670–683.

[45] 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.

[46] Kikusui, T., Winslow, J. T., and Mori, Y. (2006). Social buffering: relief from stress and anxiety. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476):2215–2228.

[47] Klucharev, V., Smidts, A., and Fernández, G. (2008). Brain mechanisms of persuasion: how 'expert power' modulates memory and attitudes. *Social cognitive and affective neuroscience*, 3(4):353–366.

[48] Krahé, C., Springer, A., Weinman, J. A., and Fotopoulou, A. K. (2013). The social modulation of pain: others as predictive signals of salience–a systematic review. *Frontiers in human neuroscience*, 7:386.

[49] Krout, K. E. and Loewy, A. D. (2000). Periaqueductal gray matter projections to midline and intralaminar thalamic nuclei of the rat. *Journal of Comparative Neurology*, 424(1):111–141.

[50] Lawrence, M. A. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments*. R package version 4.4-0.

[51] LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73(4):653–676.

[52] LeDoux, J. E. and Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: a two-system framework. *American journal of psychiatry*.

[53] Malvaez, M., Greenfield, V. Y., Wang, A. S., Yorita, A. M., Feng, L., Linker, K. E., Monbouquette, H. G., and Wassum, K. M. (2015). Basolateral amygdala rapid glutamate release encodes an outcome-specific representation vital for

reward-predictive cues to selectively invigorate reward-seeking actions. *Scientific reports*, 5(1):1–17.

[54] 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.

[55] Mathews, A. (1990). Why worry? the cognitive function of anxiety. *Behaviour research and therapy*, 28(6):455–468.

[56] Mathews, A. and Mackintosh, B. (1998). A cognitive model of selective processing in anxiety. *Cognitive therapy and research*, 22(6):539–560.

[57] McCoy, A. N. and Platt, M. L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nature neuroscience*, 8(9):1220–1227.

[58] McHugh, S., Deacon, R., Rawlins, J., and Bannerman, D. M. (2004). Amygdala and ventral hippocampus contribute differentially to mechanisms of fear and anxiety. *Behavioral neuroscience*, 118(1):63.

[59] McLaren, D. G., Ries, M. L., Xu, G., and Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gppi): a comparison to standard approaches. *Neuroimage*, 61(4):1277–1286.

[60] McNaughton, N. and Corr, P. J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience & Biobehavioral Reviews*, 28(3):285–305.

[61] McNaughton, N. and Corr, P. J. (2018). Survival circuits and risk assessment. *Current Opinion in Behavioral Sciences*, 24:14–20.

[62] McNaughton, N. and Gray, J. A. (2000). Anxiolytic action on the behavioural inhibition system implies multiple types of arousal contribute to anxiety. *Journal of affective disorders*, 61(3):161–176.

[63] Meacham, F. and T. Bergstrom, C. (2016). Adaptive behavior can produce maladaptive anxiety due to individual differences in experience. *Evolution, medicine, and public health*, 2016(1):270–285.

[64] Milgram, S. (1964). Group pressure and action against a person. *The Journal of Abnormal and Social Psychology*, 69(2):137.

[65] Mineka, S. and Cook, M. (1986). Immunization against the observational conditioning of snake fear in rhesus monkeys. *Journal of abnormal psychology*, 95(4):307.

[66] Mobbs, D. (2018). The ethological deconstruction of fear (s). *Current opinion in behavioral sciences*, 24:32–37.

[67] Mobbs, D., Hagan, C. C., Dalgleish, T., Silston, B., and Prévost, C. (2015a). The ecology of human fear: survival optimization and the nervous system. *Frontiers in neuroscience*, 9:55.

[68] Mobbs, D., Hagan, C. C., Yu, R., Takahashi, H., FeldmanHall, O., Calder, A. J., and Dalgleish, T. (2015b). Reflected glory and failure: the role of the medial prefrontal cortex and ventral striatum in self vs other relevance during advice-giving outcomes. *Social cognitive and affective neuroscience*, 10(10):1323–1328.

[69] Mobbs, D., Hassabis, D., Yu, R., Chu, C., Rushworth, M., Boorman, E., and Dalgleish, T. (2013). Foraging under competition: the neural basis of input-matching in humans. *Journal of Neuroscience*, 33(23):9866–9872.

[70] 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.

[71] Mobbs, D., Marchant, J. L., Hassabis, D., Seymour, B., Tan, G., Gray, M., Petrovic, P., Dolan, R. J., and Frith, C. D. (2009). From threat to fear: the neural organization of defensive fear systems in humans. *Journal of Neuroscience*, 29(39):12236–12243.

[72] 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.

[73] Mobbs, D., Yu, R., Rowe, J. B., Eich, H., FeldmanHall, O., and Dalgleish, T. (2010). Neural activity associated with monitoring the oscillating threat value of a tarantula. *Proceedings of the National Academy of Sciences*, 107(47):20582–20586.

[74] Moscarello, J. M. and Maren, S. (2018). Flexibility in the face of fear: Hippocampal–prefrontal regulation of fear and avoidance. *Current opinion in behavioral sciences*, 19:44–49.

[75] Moser, M.-B. and Moser, E. I. (1998). Functional differentiation in the hippocampus. *Hippocampus*, 8(6):608–619.

[76] Nesse, R. M. et al. (1994). Fear and fitness: An evolutionary analysis of anxiety disorders. *Ethology and sociobiology*, 15(5-6):247–261.

[77] Newsome, W. T., Britten, K. H., and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341(6237):52–54.

[78] Oehrn, C. R., Baumann, C., Fell, J., Lee, H., Kessler, H., Habel, U., Hanslmayr, S., and Axmacher, N. (2015). Human hippocampal dynamics during response conflict. *Current Biology*, 25(17):2307–2313.

[79] Padilla-Coreano, N., Bolkan, S. S., Pierce, G. M., Blackman, D. R., Hardin, W. D., Garcia-Garcia, A. L., Spellman, T. J., and Gordon, J. A. (2016). Direct ventral hippocampal-prefrontal input is required for anxiety-related neural activity and behavior. *Neuron*, 89(4):857–866.

[80] Panksepp, J. (1998). The periconscious substrates of consciousness: Affective states and the evolutionary origins of the self. *Journal of consciousness studies*, 5(5-6):566–582.

[81] Panksepp, J. (2011). The basic emotional circuits of mammalian brains: do animals have affective lives? *Neuroscience & Biobehavioral Reviews*, 35(9):1791–1804.

[82] Parent, M. A., Wang, L., Su, J., Netoff, T., and Yuan, L.-L. (2010). Identification of the hippocampal input to medial prefrontal cortex in vitro. *Cerebral cortex*, 20(2):393–403.

[83] Paulus, M. P., Rogalsky, C., Simmons, A., Feinstein, J. S., and Stein, M. B. (2003). Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. *Neuroimage*, 19(4):1439–1448.

[84] Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., and Platt, M. L. (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends in cognitive sciences*, 15(4):143–151.

[85] Peck, C. J., Lau, B., and Salzman, C. D. (2013). The primate amygdala combines information about space and value. *Nature neuroscience*, 16(3):340.

[86] Perkins, A. M., Arnone, D., Smallwood, J., and Mobbs, D. (2015). Thinking too much: Self-generated thought as the engine of neuroticism. *Trends in cognitive sciences*, 19(9):492–498.

[87] Perkins, AM & Corr, P. J. (2014). Anxiety as an adaptive emotion. In Parrott, G., editor, *The Positive Side of Negative Emotions*, page 37. The Guilford Press.

[88] Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., and Frith, C. D. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science*, 316(5826):904–906.

[89] Petrovic, P., Dietrich, T., Fransson, P., Andersson, J., Carlsson, K., and Ingvar, M. (2005). Placebo in emotional processing—induced expectations of anxiety relief activate a generalized modulatory network. *Neuron*, 46(6):957–969.

[90] Price, J. L. (2005). Free will versus survival: brain systems that underlie intrinsic constraints on behavior. *Journal of Comparative Neurology*, 493(1):132–139.

[91] R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[92] Raymond, J. E. (2000). Attentional modulation of visual motion perception. *Trends in Cognitive Sciences*, 4(2):42–50.

[93] Schachter, S. (1951). Deviation, rejection, and communication. *The Journal of Abnormal and Social Psychology*, 46(2):190.

[94] Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, 12(3):154–167.

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

[96] Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology (Columbia University)*.

[97] Stachenfeld, K. L., Botvinick, M. M., and Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature neuroscience*, 20(11):1643.

[98] Stankowich, T. and Blumstein, D. T. (2005). Fear in animals: a meta-analysis and review of risk assessment. *Proceedings of the Royal Society B: Biological Sciences*, 272(1581):2627–2634.

[99] Stanley, D. A. and Adolphs, R. (2013). Toward a neural basis for social behavior. *Neuron*, 80(3):816–826.

[100] Sterzer, P., Frith, C., and Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Current Biology*, 18(16):R697–R698.

[101] Tedeschi, E., Weber, J., Prévost, C., Mischel, W., and Mobbs, D. (2015). Inferences of others' competence reduces anticipation of pain when under threat. *Journal of cognitive neuroscience*, 27(10):2071–2078.

[102] Therneau, T. M. (2015). *coxme: Mixed Effects Cox Models*. R package version 2.2-5.

[103] Todorov, A., Mandisodza, A. N., Goren, A., and Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728):1623–1626.

[104] Tovote, P., Esposito, M. S., Botta, P., Chaudun, F., Fadok, J. P., Markovic, M., Wolff, S. B., Ramakrishnan, C., Fenno, L., Deisseroth, K., et al. (2016). Midbrain circuits for defensive behaviour. *Nature*, 534(7606):206–212.

[105] Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

[106] Trimmer, P. C., Houston, A. I., Marshall, J. A., Bogacz, R., Paul, E. S., Mendl, M. T., and McNamara, J. M. (2008). Mammalian choices: combining fast-but-inaccurate and slow-but-accurate decision-making systems. *Proceedings of the Royal Society B: Biological Sciences*, 275(1649):2353–2361.

[107] Vertès, A. A. (2015). The potential of cytotherapeutics in hematologic reconstitution and in the treatment and prophylaxis of graft-versus-host disease. chapter i: current practice and remaining unmet medical needs. *Regenerative medicine*, 10(3):331–343.

[108] Vertes, R. P. (2004). Differential projections of the infralimbic and prelimbic cortex in the rat. *Synapse*, 51(1):32–58.

[109] Wilcox, C. E., Mayer, A. R., Teshiba, T. M., Ling, J., Smith, B. W., Wilcox, G. L., and Mullins, P. G. (2015). The subjective experience of pain: an fmri study of percept-related models and functional connectivity. *Pain medicine*, 16(11):2121–2133.

[110] Winston, J. S., Strange, B. A., O'Doherty, J., and Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature neuroscience*, 5(3):277–283.

[111] Wooten, D. B. and Reed II, A. (1998). Informational influence and the ambiguity of product experience: Order effects on the weighting of evidence. *Journal of consumer psychology*, 7(1):79–99.

[112] Xue, G., Lu, Z., Levin, I. P., and Bechara, A. (2010). The impact of prior risk experiences on subsequent risky decision-making: the role of the insula. *Neuroimage*, 50(2):709–716.

[113] Ydenberg, R. C. and Dill, L. M. (1986). The economics of fleeing from predators. *Advances in the Study of Behavior*, 16(C):229–249.

[114] Young, C. K. and McNaughton, N. (2009). Coupling of theta oscillations between anterior and posterior midline cortex and with the hippocampus in freely behaving rats. *Cerebral Cortex*, 19(1):24–40.

INDEX