REAL-WORLD SOCIAL COGNITION:

Context Effects in Face and Threat Processing

Thesis by
Laura A. Harrison

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California
2015
(Defended May 27, 2015)

ACKNOWLEDGEMENTS

This thesis heralds the importance of context, proclaiming that social cognition does not occur in a vacuum. This thesis was not written in a vacuum. Fittingly, I owe many people a debt of gratitude for contributing to the social-developmental milieu that enabled its production.

Academically, I have a long list of people to thank and only have space to highlight a few. Since my kindergarten teacher, Mrs. Kaplow, encouraged my youthful passion for marine biology with “Sea Shell” the sea otter puppet, several other teachers and professors have supported and encouraged me as a potential scientist. A few stand out and deserve special thanks. Mr. Davis from my high school for making physics and calculus fun and for giving me confidence in my scientific abilities. Dr. Mike Barish from City of Hope and Drs. Janette Atkinson and Oliver Braddick from University College London and Oxford University gave me my first research experiences, convincing me to become a scientist. My undergraduate thesis advisor, Dr. Michael Spezio, deserves special thanks for guiding my decision to pursue a Ph.D. and inspiring the kernel of this thesis with my undergraduate thesis work (cf. Spezio, et al. (2012)).

For actually making me a scientist, I am sincerely grateful to Dr. Ralph Adophs. Ralph has been the advisor that most graduate students dream of —insightful, knowledgeable, creative, dedicated, supportive, and inspiring. Working with him has been an honor. It has also been a lot of fun. In graduate school, I retained the excitement that led me here in the first instance. An experience that I will never forget and which encapsulates Ralph as an advisor is the planning of my thesis defense. On Catalina Island, at our annual lab camping trip, Ralph helped me lay out printed slides from my defense, weighted down against the wind with firewood and propane canisters. We then went through the slides, discarding, re-ordering, and suggesting improvements. The process was memorable and effective, although I worried about the whole thing literally going up in flames! Seriously though, Ralph has been a valuable advisor in many aspects, advising not just scientifically, but holistically. For example, through semi-regular meetings, he encourages lab members to reflect on their careers and career choices. Thanks to his help in this process, I am leaving graduate school excited to embark on my postdoc, optimistic about my career, and still passionate about neuroscience.

Alongside Ralph, the other members of my thesis committee — Drs. John O'Doherty, Pietro Perona, Shinsuke Shimojo, and Doris Tsao — deserve my sincere gratitude. Throughout the past several years, through my thesis committee meetings, they have helped shape the direction of this thesis, improving each of the included projects. In addition to technical advice, I thank them for important advice on both nurturing several projects at once and conducting my research not only diligently but also passionately.

Being a part of Caltech has been an incredible experience for which I am grateful. I have many people to thank for contributing to that experience, including members of the Adolphs lab, past and present, HSS and CNS staff, and my fellow students.

I would like to thank all the undergraduate students I have mentored over the years, and name those whose work contributed to this thesis: Curie Ahn and Poonim Daya.

A scientific career is vocational and leaks into all aspects of one's life. For his support throughout my Ph.D., as well as dedicating himself to support me throughout my life, including in my chosen vocation, I in turn dedicate this thesis to my dear husband, Ben Harrison. I sincerely thank him and the rest of my family, including my parents, Ray and Cindy Loesch, my siblings Jennifer and Stephen Loesch, and my grandparents Bill and Jo Anne Schutz, for their support and encouragement throughout the years.

Last, but not least, I would like to express gratitude for the funding that supports this research and the participants who both inspire the research and make it possible.

ABSTRACT

As borne out by everyday social experience, social cognition is highly dependent on context, modulated by a host of factors that arise from the social environment in which we live. While streamlined laboratory research provides excellent experimental control, it can be limited to telling us about the capabilities of the brain under artificial conditions, rather than elucidating the processes that come into play in the real world. Consideration of the impact of ecologically valid contextual cues on social cognition will improve the generalizability of social neuroscience findings also to pathology, e.g., to psychiatric illnesses. To help bridge between laboratory research and social cognition as we experience it in the real world, this thesis investigates three themes: (1) increasing the naturalness of stimuli with richer contextual cues, (2) the potentially special contextual case of social cognition when two people interact directly, and (3) a third theme of experimental believability, which runs in parallel to the first two themes. Focusing on the first two themes, in work with two patient populations, we explore neural contributions to two topics in social cognition. First, we document a basic approach bias in rare patients with bilateral lesions of the amygdala. This finding is then related to the contextual factor of ambiguity, and further investigated together with other contextual cues in a sample of healthy individuals tested over the internet, finally yielding a hierarchical decision tree for social threat evaluation. Second, we demonstrate that neural processing of eye gaze in brain structures related to face, gaze, and social processing is differently modulated by the direct presence of another live person. This question is investigated using fMRI in people with autism and controls. Across a range of topics, we demonstrate that two themes of ecological validity — integration of naturalistic contextual cues, and social interaction — influence social cognition, that particular brain structures mediate this processing, and that it will be crucial to study interaction in order to understand disorders of social interaction such as autism.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Chapter 1

INTRODUCTION

"I had," said he, "come to an entirely erroneous conclusion, which shows my dear Watson, how dangerous it always is to reason from insufficient data."

Sherlock Holmes character (Doyle, 1892/2002)

Sherlock Holmes' observation that it is dangerous to draw conclusions from insufficient data is an extreme statement of an important reminder for cognitive and social neuroscientists. It is unclear when data is sufficient, but it is equally clear that removal of crucial variables from a model will still elicit a model, albeit one that is profoundly different from one that would have included the ignored variables. Currently, the predominant scientific approach is that of systematic, as opposed to representative, design; variables are carefully isolated, and their effect on behavior, cognition, and neural activity tested. While this approach undeniably generates valuable findings, a lingering concern is that it reveals the brain's capacity for processing those stimuli when tested in isolation in the laboratory, rather than its real-world processing tendencies when those same variables are contextualized alongside a myriad array of co-occurring variables.

In an attempt to ameliorate these concerns, "ecologically valid" or naturalistic methods are currently *en vogue*. A movement towards interactive social neuroscience is especially popular. This first chapter provides a framework for evaluating these methods, and concludes with an overview of the second and third chapters of this thesis. The second chapter examines the role of the amygdala in threat processing, beginning by proposing a general function of the amygdala in assessing saliency and self relevance, showing that amygdala damage promotes a general rather than stimulus-specific default approach bias, and finally explores how situational cues are hierarchically assessed to elicit specific defensive behaviors to both physical and psychological social threats. This hierarchical model is proposed as a tool that could probe the specific contribution of the amygdala and other structures to guiding defensive behavior. Finally, assessing a profound experimental manipulation, the third chapter tests the fundamental question of whether non-interactive methods adequately elicit the social processing we seek to study.

Chapter 1.1

REAL LIFE OR FANTASY? A FRAMEWORK FOR EVALUATING NATURALISTIC METHODS IN SOCIAL NEUROSCIENCE

Imagine you're an undergraduate student at your university. Walking through campus, you see a sign:

PARTICIPATE IN A PSYCHOLOGY EXPERIMENT!

EARN $20/HR!

$20 can buy you 200 packets of ramen noodles. You approach the sign and tear off a tab with the experimenter's contact information.

You decide to participate in the experiment. Later that night, you sign a few forms and sit down in front of a computer in a university laboratory. A psychology grad student turns to you:

> "In this experiment, you'll look at pictures of people and judge them. We're interested in which faces you find threatening. After seeing a face, please rate it on a scale of 1 to 7: use 1 if the face isn't at all threatening and 7 if it's very threatening. 4 is neutral. Enter your response as quickly as possible. Do you have any questions?"
>
> You shake your head 'no'.
>
> "Okay, press the space-bar to begin."

A hairless, computer-generated face appears. It stares blankly at you. Is it threatening? Yeah, he's a bit dodgy. You don't feel too strongly about it though. Maybe he's a 6? Then, again, he's not too bad — maybe a 4? You decide to compromise.

Tentatively, you press 5 on the keyboard.

An hour and countless faces later, with $20 in your pocket, you leave the laboratory and head for the parking lot.

On the way home, you stop at Walmart to exchange the cash for ramen. While you're getting back in your car, a man approaches you. Unlike in the lab, you don't hesitate. He's definitely dodgy. Decisively, you jump in your car, jab the lock button, and drive off.

❧❧❧

As neuroscientists, we have our experimental participants complete experiments like this one every day. What is the difference between the two described experiences of fear? There's the obvious difference: one was real and the other was a laboratory experiment. Other differences are subtler. The real man had hair. He moved. He was in a parking lot. Which of these differences matter more to our brains?

As scientists, we propose that the cognitive principles and models we uncover relate to *in situ* cognition. However, it is becoming clear that this belief is often unrealistic. A renewed focus on ecological validity aims to remedy this disconnect between the laboratory and real life, between the *capacity* of the brain to behave in a certain way under specific laboratory conditions and a *tendency* to react to those same factor manipulations in a completely different way in the unconstrained outside world.

Currently, the field is rife with discussion paragraphs and reviews dedicated to the topic of ecological validity, broadly and widely construed[1]. Despite heavy theoretical interest, empirical attempts remain relatively scattershot. To be sure, more than lip service is paid to the concept, but the overall attempt is minimally organized. In this chapter, a framework for studying ecological validity is proposed and used as motivation for the set of studies that follow in this thesis.

Ecological Validity Defined

The first step in providing a framework for this area of research is to define ecological validity and highlight motivations for its pursuit. While currently fashionable, the idea of ecological validity is not new:

> Increasing numbers of cognitive psychologist and neuropsychologists are moving beyond the laboratory and attempting to understand human cognitive abilities as they are manifest in natural contexts (p. xi) (Poon, 1989).

That statement was made in the 1980s when psychologists researching memory debated the merits of studying “everyday” cognition – analogous to the “real world” social cognition investigated here. That debate generated two important observations to keep in mind throughout this thesis. The first is that there are “theoretical and methodological trade-offs and dilemmas” (Poon, 1989) involved in this kind of research; as such, we should consider the “how, when, and why” for studying real-world as opposed to traditional laboratory cognition (Rubin, 1989). The second observation is that laboratory and real-world research lie on a continuum (Poon, 1989). At one far end of the spectrum lies experimental manipulation of isolated, low-level variables (e.g., edge

---

[1] Including but not limited to the following concepts: cognitive ethology, real-world, real-life or everyday life, naturalistic, interactive, first and second-person neuroscience, embodied cognition, representative design, and

detection); ethology and pure observation lie at the other. Excitingly, mixing this continuum, computational ethology (Anderson & Perona, 2014) is an emerging field, employing automated behavioral classification and detailed environmental manipulation, that makes ethology more experimental. Combined with tools to manipulate neural systems in organisms, this field has significant potential for detailing neural circuits underlying ecologically valid behavior.

Within this thesis, ecological validity is conceived as capturing the tendency rather than capability of the brain, and of utilizing stimuli[2] that elicit patterns of response that capture the most variance in actual everyday behavior. Note that this definition does not automatically make unnatural stimuli ecologically invalid and naturalistic stimuli ecologically valid. This distinction draws on the Bruswikian concept of representative design, developed in the 1950s, which is utilized within the strategy of probabilistic functionalism. Probabilistic functionalism assumes that “behavior takes place in a semichaotic medium that contains cues of limited trustworthiness, expressed vicariously,” (Petrinovich, 1989) requiring a research strategy that samples a broad array of cues. Petrinovich posits that Brunswik’s approach of representative design can be “conceived as an exercise is sampling theory” (1989), which pits representative design against systematic design; both make different assumptions about experimental sampling but share a common goal of making generalizable theoretical conclusions about the mind (and, in our case, the brain). Under systematic design, experimenters “systematically include and exclude factors and manipulate variables systematically as the investigator

---

[2] While this definition focuses on ecological validity of the experimental manipulation, it is of course also important to consider the validity of our behavioral measures and generalizability of our sample population.

deems useful and appropriate.” Alternatively, representative design assumes it is necessary to “understand the situation in which stimuli are encountered, and then representatively sample stimuli from that population of situations.” To do this, representative design requires sampling “distal” stimuli, i.e., situational or contextual cues that influence the more proximal cause of behavior. Petrinovich makes a strong claim that the former is nomothetic, seeking general laws, while the latter is idiographic, observing behavior of isolated cases, but that divide is too severe, especially in modern treatment: for example, even strict proponents of systematic design consider, and model, individual differences, something Petrinovich (1989) did not consider compatible with the traditional systematic approach; similarly, proponents of representative design today still seek relatively general laws.

To reiterate: the concept of ecological validity used in this thesis is similar to Bruswikian representative design, which refers to an experiment’s “quality of naturalness, or lifelikeness” – some of these cues may be strictly “ecologically valid” in Brunswik’s conceptualization, in which ecological validity relates to the “potential utility of various cues for organisms in their ecology” (Petrinovich, 1989). Simply put, we are interested in contextual cues, often naturalistic or lifelike, that have the most influence on patterns of behavior and cognition as they occur in the real world. While the terms we use are admittedly “fuzzy”, that imprecision is intentional in order to preserve a broad concept that encompasses three main classes of ecological validity, discussed below.

Framework for Ecologically Valid Social Neuroscience

Subsequent to defining the concept of ecological validity, a second goal is to introduce a framework for ecological validity in social neuroscience. There are two main lines of work conducted in ecological validity in social neuroscience, and a third important area to consider. The first centers around the theme of context, the second around interaction, and the third around believability.

Adding Contextual Cues. The first and most prominent theme of ecological validity is to add cues to make a laboratory stimulus perceptually realistic. Often, this involves making the stimulus as naturalistic as possible (e.g., returning to our original example, using photographs, or, even better, dynamic videos, instead of computer-generated facial images). While stimulus manipulation can make stimuli as close to a real-world representation as possible, another way to manipulate context to make a cue "realistic" is to instead simply influence high-level representations and beliefs to create an artificial percept of "realness." For example, the neural response to manipulation of interpersonal space (know as "proxemics") was successfully investigated by convincing participants that someone was approaching them in the scanning room, not with any change in perceptual cues, but through verbal instruction, telling participants that someone was approaching them (Kennedy, Glaescher, Tyszka, & Adolphs, 2009). Notably, this second type of contextual manipulation (i.e., non-perceptual) can only be manipulated in humans, whose beliefs can be verbally informed.

Contextual cues influence all levels of cognition. For example, starting with perception, the fusiform face area, a region that exhibits specialization for faces relative to other objects, is actually most responsive to the context of a face, rather than an actual face (Figure 1.1.1) (Cox, Meyers, & Sinha, 2004).

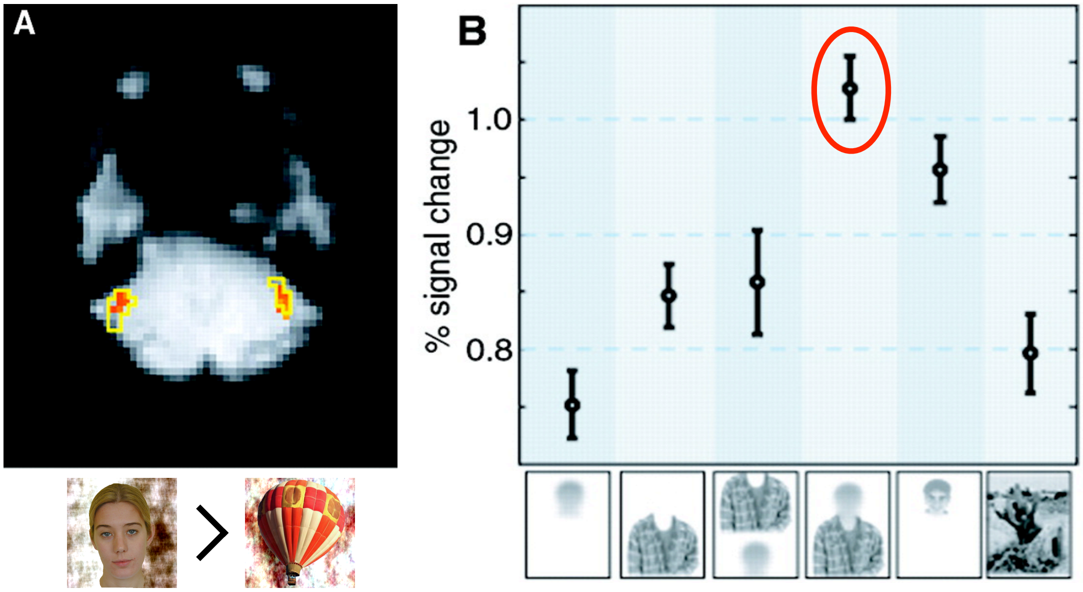


Figure 1.1.1. The Fusiform Face Area is Preferentially Sensitive to the Context of a Face. Reproduced from Cox et al., (2004). Panel a: the fusiform face area of the brain (top) is localized in contrasts of responses to faces and objects. Panel b: changes in activity in this functionally localized region is greatest for the context of a face (highlighted in red).

At the level of social cognition, an example of the power of context influencing processing tendency rather than capacity is demonstrated by comparing judgments to the two panels shown in Figure 1.1.2. While individuals can quickly and reliably make spontaneous trait inferences from photographs of facial features (Willis & Todorov, 2006) as well as judge emotions, (e.g., the woman in panel a is upset), with the addition of contextual cues from that woman's environment (Figure 1.1.2b) that social inference becomes much richer.

Figure 1.1.2. Example of Contextualized Social Emotional Inference. Contrasts of emotion judgments elicited by photographs of the same woman, cropped in to only reveal the face (panel a) and showing the environment surrounding her face (panel b), demonstrate that a deeper level of social emotional inference is permitted by the availability of contextual cues.

An experimental example of contextual cues influencing social cognition is the finding that judgments of politicians based on extra-facial visual cues correspond with real-world election outcomes (Spezio et al., 2012). It had previously been shown that subjects' laboratory two-alternative-forced-choice judgments of who was more competent and trustworthy of two actual political opponents corresponded with election outcomes — for example, the candidate more often chosen as more competent in the laboratory was more often the electoral winner (Ballew & Todorov, 2007). While the literature surrounding spontaneous trait inferences (rapid judgments of people based on photographs of their face) traditionally assumes those judgments are driven by facial features (Todorov & Uleman, 2002), and has manipulated computer-generated faces to demonstrate visual features that contribute to certain trait judgments (Todorov, Baron, & Oosterhof, 2008), we showed that judgments of extra-facial cues alone most strongly correspond with electoral outcomes (Spezio et al., 2012) (Figure 1.1.3).

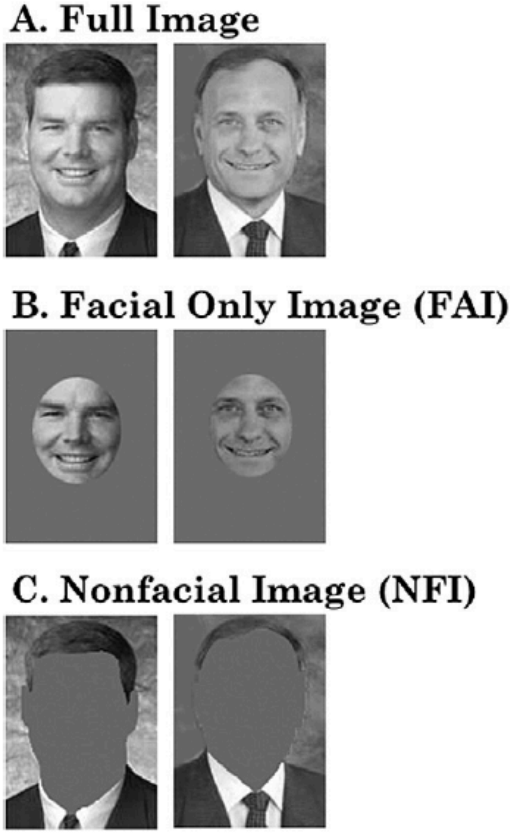


Figure 1.1.3. Sample Stimuli from Study on Judgments of Politicians' Occluded Faces. Reproduced from Spezio, et al., (2012). Subjects made two-alternative forced choice decisions between candidate pairs from real elections, indicating which face they found more trustworthy, more competent, and less threatening. Faces were presented in three conditions: (a) unaltered full facial images, (b) facial-only images that windowed facial features while excluding external cues, and (c) non-facial images that occluded facial information.

While it is unclear which of the available visual cues (dress, hairline, posture, face width, etcetera) influence these judgments, it is clear that some of these contextual, non-facial features meaningfully influence our appearance-based judgments of others.

So far, we have considered examples of considering contextual cues to improve ecological validity. A second theme of ecological exists that requires a fundamentally novel approach to experimentation: interaction.

Social Interaction. In 2007, technology blogger Kathy Sierra described a surprising phenomena: despite several social media tools for remote communication being made available at a conference she attended, participants in that conference preferred travelling and meeting in person over staying home and communicating through those tools (2007) (Figure 1.1.4). She asked why this was the case. In her conceptualization, moving up a curve of interactive richness, social communication tools of increasing representational fidelity (for example, email, to phone, to video-chat), enabled better communication, yet never approached the case of face-to-face interaction where something “ineffable happens [and] turns the brain on.”

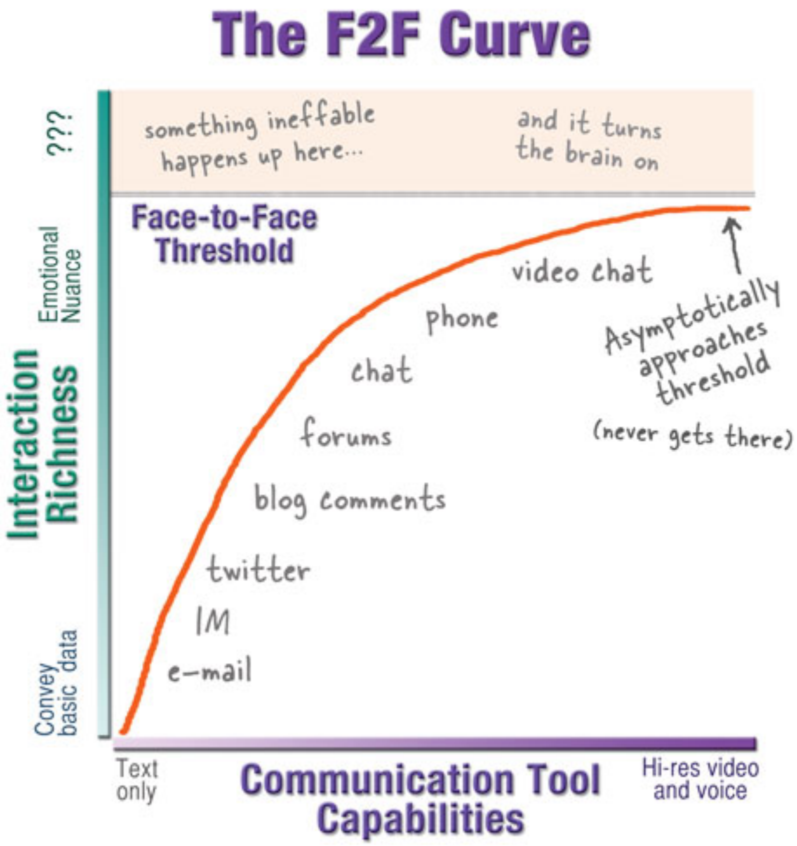


Figure 1.1.4. Face-to-Face Curve. Reproduced from Sierra (2007).

More scientifically, neuroscientists also propose that distinct neural processing occurs in interaction (e.g., (Redcay et al., 2010; Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012; Schilbach, Eickhoff, Cieslik, Kuzmanovic, & Vogeley, 2012; L. Schilbach et al., 2013; Leonhard Schilbach et al., 2013; Tanabe et al., 2012)), in that social cognition is different when interacting with rather than observing someone. Interaction can be viewed as a second-person account of social cognition (Schilbach et al., 2013); in contrast, the two dominant networks described in social neuroscience, the so-called Mirror Neuron System — a macroscopic brain network that can be recorded with fMRI, and which is thought to perform a function analogous to that of mirror neurons — which offers support for a simulationist account, and the inferential mentalizing/Theory of Mind (TOM) network, which supports theory-theory accounts, offer a first and third person account, respectively (Schilbach et al., 2013). It is possible that evidence in favor of either of these prominent networks is strongly influenced by the experimental paradigms used: "it has remained unclear whether and how activity in the[se] networks […] is modulated by the degree to which a person does or does not feel actively involved in the ongoing interaction and whether the networks might subserve complimentary or mutually exclusive roles in this case" (Schilbach et al., 2013). To solve this dilemma, to find the "dark matter" of the neural mechanisms underlying social interaction, we should study the context of interaction as a potentially necessary form of ecological validity. A proposed strong contender for explaining second person accounts of neuroscience is embodied cognition. In contrast to spectatorial accounts, embodied cognition sees social perceptions as "an active process executed by an organism situated in the environment, wherein subjects are not isolated from but embedded in and couple with the perceived world" (Schilbach et al., 2013). Both a hybrid account of the two spectatorial views (theory-theory TOM/mentalizing and simulation theory mirroring accounts), as well as accounts emphasizing embodiment are gaining credence, and second person paradigms (requiring interaction and emotional engagement) may shine light on them.

Believability. In addition to contextual cues and interaction, a third theme of ecological validity (which I do not manipulate in this thesis) is believability — making a participant believe a stimulus or experimental manipulation is ecologically valid and that the experimenter is not deceiving them and that the manipulation is real. Neuroeconomics studies often consider this element of ecological validity, convincing subjects that they are actually interacting with other real people and that they are making “real” decisions, at least one of which will actually be randomly selected and realized.

Ecological Validity in Non-Human Species. Notably, believability is a high-level manipulation that only works in humans. How can ecological validity be manipulated in other species? One approach is that ethological approaches can be scaled-up and computationally assessed (Anderson & Perona, 2014). In our three themes of ecological validity — (1) contextual cues, (2) interaction, and (3) believability — the following are possible:

(1) Making the stimulus as rich, dynamic, and contextualized as possible is feasible in all species. This theme acts on perception.

(2) Making a stimulus interactive, or contingent, on an animal’s actions can be done in other species. This theme acts at many levels, not just perception. In addition to other species, this theme would be important to test in development.

(3) Believability can only be manipulated in humans.

An important gap to bridge between social neuroscience in humans and other animals is that other animals usually do not know that they are in an experiment, and studies with them are therefore ecologically valid in that sense (Stanley & Adolphs, 2013). On the other hand, verbal report and explicit instruction are impossible in animals, and non-ethological testing approaches often require extensive training.

In animal experiments, it is beginning to be recognized that the animal’s rearing and comfort are crucial aspects of ecological validity in social neuroscience. Tetsuro Matsuzawa’s novel participant observation method (Matsuzawa, 2013) in chimpanzee research begins to address this issue. In his approach, group-housed, mother-reared chimpanzees “volunteer” to participate in experiments, coming to testing areas of their living space as they wish. Further, testing and social life are not separated — chimpanzees are raised to interact with and trust the experimenter, with their mother present for interactions with the experimenter. This trusting and fairly natural laboratory rearing has permitted studying other themes of ecological validity, such as interaction, in the finding that chimpanzees are capable of altruistic behavior, helping conspecifics based both on a perceptual understanding of their need and being “asked” for help (Yamamoto, Humle, & Tanaka, 2012). Observing this altruistic behavior required an interactive paradigm since the behavior required being asked for help, an interactive process.

Evolutionary View of Ecological Validity. The following question could be asked: what does ecological validity actual gain? Are differences just attentional? Aside from the theoretical reasons put forth thus far, our consideration of other species segues to answering that question with another question: what did the brain actually evolve to process? While we do not actually know, there are two very different views. The first is that it evolved to process real stimuli, in the real world, in an interactive context (a view espoused by ethologists). A second view, supported by strong proponents of computational models of the mind, is that the brain evolved to process simpler computations that can be brought to bear on real-world processes. At minimum, these real-world processes require incorporating more information or computations to determine behavior. Regardless of which answer one supports, ecological validity is useful. Aside from these theoretical rationales, there is empirical evidence that contextual manipulations do influence behavior.

Further Aspects of Ecological Validity. While three main themes of ecological validity have been discussed, there are of course still other aspects of ecological validity not covered here, including, but not limited to (1) individual variation, (2) the temporal dynamics of everyday versus laboratory behavior were response times are experimentally circumscribed, and (3) the need for more varied and comprehensive ways to measure behavior, including more implicit measures.

Moving Forward

Returning to the concept of representative design, with any factor relevant to ecological validity tested, it is important to be keenly aware of issues of generalizability. Additionally, as with all neuroscience, studies sampling many different methods and providing convergent evidence are preferable. Also, while ethological approaches are valuable and information, as neuroscientists, we need to not only observe behavior, but also develop and test models that explain neural processing thereof.

Moving forward, working within the framework of two main themes of ecological validity, contextual cues, and interaction (with a parallel them of believability), we test the importance of the those themes in the next two chapters by (1) investigating the sensitivity of the amygdala to contextual cues, including ambiguity, in threat perception, and (2) investigating the influence of interaction on the neural processing of gaze in individuals with autism spectrum disorder.

Chapter 2

THE AMYGDALA AND SOCIAL THREAT

The first chapter of this thesis claimed that ecologically valid social cognition is influenced by a constellation of situational contextual cues. While offering support for that claim, this chapter also offers evidence in support of the amygdala serving a general role; this in turn helps explain the amygdala's generalized processing of diverse inputs.

This three sections of this chapter all focus on the amygdala and threat; two of the three sections are published or in press and one is currently under review. The rationale for focusing on the amygdala is twofold: (1) arguably, it is the most investigated brain structure in social neuroscience, and will feature later in this thesis as a neuroimaging region of interest, partially because it is hypothesized to contribute to dysfunctional social cognition in autism spectrum disorder (cf. Chapter 3.1). Within this chapter, the first section provides a brief review of what we know about the function of the amygdala; the second and third sections are empirical papers. A further rationale (2) for studying the amygdala is that I had access to three extremely rare neurological patients who had focal bilateral lesions to the amygdala. Because I investigated approach-avoidance behavior to faces, an important category of social stimuli, prior work that linked face processing to the amygdala, including in these same patients, increased the relevance of focusing on this brain structure.

Summarizing our results, this chapter begins exploring the role that the amygdala plays in social perception, suggesting a general role related to evaluating the saliency or self-relevance of social and non-social cues for an organism.

In the second section, disruption of this evaluation is shown to promote an enhanced default approach bias in bilateral amygdala lesion patients – instantiated as a tendency to default to rating impoverished facial images in which facial features were occluded as more approachable, i.e., more trustworthy and less threatening than their unaltered, whole-face counterparts. This second section marks an important reminder that information *content,* i.e., ambiguity, rather than specific information, can be an important contextual dimension along which our brains evaluate stimuli to guide our social behavior.

This chapter culminates by assessing how ambiguity, alongside nine other contextual factors, hierarchically guides human defensive responses to both physical and psychological threat scenarios, mirroring patterns of threat evaluation observed in other species. In line with the first section of this chapter, this hierarchical guidance of defensive responses relies upon appraisal of the contextual relevance of potential threats to a person.

Chapter 2.1

THE AMYGDALA AND SOCIAL PERCEPTION[3]

The amygdala is a key structure connecting sensory representations with valuation, social inference, attentional modulation, and memory encoding. As such, it plays a prominent role in one particular aspect of social perception: the ability to infer the meaning of social communicative signals. While this role is best studied in regard to the recognition of emotions from facial expressions, it extends to more complex social judgments, other sensory modalities, as well as the incorporation of context. Recent work attempts to synthesize a more unitary function from all these findings, possibly related to aspects of evaluating biological significance or saliency. Considerable future work is required to situate the amygdala's function within a more extended neural system, likely featuring close interactions between the amygdala, temporal and prefrontal cortex, as well as parts of the basal ganglia and brainstem.

Introduction

Bilaterally buried in the medial temporal lobe, the human amygdala is a compact subcortical structure, unilaterally occupying just under 1400 mm^3 in postmortem histology (Schumann & Amaral, 2006) and around 2000 mm^3 in live MRI scans (Schumann et al., 2004). Our understanding of the amygdala's functional role has evolved considerably over the past 50 years, most strikingly shifting from a role in social behavior (in the mid-1900s) to one related to fear (in the late 1900s) to one more

---

[3] A version of this chapter has been published (Harrison & Adolphs, 2015).

abstractly related to value, saliency, and relevance at the current time.

Historically, and at first glance, the role of the amygdala may appear deceptively simple — it detects threat through a specific role in associative memory (LeDoux, 1996). While this function’s simplicity and biological significance constitute an attractive explanation, decades of investigation support a considerably more complex and diverse view (Aggleton, 2000; Whalen & Phelps, 2009), to some extent leaving in question what might be its “basic” function — topics we take up below.

Lesion studies, including the early classic work by Kluver and Bucy (1939), implicated the amygdala in social processing. Subsequent research focused on detection of social and non-social threat. In tandem with increased interest in social neuroscience, social stimuli, especially faces and facial emotions, were studied in detail. Outside of social neuroscience, threat processing studies evolved into a healthy branch of reward learning research, supplemented by neuroeconomics. This accumulation of evidence made it clear that the amygdala does much more than detect threat; in an early synthesis of these diverse areas of research, we (Adolphs, 2010) summarized the emerging view that the amygdala processes “a psychological stimulus dimension related to saliency or relevance.” Clearly, this role would be important to social perception, both at the direct perception end, and in a modulatory attentional role (see Box 2.2.1).

Box 2.2.1. Social Perception

Perception is the transduction of sensory stimuli to action potentials, which elicit early neural processing, resulting in an organism detecting, discriminating, and categorizing an attended environmental stimulus. A perceived object is processed to the extent that it can be discriminated from other stimuli, and categorized (further recognition and judgments that rely on memory, and, in the case of humans, naming, typically follow but need not be included under the rubric "perception").

Social perception is the perception of "social" objects. Notably, "social perception refers to *initial* stages in the processing of information that culminates in the accurate analysis of the dispositions and intentions of other individuals" (Allison, Puce, & McCarthy, 2000). Thus, while emotion recognition and face processing are typically considered perceptual, inferring the beliefs of other people ("mentalizing") is not. The boundary between "initial" and "later" processing is fuzzy and can be debated.

Similarly, the bounds of the amygdala's role in social perception are not clear. Antecedent to perception, it guides attention. Subsequent to perception, the amygdala helps evaluate social stimuli to guide our behavior. While working within a classic three-stage cognitivist model that separates perception, attention, and behavior can be useful methodologically, attempting to isolate the amygdala's role in perception ultimately fails to provide a unitary understanding of the amygdala's function. The amygdala likely contributes to rapid detection, more extended processing, and behavior - all at different points in time. In fact, it participates *all* the time, likely through interactions of its various sub-nuclei with different networks, and we should view its perceptual role as restricted to a particular slice in processing time within a specific network. Its broader role is much more pervasive.

How tenable is the hypothesis that the amygdala processes salience or relevance? Attractively, it predicts a rather broad processing role, which is borne out by the amygdala's anatomy. The amygdala features dense afferent and efferent connections to other cortical and subcortical regions, and features interconnections between its constituent subnuclei (Amaral, Price, Pitkänen, & Carmichael, 1992; Pitkänen, Savander, & LeDoux, 1997). Interconnections within the amygdala highlight the complexity packed into this small structure. The human amygdala is a collection of over a dozen nuclei and sub-nuclei (Figure 2.1.1). Broadly speaking, it consists of two main regions: the basolateral amygdala and centromedial amygdala. Ultimately, an understanding of the role of the amygdala in social perception cannot ignore distinctions between its subnuclei.

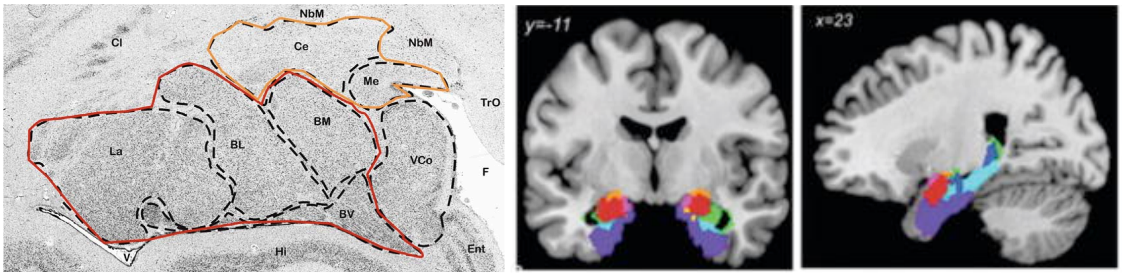


Figure 2.1.1. Amygdala nuclei. A coronal section of a postmortum human brain (left) shows the centromedial nucleus outlined in orange and basolateral complex outlined in red. An MRI scan (right) shows coronal and parasaggital probabilistic locations of amygdala nuclei in humans. Lefthand panel: BL = basolateral nucleus; BM = basomedial nucleus; BV = basoventral nucleus; CE = central nucleus; La = lateral nucleus; Me = medial nucleus; VCo = ventral cortical nucleus; CL = Claustrum; Ent = entorhinal cortex; F = endorhinal sulcus; Hi = hippocampus; NbM = Nucleus basalis of Meynert; TrO = Tractus opticus; V = lateral ventricle. Righthand panels: Orange = centromedial amygdala; red = basolateral amygdala; magenta = superficial amygdala; yellow = hippocampal-amygdaloid transition area; light blue = subiculum; green = CA regions of the hippocampus; dark blue = dentate gyrus; purple = entorhinal cortex. Reproduced from Amunts, et al. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region, and entorhinal cortex: intersubject variability and probability maps. *Anatomy and Embryology*, 210(5-6), 343-352, with permission from Spring Science + Business Media.

Here we readdress the question of what the amygdala contributes to social perception. We do so by (1) reviewing the three main areas of amygdala research - fear, social judgment, and reward processing; (2) conducting a meta-analysis of human fMRI studies; and (3) considering anatomy and connectivity.

The Amygdala and Fear

Behaviorally, some of the most striking consequences of focal amygdala lesions in nonhuman animals are a lack of fear-like behavior in response to normally threatening stimuli. This has been shown particularly clearly in the case of rats (Choi & Kim, 2010), where reversible pharmacological lesions revealed a parametric relationship between amygdala activity and fearful behavior: muscimol-induced lesions were associated with a lack of fearful behavior, whereas bicuculline-induced excitation was associated with exaggerated fearful behavior.

Recently, optogenetics has been used to study fear behavior and conditioning in rats. Optogenetics uses molecular biology tools to express synthetic ion channels in particular neurons of an animal's brain. These ion channels are opened or closed by light, allowing light pulses to modulate neural activity in real-time in freely moving animals. Optical stimulation of amygdala neurons can directly serve as a conditioned stimulus in fear conditioning (Johansen et al., 2010). Meanwhile, optical silencing of basolateral neurons that project to the centromedial nucleus affects anxiety behavior (Tye et al., 2011). Optogenetics is a powerful technique, with exciting potential to elucidate functional interactions of neural populations.

A complementary approach to neuromodulation consists of irreversible chemical lesions via ibotenic acid, an approach taken in several monkey studies (Emery et al., 2001; Mason, Capitanio, Machado, Mendoza, & Amaral, 2006; Prather et al., 2001). Here, too, there is a consensus that amygdala lesions reduce the normally cautionary break on behavior that is elicited by stimuli that signal potential threat. Interestingly, the behavioral effect (which consists essentially of an unmasking of exploratory or approach behaviors that are normally inhibited) is seen both for overtly threatening stimuli (e.g., snakes) as well as for stimuli that merely signal uncertainty (e.g., novel objects next to food). The consequences on behavior for complex social stimuli are rather varied, and depend to some extent on context and individual differences (Mason et al., 2006).

In humans, only a single complete lesion case has been well studied (although there are a handful of studies on other patients with similar lesions): SM, a patient with developmental bilateral amygdala lesions, also showed a dramatic lack of fear behaviors (Feinstein, Adolphs, Damasio, & Tranel, 2011), as well as impairment in directing gaze and attention to relevant facial features (Figure 2.1.2). Especially valuable in this human case was the opportunity to (a) test fear elicitation across a wide range of stimuli (autobiographical recall, actual snakes and spiders, a haunted house, fear experience in everyday life, horror movies) and (b) investigate other basic emotions as well. A highly selective absence of fear was found across all the stimuli. Interestingly, while the amygdala seems necessary to elicit fear from external stimuli, following lesioning of the amygdala, interoceptive fear can be induced by carbon dioxide inhalation (Feinstein et al., 2013).

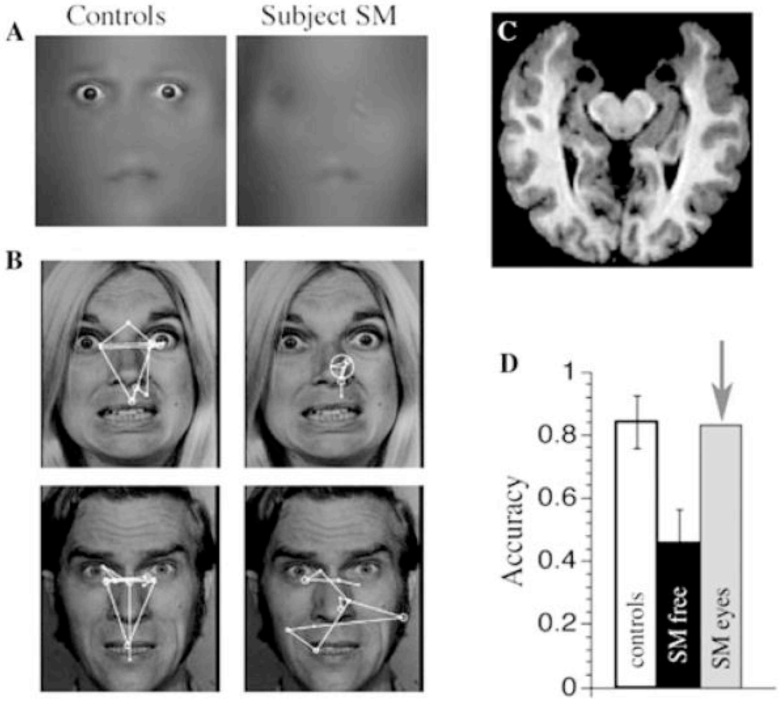


Figure 2.1.2. The brain and face processing of patient SM. Bilateral amygdala lesions impair the use of the eyes and gaze to the eyes during emotion judgment. (A) A patient with bilateral damage to the amygdala made significantly less use of information from the eye region of faces when judging emotion. (B) While looking at whole faces, the patient (right column of images) exhibited abnormal face gaze, making far fewer fixations to the eyes than did controls (left column of images). This was observed across emotions (free viewing, emotion judgment, gender discrimination). (C) MRI scan of the patient's brain, whose lesion was relatively restricted to the entire amygdala, a very rare lesion in humans. The two round black regions near the top middle of the image are the lesioned amygdalae. (D) When the subject was instructed to look at the eyes ("SM eyes") in a whole face, she could do this, resulting in a remarkable recovery in ability to recognize the facial expression of fear. The findings show that an apparent role for the amygdala in processing fearful facial expressions is in fact more abstract, and involves

the detection and attentional direction onto features that are socially informative. Reproduced from Adolphs, (2010). What does the amygdala contribute to social cognition? In: A. Kingstone & Miler (Eds.), *Year in Cognitive Neuroscience 2010* (pp. 42-61). Malden: Wiley, with permission from Wiley.

A few other such amygdala lesion patients have also been studied, and bear out these main findings (J. Feinstein, personal communication). A point to note is that they all suffer from a developmental and progressive disease, leaving open questions about when in development lesions first manifest as well as whether any tissue or passing fibers might be spared. Some of these patients appear to have damage mostly restricted to the basolateral nucleus, and also exhibit behavioral variability (Terburg et al., 2012). Systematic future investigations of these patients will be critical to begin to address the difficult question of the causal roles of particular amygdala subnuclei in humans.

These findings, and in particular the human case (Feinstein et al., 2011), raise two broad questions for how to conceive of the amygdala's role in social perception. First and foremost: at what point in processing (see Box 2.2.1) does the amygdala come into play? Does it impair perception as such (detection, discrimination)? Or does it impair the ability to associate meaning with what is perceived, or to modulate cognition based on the associated value of what is perceived (e.g., recognition, social judgment)? Across all of the studies, human and nonhuman, the answer here has been relatively clear: amygdala lesions impair the associative ability, but not basic perception as such. Yet even this conclusion is now being modified in light of the amygdala's role in attentional processing. The second question is how selective the role is to the domain of fear. Is the amygdala's role in more complex social judgment reducible to a role in recognizing fear?

Or, alternatively, is its role in fear reducible to a more computationally abstract or basic function? We take up these two related questions next.

The Amygdala in Social Judgment

Monkeys with amygdala lesions show complex deficits in their interactions with other monkeys that are not obviously reducible to an absence of fear (Emery et al., 2001). In general, the deficits are surprisingly subtle, although they are invariably associated with negative social consequences (loss of social status, abandonment by the troop and death in the wild) (Kling & Brothers, 1992). In fact, neonatal amygdala lesions in infant monkeys result in an exaggerated social fear behavior, something not seen following adult lesions (Prather et al., 2001).

The deficits in social behavior in humans with amygdala lesions are also very subtle, although this is perhaps not too surprising given the complex and interactive nature of the social environment and, specifically in humans, the compensatory aspects provided by others who know that a patient has an amygdala lesion (Adolphs, 2010). More experimental control can be obtained in studies that assess social judgments through ratings given to depictions of scenes, faces, or descriptions of scenarios. In these studies, while there is also considerable variability, the deficit has been largely consistent with a lack of fear: patients with bilateral amygdala lesions tend to judge other people as more trustworthy and more approachable and have difficulty recognizing fear (Adolphs, Tranel, & Damasio, 1998; Adolphs et al., 1999).

This conclusion in humans is to some extent borne out by functional neuroimaging studies. For instance, face stimuli designed to parametrically vary in terms of their perceived untrustworthiness elicit a corresponding parametric activation of the amygdala in healthy viewers (Winston, Strange, O'Doherty, & Dolan, 2002). Yet there are notable discrepancies here. In particular, the human amygdala is also potently activated by appetitive stimuli (e.g., erotic pictures (Hamann, Ely, Hoffman, & Kilts, 2002)); in fact, it is variably activated by all facial expressions across the board (Fitzgerald, Angstadt, Jelsone, Nathan, & Phan, 2006), although it may respond in particular to certain underlying dimensions such as valence (Todorov et al., 2008).

Beyond imaging studies focusing on social stimuli, a unifying attempt comes from the reward learning literature. There is an extensive literature from lesion and electrophysiological studies in animals that argues for a role in reward learning and attention. That literature historically emphasized fear or arousal, in the form of Pavlovian fear conditioning as well as instrumental avoidance learning (LeDoux, 1996) and modulation of declarative memory (McGaugh, 2004), but has now been extended to encompass both appetitive and aversive forms of learning, and both Pavlovian and instrumental mechanisms (Murray, 2007). Here, attempts to find a simple role have been frustrated by the sheer range of findings. Results from neuroimaging studies have been similarly diverse.

Given the very large number of human neuroimaging studies of the amygdala, it is possible to attempt to extract some possibly basic underlying dimensions from stimuli and tasks, across several sub-domains of research, which might most consistently activate the amygdala.

Social Perception Meta-Analysis

To synthesize the neuroimaging research, we turned to Neurosynth (http://neurosynth.org), an automated meta-analysis approach that converges with manual results (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). At a basic level, Neurosynth produces forward and reverse inference maps that relate a certain psychological construct to a brain region. Specifically, *forward inference* maps represent the probability of observing activation in a brain region (within 6mm of a coordinate), given the presence of a particular term, e.g., “social”, in a research article at a certain frequency (default is once every thousand words); *reverse inference* maps represent the probability of a term occurring in an article given activation in a particular brain region. Probabilities are calculated based on a database of nearly 6000 studies.

Because Neurosynth uses a large database to create these reverse and forward inference term-to-activation mappings, its developers argue that their tool renders the reverse inference problem tractable. The reverse inference problem occurs when one incorrectly assigns a functional role to a region, as is easily done when only looking at results from a few hand-selected studies**.** Much more data, which is provided by Neurosynth, is necessary to make the probabilistic statements necessary to make functional inferences: a region should be both consistently (forward inference) and selectively (reverse inference) associated with a functional term.

Confirming the premise of this chapter that the amygdala is involved in social perception, a “social perception” reverse inference map included activation in the amygdala as well as other regions associated with social perception (Figure 2.1.3). Interestingly, this map exhibited some right lateralization, both in the amygdala, and across hemispheres. In the whole brain, 60% of voxels in the reverse inference map (Figure 2.1.3) are in the right hemisphere. Strong lateralization was observed in the amygdalae: 60% of voxels in the right amygdala had activation exceeding FDR 0.10

correction, compared to only 3% of voxels in the left amygdala.

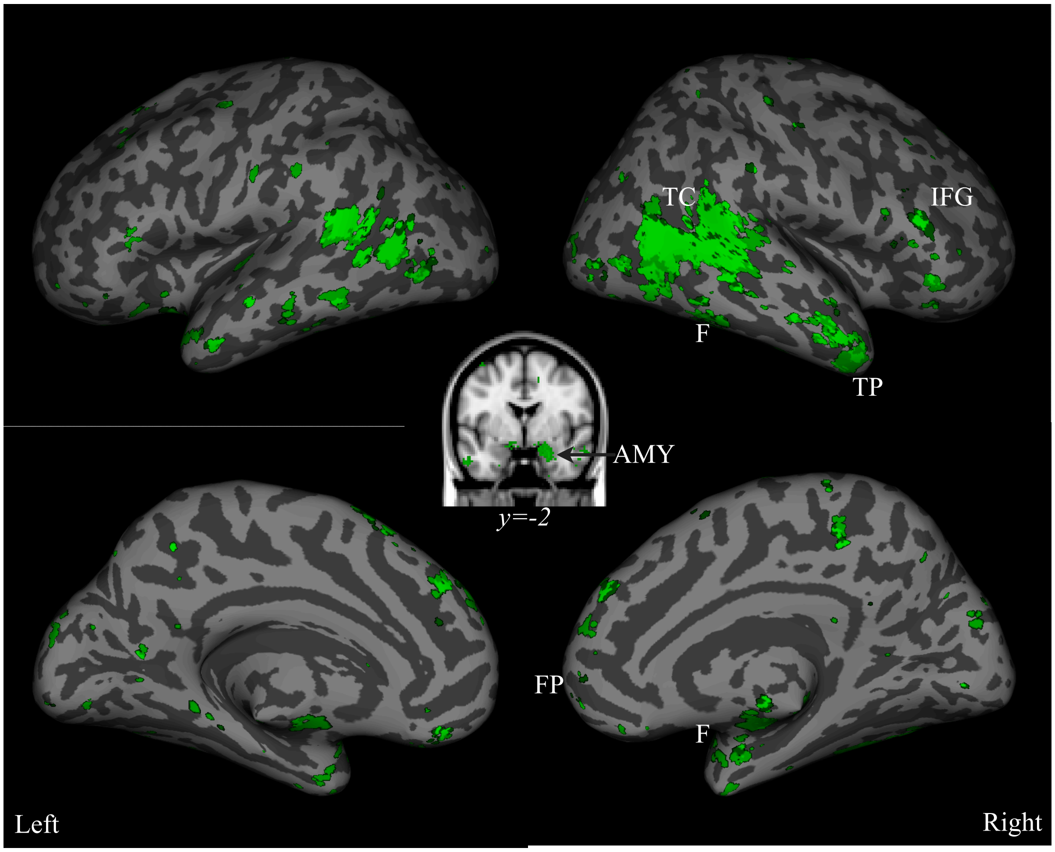
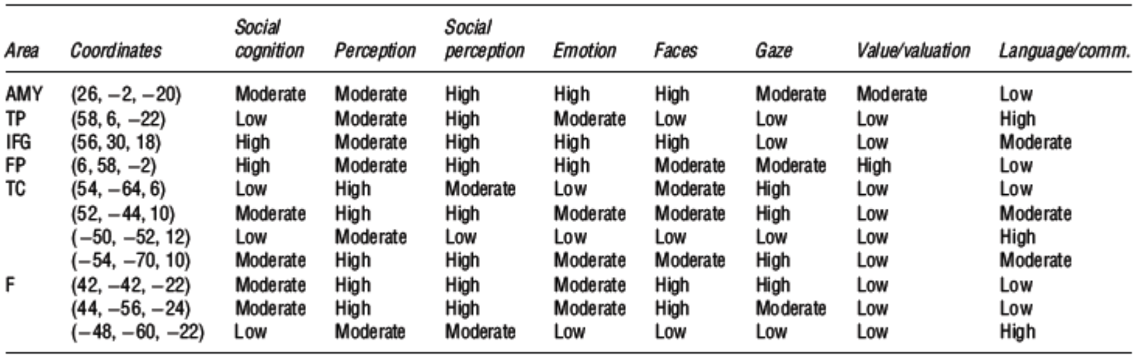


Figure 2.1.3. Social perception reverse inference map from fMRI studies. Combination of FDR 0.1 (dark green) and FDR 0.05 (bright green) corrected z-score maps of the probability, given observed activation at a particular voxel, that a study in the Neurosynth database is related to social perception. These reverse inference maps were created by searching the database of nearly 6000 papers for those in which the terms "social" and "perception" both occurred at least once in every 1000 words. 128 papers met this criteria. AMY = amygdala; TP = temporal pole; FP = frontal pole; F = fusiform gyrus; IFG = inferior frontal gyrus; TC = temporal cortex. Surface rendering and visualization for this and following images were completed using the SPM surfrend toolbox (http://spmsurfrend.sourceforge.net) and Neurolens (neurolens.org).

While the amygdala's involvement in social perception that emerges from this data-driven analysis is unsurprising, how does its role differ from that of other "social perception" areas identified? We tested the specificity of the amygdala's role in social perception by calculating association scores between 8 functional keywords and 11 regions identified in the social perception reverse inference map (Table 2.1.1). While Neurosynth produced a long list of terms associated with activation in each of these regions, we wanted to summarize those lists to determine whether they were strongly associated with a small number of social perception concepts. These social perception "keywords" onto which to map the neurosynth list of terms were obtained by presenting neuroscientists (n=6) with the top 50 terms associated with each social perception region by Neurosynth, and asking them to generate words to summarize each list of terms. From these responses, 8 keywords were distilled and appear as the column headers in Table 2.1.1. A second group of neuroscientists (n=5) rated (0= "No"; 1 = "Maybe"; to 2 = "Yes") the relationship between each word associated with a region and these 8 keywords. Neurosynth lists of words were truncated to only consider words receiving a z-score>3; this resulted in a total of 118 unique terms being rated across all the brain regions. These ratings were converted into association scores, reported in Table 2.1.1.

Table 2.1.1. Association score between each of 11 social perception regions (Figure 2.1.3) and 8 social perception keywords obtained from fMRI studies. An association score between each keyword and region was derived by creating a region x term association score for each subject, averaging across subjects, and thresholding High/Moderate/Low ratings by a third split of all mean scores. Each subject's region x term association score was calculated as a weighted combination of ratings, normalized

by the number of terms considered. Ratings of 3 (“Yes”) were weighted by 2, ratings of 2 (“Maybe”) were weighted by 0.5, and ratings of 1 (“No”) were weighted by -1. AMY = amygdala; TP = temporal pole; IFG = inferior frontal gyrus; FP = frontal pole; TC = temporal cortex; F = fusiform gyrus.



Area	*Coordinates*	*Social cognition*	*Perception*	*Social perception*	*Emotion*	*Faces*	*Gaze*	*Value/valuation*	*Language/comm.*
AMY	(26, −2, −20)	Moderate	Moderate	High	High	High	Moderate	Moderate	Low
TP	(58, 6, −22)	Low	Moderate	High	Moderate	Low	Low	Low	High
IFG	(56, 30, 18)	High	Moderate	High	High	High	Low	Low	Moderate
FP	(6, 58, −2)	High	Moderate	High	High	Moderate	Moderate	High	Low
TC	(54, −64, 6)	Low	High	Moderate	Low	Moderate	High	Low	Low
	(52, −44, 10)	Moderate	High	High	Moderate	Moderate	High	Low	Moderate
	(−50, −52, 12)	Low	Moderate	Low	Low	Low	Low	Low	High
	(−54, −70, 10)	Moderate	High	High	Moderate	Moderate	High	Low	Moderate
F	(42, −42, −22)	Moderate	High	High	Moderate	High	High	Low	Low
	(44, −56, −24)	Moderate	High	High	Moderate	High	Moderate	Low	Low
	(−48, −60, −22)	Low	Moderate	Moderate	Low	Low	Low	Low	High

First, this keyword association analysis (Table 2.1.1) confirms that the amygdala is indeed involved in social cognition and perception, and has an especially strong association for domain-specific “social perception”, compared to domain general “perception”. This domain specificity may be skewed by the database of human neuroimaging articles used by Neurosynth. Beyond confirming the amygdala’s involvement in social perception, we demonstrate it is moderately to highly associated with all the other social perception functional concepts, with the sole exception of the concept language/communication. Appropriately, this last concept was strongly associated with the temporal cortex. As predicted by the large literature on the amygdala and faces, emotion, and gaze, the region was highly associated with those three concepts. However, these associations were not very specific, as several other regions were also associated with faces, emotion, and gaze. Meanwhile, the amygdala’s moderate association with the concept of value/valuation was more specific: the amygdala and the frontal pole were the only regions in the Neurosynth social perception network associated with valuation. To a degree, these findings agree with the idea that the amygdala, anatomically intermediary between perceptual and cognitive regions, plays an intermediary role between perception and cognition by assigning value to (social and non-social) perceptual objects.

Moving Forward: Anatomical Considerations

As a whole, the above reviews present a puzzle for the amygdala's role in social perception. To a degree, the amygdala's role in this domain can be explained in terms of its role in basic reward and attentional processes. However, open questions remain, and a key challenge is resolving several potential explanations. The findings from our Neurosynth analysis begin to provide a comprehensive answer and highlight a promising direction.

One solution to the above predicament is to acknowledge that the question "what does the amygdala contribute to social perception" was rather ill-posed to begin with, since it leaves out the rest of the social perception network within which the amygdala is embedded, and whose functions we need to understand if we are to thus situate the amygdala's.

One such network approach was taken in a study (Bickart, Hollenbeck, Barrett, & Dickerson, 2012) that seeded three regions of the amygdala in a resting-state functional connectivity analyses that elicited three main networks correlated with activity in those regions: a perception network associated with the ventrolateral amygdala, an affiliation network with the medial amygdala, and an aversion network with the dorsal amygdala. These networks mirrored known anatomical connectivity from monkey and rodent literature; the large temporal component of the perception network was evocative of Freese and Amaral's (2005) well characterized dense temporal connections (Figure 2.1.4).

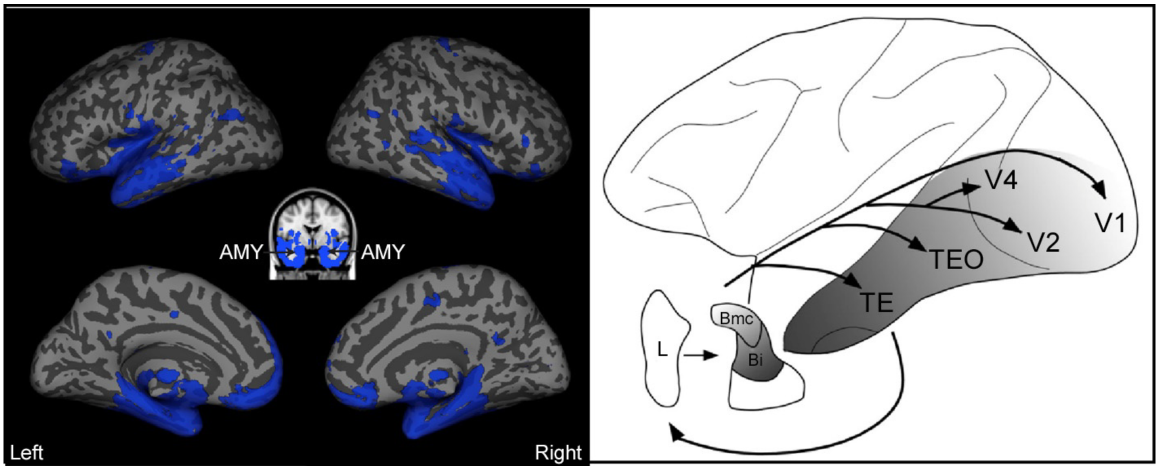


Figure 2.1.4. Amygdala connectivity. Amygdala resting-state functional connectivity perceptual network map (left) from Figure 5 in (Bickart et al., 2012). Map kindly provided by Brad Dickerson. AMY= amygdala. This connectivity with temporal regions is reflected in a summary of known anatomical connections between the amygdala and ventral visual stream areas in the monkey (right). Projections are topographically organized, with ventral and rostral amygdaloid regions projecting most to rostral visual areas, and dorsal and caudal amygdaloid regions projecting most to caudal visual areas. A feedforward/feedback loop exists between area TE and the basal and lateral nuclei of the amygdala. Right-hand panel reproduced from Freese & Amaral (2005). The organization of projections from the amygdala to visual cortical areas TE and V1 in the macaque monkey. *Journal of Comparative Neurology, 486*(4), 295-317, with permission from Wiley-Liss, Inc.

Additionally, the results from Bickart and colleagues (2012; Figure 2.1.4) partially converged with our meta-analysis results (Figure 2.1.3): their social perception network, derived from a region similar to the basolateral nucleus, and our social perception regions partially overlapped, with 27% of the voxels in our Neurosynth social perception map (combination of FDR 0.1 and FDR 0.05) (Figure 2.1.3) also present in their map (Figure 4). While the different amygdala nuclei interact, these converging results suggest that the core social perception function might be assigned mostly to the basolateral amygdala, which is also the region of the amygdala that is lesioned most commonly in Urbach-Wiethe disease.

These converging anatomical observations are indicative of one of the most promising means of advancing our understanding of the amygdala's role in social perception: a network approach. The amygdala's role will not be understood in isolation. We should strive for a biologically-constrained role that seeks to understand the function of the amygdala as a group of structures that interact with one another and within networks. The field of optogenetics has much potential for advancing our network-level understanding of amygdala function. However, findings must be related across species (optogenetics research currently being limited mostly to mice) and subfields: here, it is crucial to conduct meta-analyses within species, and, across species, to carefully consider how sub-field-defined cognitive functions are related.

Moving forward, especially in the social domain, we must consider environmental context. Our social environment is rich and complex, and social cognition and perception are largely inferential. They rely upon correct incorporation of contextual cues. This context relates to stimulus value, which we already know is partially assigned by the amygdala, differs among individuals, and is a product of stimulus history. Approaching our investigation within this framework is necessary for fleshing out a nuanced understanding of the amygdala's network role in modulating perception and assessment of our complex social environment.

Chapter 2.2

PEOPLE WITH AMYGDALA LESIONS SHOW A BIAS TO APPROACH FACES EVEN WHEN THEY ARE OCCLUDED[4]

Approach and avoidance constitute a basic dimension of all animal behavior. A large literature documents approach and avoidance elicited by specific sensory stimuli, yet comparatively little is known about default approach biases when stimulus information is reduced. The amygdala is well known to contribute to approach and avoidance behaviors in response to specific sensory stimuli, and here we test whether the amygdala's role might extend to situations where stimulus information is reduced. A novel task asked three rare patients with bilateral amygdala lesions to make approach-related judgments about photos of faces when intact, and with all internal facial features occluded. Direct comparisons of these stimuli isolated a stimulus-independent bias. The patients showed a greater tendency than controls to default to rating occluded faces as more approachable than whole faces. These findings suggest that the amygdala's role in approach behavior extends beyond responses to specific stimuli.

Introduction

From single-celled organisms to humans, all mobile species exhibit approach-avoidance behavior. In humans, approach-avoidance behavior is regulated by motivation and influenced by emotion (Elliot, Eder, & Harmon-Jones, 2013); at a more primitive

---

[4] A version of this chapter is accepted for publication as Harrison, L, Hurleman R, & Adophs, R (In press). *Psychological Science.*

level, it is related to instinctive defensive behaviors (Blanchard, Griebel, Pobbe, & Blanchard, 2011; McNaughton & Corr, 2004).

While basic, approach-avoidance behavior shows large individual differences. Whereas some people would walk into and explore an unfamiliar dark room, others would pause and gather more information, and some might even flee. What accounts for this behavioral variability? Prior experience might sway one's response, but the example situation offers little information, and may not have been encountered previously. Nevertheless, a behavioral tendency will be observed. The amygdala is a brain structure known for its role in memory, learning, and emotion, and implicated in psychiatric disorders including anxiety. We investigate what role the amygdala might play in regulating stimulus-independent behavior, termed a "default bias." A default approach bias may be normal in certain contexts; here, we ask whether amygdala lesion patients exhibit an abnormally large approach tendency to low-information stimuli.

An abnormal tendency to approach others and normally threatening stimuli in amygdala lesioned monkeys (Klüver & Bucy, 1939), rodents (Choi & Kim, 2010) , and humans (Feinstein et al., 2011; Kennedy et al., 2009) points to the amygdala as being important in regulating approach-avoidance behavior. However, the basis of this approach tendency is unclear.

On the one hand, the bias may be specifically tuned for certain stimuli: much of what we know about the amygdala's contribution to cognition and behavior has come from studies investigating faces. Single-unit amygdala response selectivity has been found for faces in humans (Rutishauser et al., 2011) and monkeys (Gothard, Battaglia, Erickson, Spitler, & Amaral, 2007) and fits with the known connectivity of the amygdala with anterior temporal neocortex (Amaral et al., 1992), a region containing face-selective cells (Perrett, Rolls, & Caan, 1982; Tsao, Freiwald, Tootell, & Livingstone, 2006). Lesions of the human amygdala can result in a remarkably specific impairment in recognizing fear (Adolphs et al., 1999; Adolphs, Tranel, Damasio, & Damasio, 1994; Broks et al., 1998) and trustworthiness from faces (Adolphs et al., 1998).

On the other hand, the amygdala might contribute to stimulus-independent baseline or default biases, similar to the "tonic influence on behavior" theorized over two decades ago (Amaral et al., 1992). In amygdala lesion patients, some preliminary evidence for a general approach bias includes a propensity to approach other people or potentially dangerous situations, regardless of context (Feinstein et al., 2011; Kennedy et al., 2009).

The amygdala's roles in a face-specific or default approach bias are not mutually exclusive. Trustworthiness and approachability judgments to whole faces in SM, a widely-tested amygdala lesion patient, already hint at two distinct processes: the first is a deviation from normal judgments that increases as faces become less trustworthy (i.e., an inability to process facial cues to untrustworthiness); the second is a more uniform positive bias across all faces irrespective of their perceived trustworthiness (i.e., a global bias towards trust; Adolphs et al., 1998; cf. Figure 2.2.8). In the present study, we attempt to disentangle these two components, while also making two improvements: using more than a single case-study and controlling for regression-to-the-mean.

One mechanism by which a default bias could be achieved is through sensitivity to ambiguity. Between detecting a dangerous stimulus and launching a defensive response, the amygdala, in conjunction with the cortex, must contextually assess potential danger (Davis & Whalen, 2001). When signals are ambiguous, the amygdala may increase vigilance and predictive information by lowering sensory thresholds (Whalen, 2007). Sensitivity to ambiguity has been implicated across species; in mice and humans, temporal unpredictability in stimulus presentation elicits anxious behavior and amygdala activity (Herry et al., 2007). At a sufficiently long stimulus duration to permit appraisal (van der Zwaag, Da Costa, Zürcher, Adams, & Hadjikhani, 2012), ambiguous fearful faces with direct gaze elicit greater fMRI amygdala activation than unambiguous averted-gaze fearful faces (Adams, Gordon, Baird, Ambady, & Kleck, 2003). Individual variation in state anxiety correlates with amygdala BOLD response to potentially ambiguous neutral faces (Somerville, Kim, Johnstone, Alexander, & Whalen, 2004). To our knowledge, we are the first to explore whether human amygdala lesions produce an abnormal approach tendency as a function of ambiguity.

We tested the hypothesis that amygdala damage produces a stimulus-independent default bias by directly contrasting approach-related judgments for two sets of otherwise identical face stimuli, with one set modified such that the inner part of the face is erased (cf. Figure 2.2.2a). Participants made trust and threat judgments, which are known to be processed relatively automatically from faces (Willis & Todorov, 2006). In the real world, distance (Sinha, Balas, Ostrovsky, & Russell, 2006), accessories (scarves, sunglasses), and objects naturally occlude facial features. In both our whole and occluded stimulus conditions, the same external-facial cues (hairline, shape) were available, but facial feature information was only available in the whole condition. We operationalized a default occluded-approach bias as a tendency, within-subject, to judge the low-information occluded face as more trustworthy and less threatening than the corresponding whole face.

Materials and Methods

Participants

Amygdala Patients. Selective bilateral damage to the human amygdala is extremely rare, but can arise from the genetic disease, Urbach-Wiethe disease (Hofer, 1973). Three women with bilateral amygdala calcification lesions (Figure 2.2.1) caused by Urbach-Wiethe disease were tested. Two patients, AM and BG, are identical twin sisters from rural southern Germany. They were 36 years of age at testing, are married with children, and have been in full-time employment since they received their 13 years of education in Germany. The third patient, AP, is American, was 27 years of age at testing, and has worked since she obtained her Bachelor's degree. All three patients have an IQ in the average range (HAWIE-R: AM, 101; BG, 96; WASI: AP, 98), (Becker et al., 2012). Their lesions are all similarly symmetric and confined to the amygdala (AM, 1.12 cubic centimeter bilaterally; BG, 1.15 cc; AP, 0.71 cc). The damage includes complete ablation of the basolateral amygdala with minor damage of other amygdaloid regions, including anterior and ventral cortical regions at the rostral level and lateral and medial parts of the central nucleus and amygdalo-hippocampal area at the caudal level (Figure 2.2.1). All amygdala patients were tested individually in the laboratory.

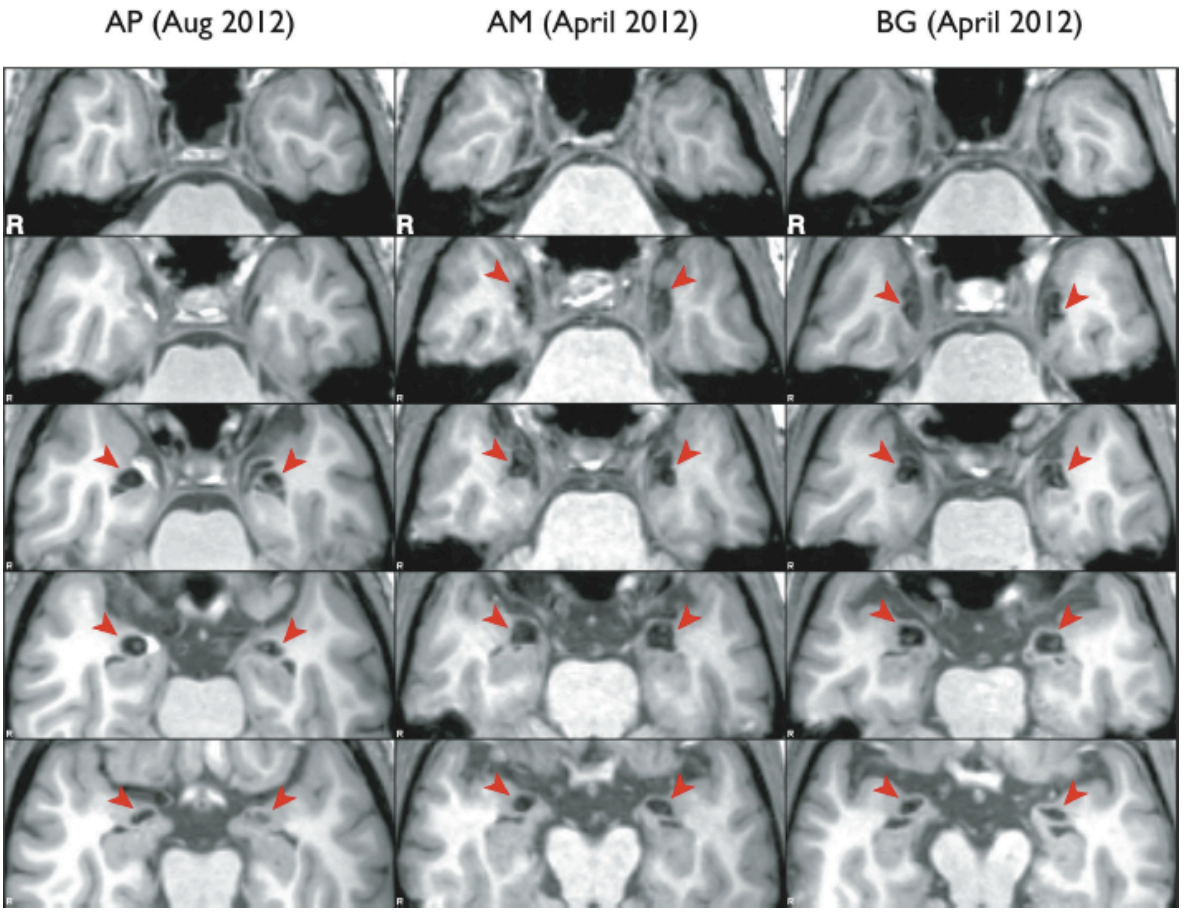


Figure 2.2.1. Anatomical MRI scans of the patients' amygdala lesions. Displayed are 1mm isotropic T1-weighted magnetic resonance imaging coronal sections of the patients' anterior medial temporal lobes. Red arrows highlight AP, AM, and BG's focal bilateral amygdala calcification damage. Images obtained at 3T in the Caltech Brain Imaging Center at the specified timepoints. R: right.

Healthy Comparison Subjects. 81 age, gender, and education-matched controls with no current mental health diagnosis were tested. Of the control group, 61 participants were American (mean age: 30.5 ± 8.0 years) and 20 were German (mean age: 35.1 ± 6.1 years). Americans were recruited through Amazon's Mechanical Turk and Germans were recruited through emails forwarded to acquaintances of the authors' German colleagues.

All control participants completed the experiment using Qualtrics, an online survey-hosting platform, under conditions that were otherwise identical to the in-lab tests completed by the lesion patients. All participants, including lesion patients and controls, gave informed consent/assent in accordance with a protocol approved by the Caltech IRB.

Sample Size. For our normal controls, we tested a sample size (*N*=81) exceeding ones that have been investigated in approach/withdrawal in the past (*N*=46, e.g., Adolphs et al., 1998). With respect to the patients with amygdala lesions, these have in the past typically been reported as single case studies; here we report three. We present their results both individually and as a small group, and use bootstrap analyses to compare them statistically to the control group.

Stimuli

The stimulus set consisted of 34 high-resolution color images (16 females and 18 males between 20 to 50 years of age) showing essentially neutral facial expressions. Images were taken with the same camera, at the same angle with controlled lighting, and in front of the same plain background. After image capture, images were luminance matched on each RBG channel, using the SHINE toolbox (Becker et al., 2012). The 34 original images (“Whole”) were used to create a second set of images, which were oval-masked so that there were no inner facial features (“Occluded”), (Figure 2.2.2). All 68 images were resized such that the inter-ocular distance for all images was constant. All faces were unfamiliar to the participants.

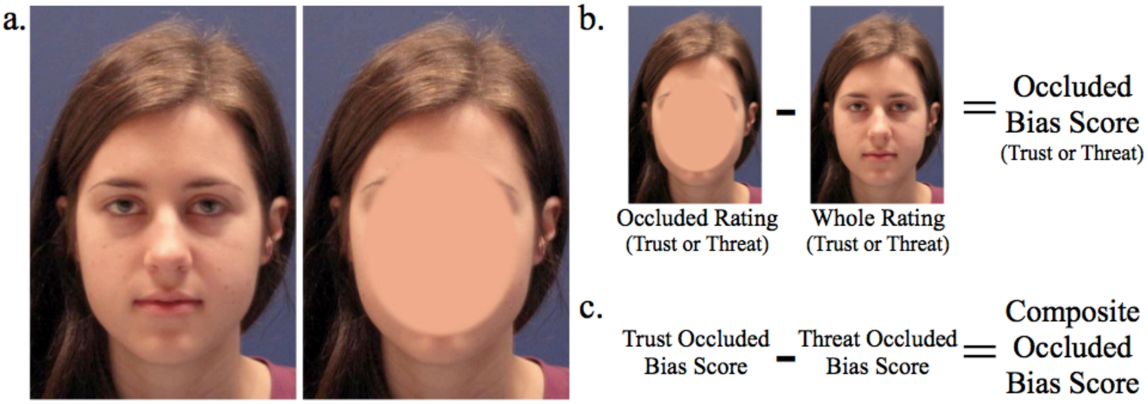


Figure 2.2.2. Stimuli and construction of bias scores. Panel a. Sample whole (left) and occluded (right) facial images; occluded face stimuli contained less information than whole face stimuli. Panel b. Occluded bias scores were calculated by subtracting the whole from occluded rating for each face. Positive occluded bias trust scores indicate a tendency to find the occluded face more trustworthy; negative occluded bias threat scores indicated a face was less threatening in the occluded condition. Panel c. Composite occluded bias scores combined trust and threat occluded bias scores, subtracting threat bias scores from trust bias scores so that the two measures had the same directionality: faces with larger positive composite bias scores were more approachable (more trustworthy and less threatening) in the occluded condition.

Experimental Design

For each face, participants indicated whether or not they found the person threatening or trustworthy. Participants rated all images in a self-paced manner on a six-point scale (“Strong No”, “No”, “Weak No”, “Weak Yes”, “Yes”, “Strong Yes”). The directionality of the rating scale was counterbalanced across participants. All occluded face stimuli were presented first. Preferably presentation order would be counterbalanced or randomized; here, counterbalancing was not possible with a patient sample of 3 and, without a much larger stimulus set, it would have been difficult to control for memory effects with randomization. While our design therefore includes a presentation order confound, the same confound is present across participant groups and therefore does not affect our main question of interest, that of whether within-subject shifts in ratings between stimulus conditions differ between participant groups. Within the constraint of presenting occluded stimuli first, the trust and threat judgments were presented in a randomized order. Each judgment (threat or trust) comprised a block, in which all 34 images (occluded or whole) were presented, also in randomized order. For the German participants, the entire experiment was translated into German: 'threatening' was translated as 'bedrohlich' (synonyms: menacing, ominous); 'trustworthy' was translated as 'zuverlässig' (synonyms: reliable, trustworthy — note that this translation was probably more ambiguous than that used for threat). Translations were independently verified by five bilingual German/English speakers.

Analysis

Rescaling German Participants' Scores. The rescaling discussed in this section does not affect the approach biases discussed below, which constitute our main effect of interest. Those biases are comprised of within-subject rating shifts, which are not affected by group-level manipulations applied uniformly to individuals' scores. This rescaling only affects separate reports of ratings for each stimulus condition (2: whole, occluded) by judgment (2: trust, threat), cf., Table 2 and Figure 2.2.3.

To control for cultural and language differences between the two nationalities, we minimally rescaled all German ratings with a fixed bias offset. The average American trust rating across all faces (whole and occluded) and all control participants was subtracted from the same average rating for German control participants; the value of this difference between the two groups was subtracted from every single German rating, including those of the two German lesion patients. German threat ratings were rescaled following the same procedure. Rescaling was small, and likely simply reflects differences in language (N.b. trustworthiness, for which the German translation was likely more ambiguous than that for threat, had a larger absolute rescaling factor). German trust ratings were rescaled by 0.429 and threat ratings by -0.188. Since there were no mean differences between American and German controls after rescaling (Figure 2.2.3), we pooled these control groups in all subsequent analyses.

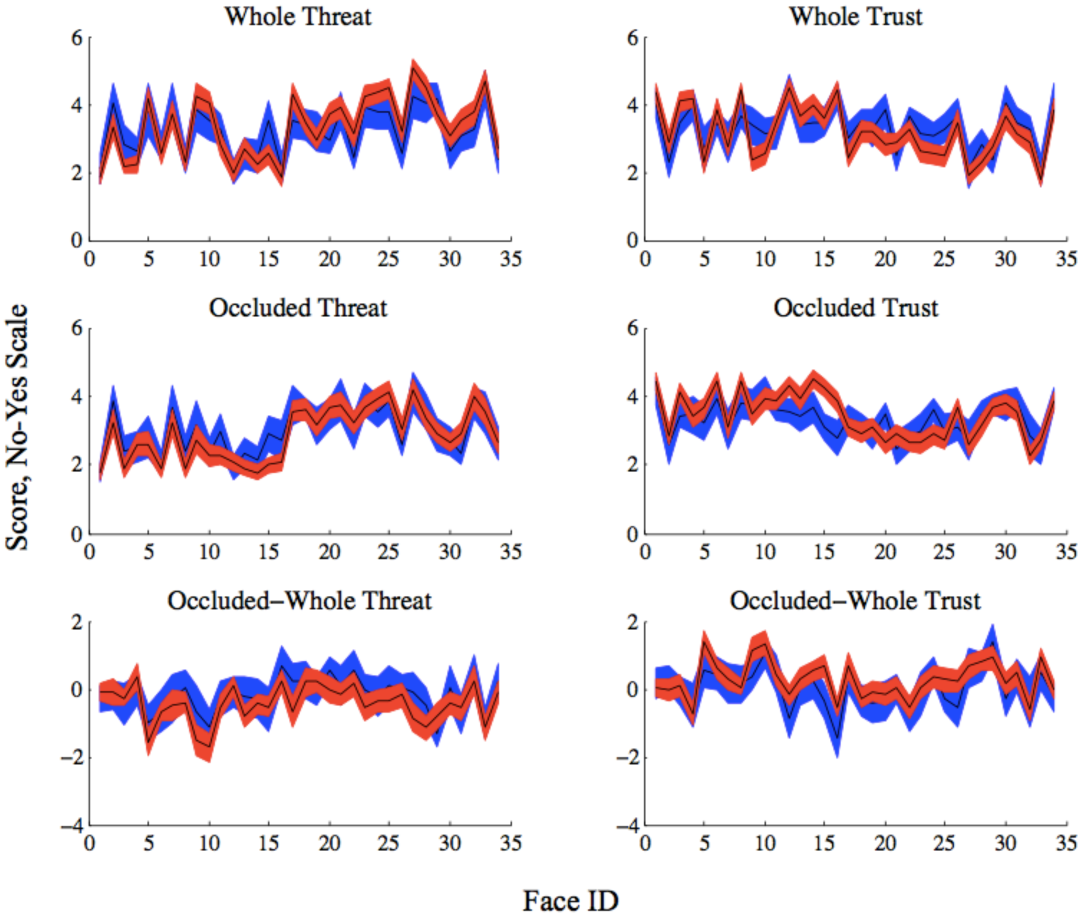


Figure 2.2.3. German and American Face Ratings. 95% CI interval of mean rescaled German (red) and American (blue) threat (left) and trust (right) ratings for whole (top) and occluded (middle) faces, and for occluded minus whole scores.

Comparing Patients to Controls: General Approach. Comparisons of the three lesion patients to controls was driven by asking how the three lesion patients would compare to three people randomly drawn from the general population. This comparison was estimated by building a bootstrapped population estimate from 100,000 bootstrap samples of the average rating of three randomly sampled controls, and calculating the proportion of this bootstrap distribution that lay in the tail of the distribution that exceeded the three patients' average rating.

Defining Occluded-Approach Bias. We defined a bias score (Figure 2.2.2b), for each face, as the occluded minus the whole face rating given to that face. Positive shifts in trust judgments and negative shifts in threat judgments represented an occluded-approach bias.

Determining Consistency of Approach Bias Across Judgments. To determine the consistency of the directionality of an approach bias across the two judgments, plots of mean control trust and threat bias scores for each face were overlaid (cf. Figure 2.2.4a), demonstrating that trust and threat occluded biases moved together in a consistent fashion. Each face's threat bias scores were subtracted from its trust bias scores to form composite occluded-bias scores (Figure 2.2.2c) for each participant. Mean negative "occluded-avoided" composite bias scores indicated faces that were avoided in the occluded condition, receiving higher threat and lower trust ratings than in the whole condition. Mean positive "occluded-approached" composite bias scores indicated faces that received lower threat and higher trust ratings in the occluded condition.

Testing Difference in Occluded-Approach Bias in Patients. Each individual patient's occluded bias scores were qualitatively compared to those of controls by overlaying their individual bias score for each face on a 95% confidence interval plot of the mean control bias score for each face, with higher scores indicating an approach bias in each patient.

The group of patients' occluded bias scores were quantitatively compared to a bootstrapped distribution of controls' ratings as described above for (1) all faces, (2) for faces with occluded-avoided (negative occluded bias) scores and (3) for faces with occluded-approached (positive occluded bias) scores. To remove any statistical dependency between the face-classification and comparison of control and patient scores, faces were reclassified as occluded-avoided or occluded-approached on each bootstrap iteration according to the mean bias score of the 78 controls who were not randomly sampled on that bootstrap iteration.

Testing difference in "positivity" bias in occluded and whole faces. To disentangle the contributions of occluded and whole face ratings to an approach bias, we separately defined a "positivity" approach bias (cf. (Norris, Gollan, Berntson, & Cacioppo, 2010)) in whole and occluded faces, as the trust minus threat rating for each whole or occluded face. More positive scores indicated faces that were approached with a "positivity" bias (more trustworthy than threatening) and more negative scores indicated faces that were avoided (more threatening than trustworthy). Bootstrap resampling compared patients' and controls' positivity bias for whole and for occluded faces.

Results

Overview

Results are fully detailed below, and have been summarized in Table 2.2.1 to quickly orient readers to our findings.

Table 2.2.1. Results Summary. Main findings listed alongside relevant Results source.

Finding	Source
A composite occluded bias score is derived by combining separate trust and threat occluded bias scores, which went in opposite directions: faces with an occluded-approach bias for trust ratings tended to have an occluded-avoid bias for threat ratings and vice versa.	Figure 2.2.4a
Patients' composite occluded minus whole bias scores tend to be more positive, representing an occluded-approach bias, than controls' scores for each face.	Figure 2.2.4b-d
Bootstrap analysis comparing 3 randomly sampled controls' mean composite occluded bias score to the patients' actual mean score shows a	Figure 2.2.5

stronger occluded-approach tendency in the patients for all faces, especially for faces controls tended to avoid in the occluded condition.	
Parametric visualization of the bias shown in Figure 3. To test whether patients' bias results from greater difficulty in rating occluded faces, we generated synthetic patient bias scores derived from actual whole-face ratings and random occluded-face ratings, which did not explain their bias.	Figure 2.2.6
Bootstrap comparison of controls' and patients' general "positivity" (trust minus threat) bias for whole and occluded faces separately shows (1) that all participants had a bias to approach occluded faces more than whole faces, but (2) also confirms this bias was greatly enhanced in the patients.	Figure 2.2.7
Full reporting of patients' individual scores and 95% CI of mean patient scores for each judgment X stimulus type.	Table 2.2.2

Lesion Patients' Composite Bias Scores Relative to Controls

We sought to demonstrate a general approach bias in amygdala lesion patients by exploring whether they tended to approach occluded stimuli more than whole face stimuli, relative to controls.

Before comparing the patients' and controls' occluded bias, we first confirmed the validity of a composite bias score. Controls' trust and threat bias scores moved together in an expected manner (Figure 2.2.4a), such that faces that tended to be rated as more trustworthy (blue dots) when occluded were also rated as less threatening (red dots) when occluded, and vice-versa. Therefore, for each subject, the threat bias score for each face was subtracted from its trust bias score, combining biases in the two judgments into a composite bias score (black line), with negative scores denoting faces avoided in the occluded condition and positive scores denoting faces approached in the occluded condition.

The patients' bias scores (teal dots) for each face were compared to the pooled controls' mean bias scores (black line) (Figure 2.2.4b-c): overall, their bias scores indicated a heightened tendency to approach faces in the occluded condition (this bias was not alternatively driven by a heightened tendency to avoid faces in the whole condition, see below and Table 2). Each patient's mean bias score was higher than those of controls (*M* = 0.59, SD = 0.76): AP's mean score was 0.46 SD higher than controls', AM's 0.85 SD, and BG's 1.28 SD.

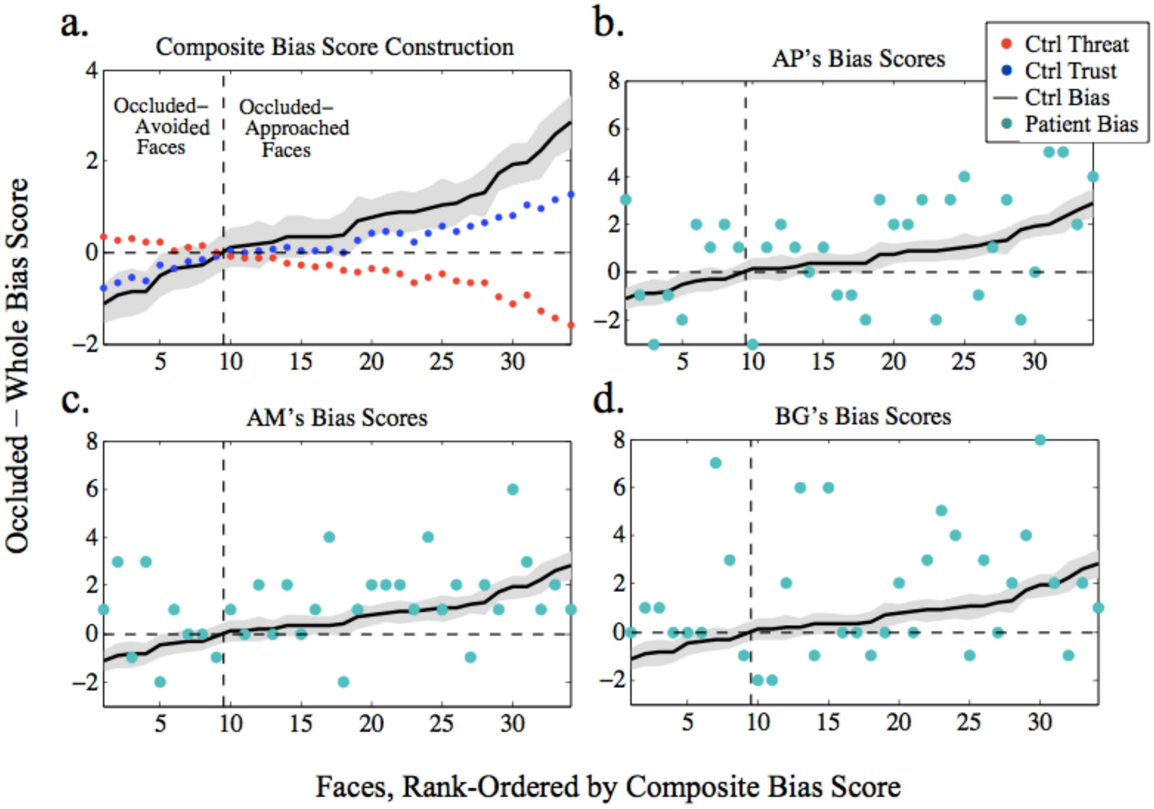


Figure 2.2.4. Composite bias score defined and compared to patients' scores. Panel a. Overlaid plots of controls (*N*=81) mean occluded minus whole difference score (y-axis) for each face for trust (blue) and for threat (red) show that faces that were not trusted in the occluded condition also tended to be found threatening relative to the whole-face condition. This consistent bias contributed to the formation of a composite bias score (black; 95% CI in gray) defined by adding the trust and reversed threat bias scores for each face. Face

stimuli (x-axis) rank-ordered according to composite bias score. Faces with a negative bias score tended to be avoided in the occluded relative to whole condition (or, conversely, approached more in the whole face condition), while faces with a positive bias score tended to be approached more in the occluded face condition. Panels b-d. Controls' mean composite bias scores (black; 95% CI in gray), with individual patients' scores overlaid in teal, demonstrate a trend for patient bias scores to exceed controls' scores. Ctrl=controls.

Further confirming the finding that patients have a general approach bias to rate occluded faces as more trustworthy and less threatening than whole faces, our bootstrap analysis showed that the patients' composite bias scores (patient average shown as teal line in Figure 2.2.5) were indeed higher than controls' scores (distribution shown by black line in Figure 2.2.5), with 7.6% of control bootsamples exceeding the mean patient bias score. In addition to this bootstrap analysis, the control and patient group averages were strongly distinct (d =1.13), with nonoverlapping 95% CIs [0.43, 0.76] and [0.90, 1.59], respectively. This difference from controls was weaker for those faces that controls also had a bias to approach in the occluded condition (bottom panel of Figure 2.2.5) (15.8% of control bootsamples exceeded the patient mean, and group averages were only moderately distinct (d = 0.74), with 95% CIs for controls [0.84, 1.20] and patients [1.19, 1.74]). As expected, this difference was, however, most pronounced for those faces that controls had a tendency to avoid in the occluded condition (middle panel of Figure 2.2.5 - 1.9% of control bootsamples exceeded the patient mean, and group averages were strongly distinct (d = 1.53), with 95% CIs for controls [-0.81, -0.37] and patients [0.04, 1.22]).

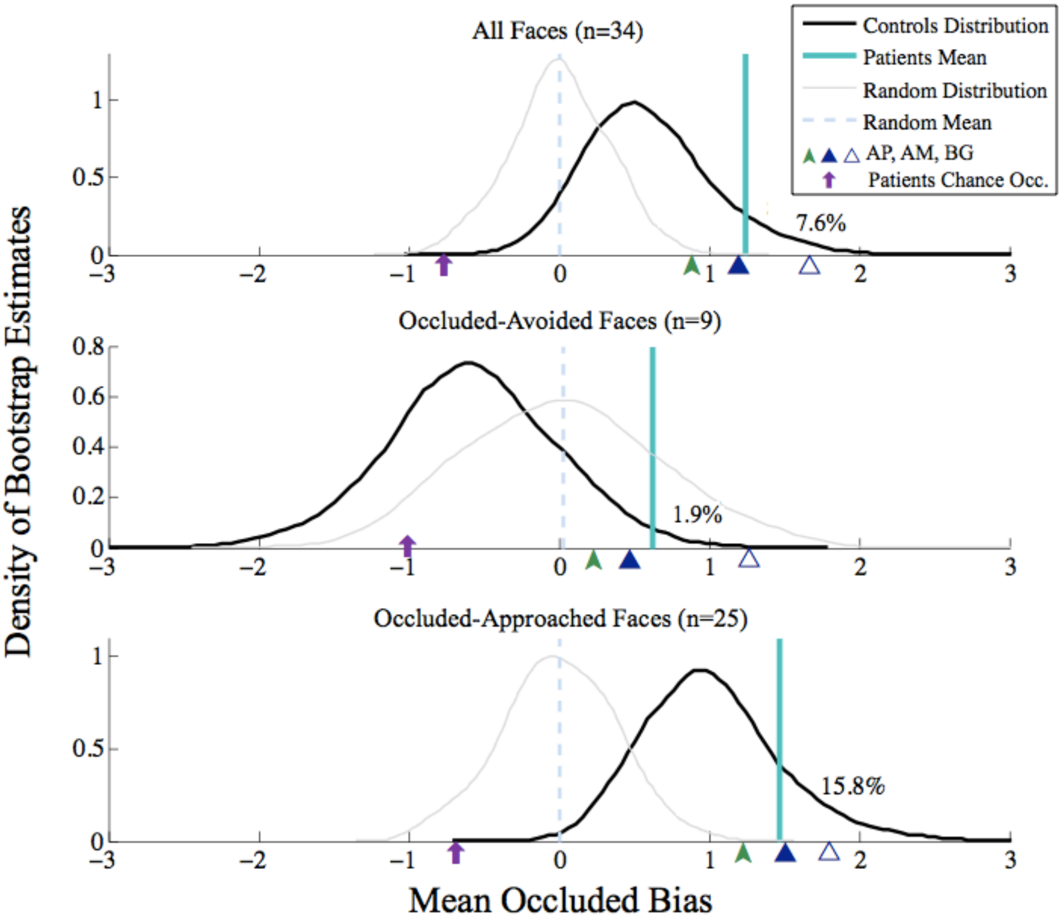


Figure 2.2.5. Mean patient occluded bias scores compared to bootstrapped mean control bias scores. Density plot estimates (black) of the bootstrap distribution of three randomly-sampled controls' mean occluded-bias scores for all faces (top), "occluded-avoided" faces as classified by control ratings (middle), and "occluded approached" faces (bottom), with the three patients' actual mean bias score overlaid in teal and the individual patients' mean scores indicated by green and blue arrows on the x-axis. To remove statistical dependency between our classification of "occluded-avoided" and "occluded-approached" faces and our comparison of 3 randomly sampled controls to the patients, for each bootstrap iteration, occluded-avoided and occluded-approached faces were classified according to the 78

remaining controls. Percent of bootstrap control estimates with a higher occluded bias rating is indicated on each plot, demonstrating that patients' occluded-approach bias scores tended to be higher than controls' scores. To test against these reported effects being driven by regression to the mean, bootstrapped estimates of 3 patients with random bias scores (built from raw trust and threat scores evenly distributed from 1 to 6) are overlaid in gray with their mean value overlaid in a pale blue dashed line. Demonstrating that the patients' deviation from controls was driven by their occluded ratings, an average synthetic patient bias rating given their actual whole face scores and chance occluded scores is indicated on each plot with a purple arrow.

To test against the possibility that these effects were driven by regression to the mean (that is, that the amygdala lesion patients simply produced noisier, more random, ratings of the faces), bootstrapped estimates of 3 simulated patients with random bias scores (built from raw trust and threat scores evenly distributed from 1 to 6) were overlaid (grey lines) on the actual controls' bootstrapped distributions (black lines) in Figure 2.2.5. Adding noise to the control data strengthened the separation of the controls' and patients ratings for all faces (top panel of Figure 2.2.5), and occluded-avoided faces alone (bottom panel), as the random distribution (grey line) moved below the actual distribution (black line) for controls, increasing the separation between controls and the patients' mean (teal line). While this increased separation was not observed for the occluded-avoid faces, the bias in the patients measured by our bootstrap analysis was strong with only 1.9% of randomly sampled control mean bias scores exceeding the actual patients' mean bias score; in tandem with the effect of noise on the other groupings of faces, the possibility of regression-to-the-mean is unlikely as an explanation for the patients' impairment.

To support the conclusion that this effect was driven by a specific bias in rating occluded faces relative to whole faces, we next derived synthetic difference ratings by subtracting actual whole face ratings from chance occluded-face patient ratings (i.e., 3.5, with a fixed bias offset for the German patients): these (indicated by purple arrows in Figure 2.2.5) tended to be lower than controls' bias scores, once again going in a direction opposite to that seen for the actual ratings given by the amygdala lesion patients.

Binary classification of faces as occluded-avoided (middle panel in Figure 2.2.5) or occluded-approached (bottom panel) indicated that the patients' occluded approach bias was strongest for faces avoided by controls in the occluded condition. We visualized how the patients differed from controls for each face by overlaying a 95% CI plot of controls' rank-ordered mean bias scores (black line with grey area in Figure 2.2.6) with a plot of the patients' mean score for each face (red dots), smoothed with a 10-face moving average to improve visualization of the general parametric trend. While the patients tended to have a higher bias scores than controls across all faces, the patients' bias scores became less distinct from controls' as a function of the controls' occluded-bias scores. To again support the hypothesis that this effect was driven by a specific deficit in rating occluded faces, we also plotted synthetic ratings derived by subtracting actual whole-face patient ratings from chance occluded-face ratings (blue dots in Figure 2.2.6). These synthetic ratings did not positively deviate from controls' ratings as did the patients' difference ratings — in fact, they were lower — indicating that the patients' deviation from controls' scores were indeed driven by abnormal, but consistent, occluded-face ratings, i.e., the bias we report.

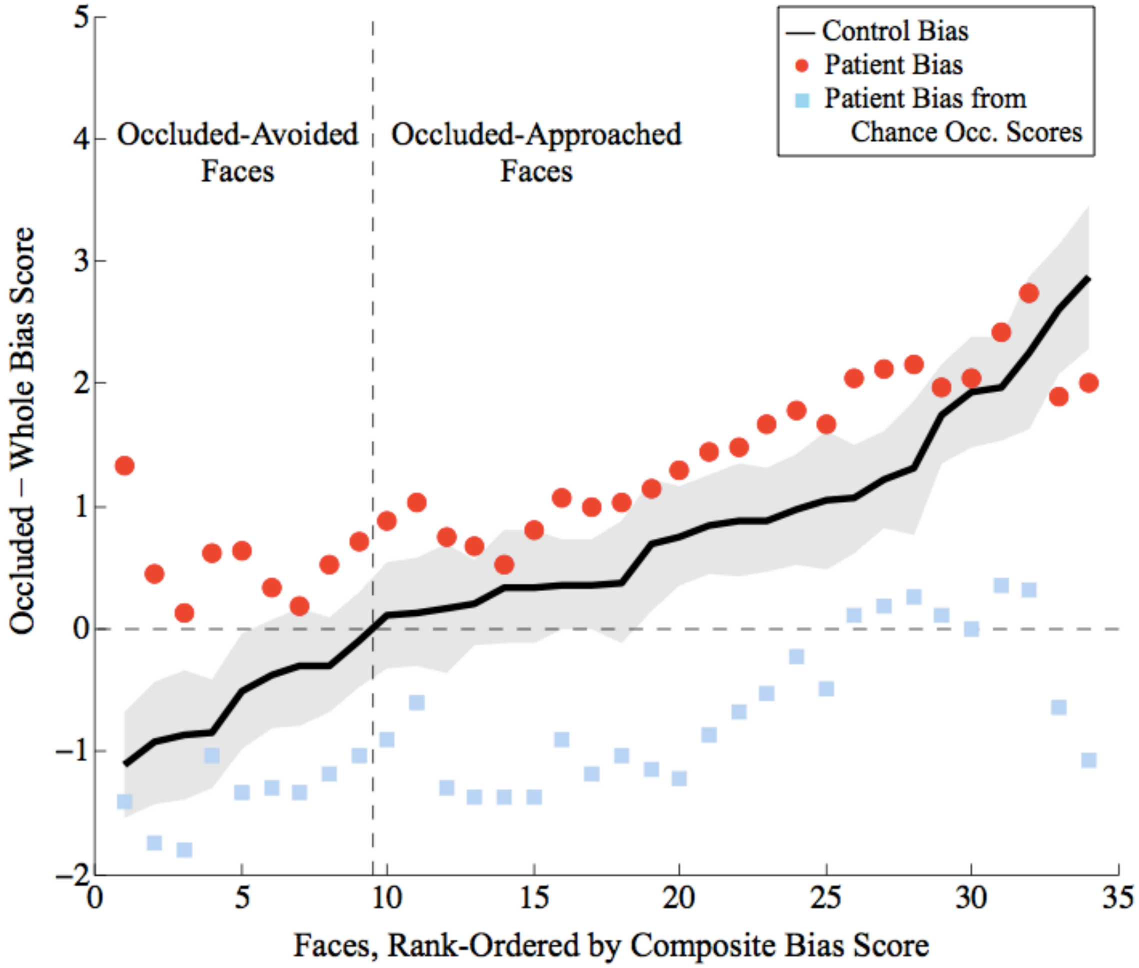


Figure 2.2.6. Parametric bias visualization. Overlaid plots of controls' (*N*=81) mean composite bias score (y-axis, black; 95% CI in gray) and patients' (*N*=3) mean composite bias score (red circles; smoothed with a 10 face moving average, with a step size of one face). The patients all tended to have a higher approach bias than controls, meaning they tended to approach occluded faces more than controls. Face stimuli (x-axis) rank-ordered according to mean control composite bias score. Synthetic patient ratings (light-blue squares, also smoothed with a 10 face moving average), indicate patient bias scores given actual whole-face ratings and chance occluded-face ratings; these scores test the hypothesis that the deviation of patients' bias scores from controls' scores was driven by abnormal

occluded-face ratings; since chance occluded ratings did not exceed controls' bias scores, the effect was driven by the patients' tendency to approach occluded faces more than whole faces.

Trust-Threat "Positivity" Bias in Whole and Occluded Stimuli Alone

Separate examination of an approach-related "positivity" bias (positive for faces with stronger trust than threat ratings) for whole and occluded faces separately revealed a default positive approach bias to occluded but not whole faces in both controls and patients (Figure 2.2.7). While a shift between whole positivity ratings (*M*=0.063) and occluded positivity ratings (*M*=0.656) was present in controls (p=5.90x10^{-10}), this shift in positivity bias for occluded faces was markedly greater in the patients (*M* =0.373 for whole vs. *M* =1.618 for occluded faces, *p*=0.0199).

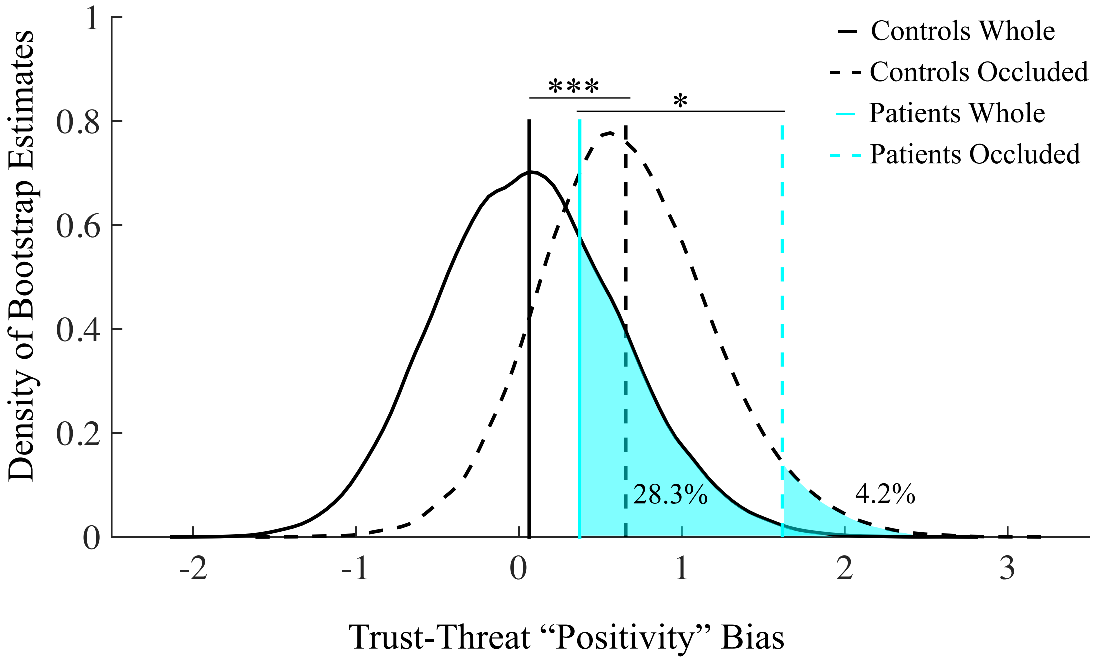


Figure 2.2.7. Bootstrap comparison of patients and controls approach-related "positivity" biases. Density plot estimates of three randomly sampled controls' mean positivity bias

(black lines), and the actual patients' mean positivity bias (blue lines) for whole (solid lines) and occluded (dashed lines) faces. Patients and controls both had a greater approach-related positivity bias for occluded than whole faces, but the magnitude of the shift was greater for the patients, who were also most distinct from controls for the occluded faces specifically. *$p<0.05$; ***$p<0.001$.

Results Separated by Judgment and Stimulus Type

Across comparisons, an occluded approach bias was defined as a tendency to give both higher trust ("Yes," trustworthy) and lower threat ("No," not threatening) ratings to occluded faces than to whole faces. While a composite approach bias was the most comprehensive way to describe our results, separate trust and threat biases obviously contribute to this construct. We thus also compared the lesion patients' mean trust and threat occluded biases separately to bootstrapped control distributions, alongside patient performance for whole and occluded faces alone (Table 2.2.2), confirming that the patients' observed approach bias was driven by abnormal ratings of occluded faces.

Table 2.2.2. Bootstrapped control samples exceeding mean and individual patient values.

Judgment	Stimuli	Controls Exceeding Patients (%)	Control Mean (95% CI)	AP	AM	BG	Summary
Threat	Whole faces	19.9	[3.2, 3.4]	3.3	2.7	3.2	AP & BG same, but AM lower
	Occluded faces	4.7	[2.8, 3.1]	2.6	2.2	2.1	All lower
	Occluded bias	12.1	[-0.5, -0.3]	-0.7	-0.4	-1.1	AP and BG lower; AM not different, but already impaired for whole faces
Trust	Whole faces	23.7	[3.1, 3.3]	3.1	2.6	4.6	AP normal; AM low; BG high
	Occluded faces	3.9	[3.3, 3.5]	3.3	3.4	5.0	AP and AM normal; BG very high
	Occluded bias	8.9	[0.1, 0.3]	0.2	0.8	0.4	AP normal; AM and BG high
Composite Bias	All faces	7.6	[0.4, 0.8]	0.9	1.2	1.6	All higher
	Occluded-avoided faces	1.9	[-0.8, -0.4]	0.2	0.4	1.2	All higher
	Occluded-approached faces	15.8	[0.8, 1.2]	1.2	1.5	1.7	AP normal; AM and BG higher

Note. Occluded bias defined as occluded minus whole face rating. Composite bias defined as occluded trust bias minus occluded threat bias. Occluded-avoided faces have a negative average composite bias score; occluded-approached faces have a positive average composite bias score. For threat, the percentage of lower bias scores were counted; for trust and composite bias, the percentage of higher scores were counted.

As a group, the patients' whole-face ratings were not different from controls' ratings, which was somewhat unexpected given a previous finding in amygdala lesion patient SM (Adolphs et al., 1998), which showed that SM gave abnormally high trustworthiness judgments to whole faces (albeit a different set of faces than the ones used in our present study). To verify that this discrepancy from SM's behavior in our amygdala lesion patients was not driven by differences in the stimulus sets, we tested two of our patients (AM and BG) also on the original stimuli from that experiment with SM (Adolphs et al., 1998). While AM and BG tended to deviate from controls, especially for the faces controls avoided most (Figure 2.2.8), they only exhibited a weak trend towards the prior finding in SM. Unfortunately, SM was not available for testing in our new task, and AP was not available for testing on the original task used by Adolphs et al. (1998).

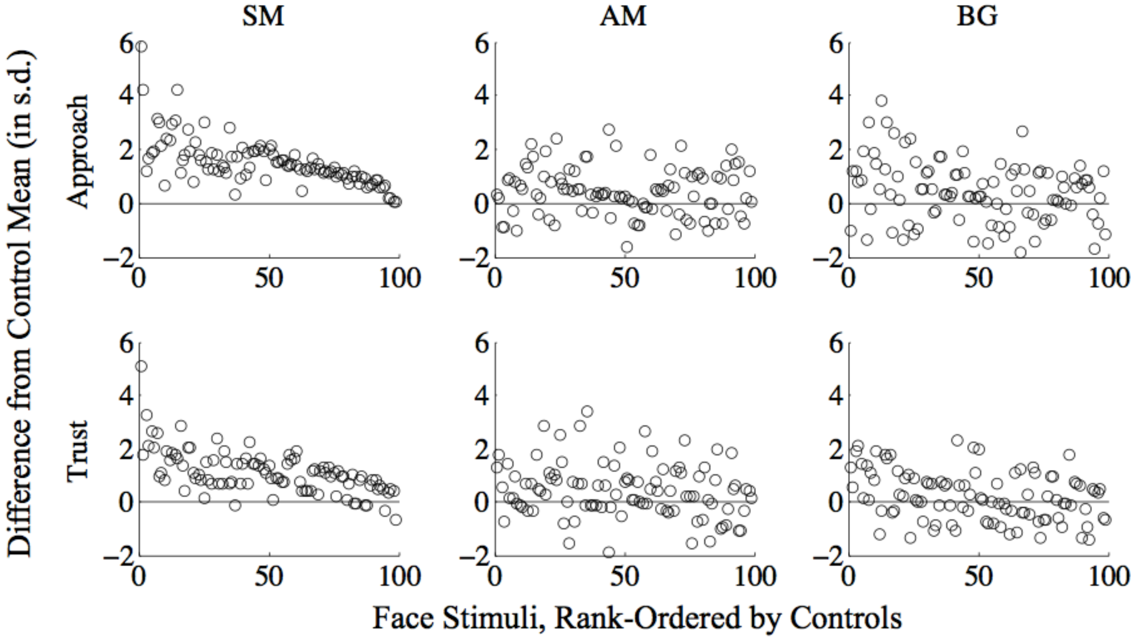


Figure 2.2.8. Amygdala lesion patients' deviations from normal judgments of approachability (top) and trustworthiness (bottom) of 100 faces (circles; y-axis). Units are standard deviations of the normal control ratings. Stimuli are rank-ordered on the x-axis according to the ratings normal controls gave them. SM demonstrated a specific exaggerated impairment in her ratings of the faces rated least approachable or trustworthy by controls, but also displayed a global positive deviation from norms' ratings for nearly

all faces. SM’s panels recreated with data from Figure 2 of Adolphs, Tranel, and Damasio, *Nature*, 1998.

Discussion

A significantly enhanced approach-related bias, relative to controls, was uncovered in three rare patients with selective bilateral amygdala lesions by comparing ratings of faces in a whole and occluded condition. A default bias was demonstrated by the patients’ greater willingness to approach a face (i.e., less threatening and more trustworthy ratings) in the low-information occluded condition.

Returning to our example situation of walking into a dark room, the normal response to an ambiguous situation is risk-assessment (Blanchard et al., 2011) — given insufficient information to determine whether a threat is present, one should pause and gather more information before proceeding; our patients’ ratings indicated that they would simply enter the room; at the other end of the spectrum, anxious individuals might flee the dark room before gathering further evidence. Future experiments exploring individual differences (e.g., trait/state anxiety, perceived dominance, history of exposure to physical/social threat or betrayal) will be important to both validate our task and determine what factors beyond amygdala damage relate to heightened approach tendencies. It is worth noting that in our sample a few control individuals had an approach bias similar to that of the patients, emphasizing the importance of future studies to determine the cause of these individual differences.

In humans, given a lack of stimulus information, an exploratory tendency may normally promote a default approach bias, the “positivity offset” in the Evaluative Space Model (Norris et al., 2010), similar to our observed shift in “positivity” approach ratings between the occluded and whole face conditions (Figure 2.2.7). This occluded positivity bias was observed both in controls and patients, but enhanced in the patients. It is worth noting that while in this class of stimuli a lack of information, which was similar to ambiguity, encouraged approach behavior in both healthy controls and the amygdala lesion patients, for some classes of stimuli, given some additional contextual cues, ambiguity in and of itself might be perceived as threatening and prompt avoidance behavior, especially in healthy controls.

While patients gave stronger approach ratings than controls, they were not completely indiscriminate: their judgments differed more in degree than direction (Table 2.2.2). Future work should test how their enhanced approach bias extends to (1) other classes of degraded stimuli (including non-linguistic tasks to better facilitate cross-cultural comparison), and (2) the real world. Showing abnormal proxemic (i.e, personal space) behavior and a tendency to approach real threatening stimuli (e.g., snakes) in these three patients, as has already been done in patient SM (Feinstein et al., 2011; Kennedy et al., 2009), would further corroborate a default approach bias. As confirmed in preliminary testing, at least BG has abnormally small personal space and fear responses (D.P. Kennedy, J. Feinstein, & R. Adolphs, personal communication). Testing participants' actual behavior is crucial – compensatory processing may allow them to give more "correct" explicit ratings: for example, although SM abnormally approached actual snakes without showing any fear, beforehand, she verbally insisted that she "hates" snakes and "tries to avoid them" (Feinstein et al., 2011).

Differences amongst the amygdala lesion patients need to be resolved. Amygdala damage can prompt two distinct approach processes — a default bias, as well as a face-specific bias — both of which can operate simultaneously. Removing facial feature information from facial stimuli allowed us to challenge participants to indicate a default bias while working within the general category of facial stimuli. Across all participants, responses to facial features were variable and the patients were similar to controls. SM's whole face ratings had been different from controls (Adolphs, et al., 1998); this deviation is in line with the heterogeneity of impairments reported in bilateral amygdala damage (Adolphs et al., 1999; Hamann et al., 1996; Siebert, Markowitsch, & Bartel, 2003) and likely reflects compensatory processing (Becker et al., 2012; Scheele et al., 2012). SM's impairment for whole faces hints at progressive amygdala damage/impairment, as is expected in Urbach-Wiethe disease (Appenzeller et al., 2006).

While patient differences in a face specific deficit need to be further explored and explained based on precise anatomical differences, the present study focused on isolating a stimulus-independent shift, which will clearly interact with responses to facial features. Sometimes, "good" facial features (determined idiosyncratically) helped occluded-avoided faces; sometimes "bad" features harmed occluded-approached faces. However, across the entire stimulus set, a general occluded-approach bias could be observed.

Mechanistically, the patients' approach bias may relate to a specific cautionary deficit, related to disrupted vigilance (Davis & Whalen, 2001; Paul J. Whalen, 2007). This viewpoint is anatomically compatible with the amygdala launching a defensive behavioral response to coincident sensory and contextual danger signals, conveyed via the temporal and prefrontal cortices, respectively (Freese & Amaral, 2009).

However, the patients' approach bias can be explained by a more general mechanism of amygdala function. A general role in processing saliency/self-relevance (Cunningham

& Brosch, 2012; Harrison & Adolphs, 2015; Sander, Grafman, & Zalla, 2003) is compatible with a wider array of known amygdala activity. The amygdala contributes to negative *and* positive reinforcement (Murray, Izquierdo, & Malkova, 2009), and processes positively and negatively valenced stimuli (Anderson et al., 2003; Hamann et al., 2002). In rats (Hatfield, Han, Conley, Gallagher, & Holland, 1996) and nonhuman primates (Izquierdo & Murray, 2007; Málková, Gaffan, & Murray, 1997), basolateral amygdala lesions interfere with reinforcer devaluation, such that an animal will indiscriminately approach devalued food items, similar to our patients' default approach bias. Hypothetically, the basolateral nucleus, damaged in our patients, updates the self-relevant value of a stimulus (Murray et al., 2009).

A saliency/relevance explanation binds our default bias finding with prior findings in amygdala lesion patients: amygdala lesions do not preclude the ability to experience fear - indeed, CO_2 inhalation can induce fear and panic in amygdala lesion patients (Feinstein et al., 2013), but instead inhibit proper orienting to stimuli (Spezio, Huang, Castelli, & Adolphs, 2007), which often results in a diminished ability to experience (Feinstein et al., 2011) or recognize (Adolphs et al., 2005) fear. Proper orienting can recover this ability: in SM, fear is correctly identified following explicit top-down instruction to look at the eyes (Adolphs et al., 2005).

Our finding of an enhanced default positivity bias suggests a further role for the amygdala in setting a default on what is potentially relevant or salient, normally preventing us from approaching situations that may be threatening, while simultaneously permitting exploration. In our patients, this balance is shifted. Similarly, in psychiatric disorders featuring dis-regulation of the amygdala (e.g., anxiety disorders (Davis, 1992; Etkin &

Wager, 2007) and autism (Baron-Cohen et al., 2000; Castelli, Frith, Happe, & Frith, 2002; Dalton et al., 2005)), stimuli are not correctly evaluated, from shifted baseline biases as well as under- or over-weighting the threat, social importance, or relevance of stimuli.

In summary, contrasting judgments of occluded and whole faces, we uncovered a stimulus-independent approach bias following bilateral amygdala damage. Future directions include (1) testing for a default approach or avoidance bias in psychiatric disorders for which the amygdala is implicated, as well as (2) developing implicit tests of an approach bias to circumvent potential compensatory mechanisms, and (3) devising tests to provide a clearer mechanistic account of our findings.

Chapter 2.3

ECOLOGICAL STRUCTURING OF HUMAN DEFENSIVE RESPONSES: EVIDENCE FROM JUDGMENTS OF PHYSICAL AND PSYCHOLOGICAL THREAT SCENARIOS

How humans react to threats is a topic of broad theoretical importance, and also relevant for understanding anxiety disorders. Many animal threat reactions exhibit a common structure, a finding supported by human evaluations of written threat scenarios that parallel patterns of rodent defensive behavior to actual threats. Yet the factors that underlie these shared behavioral patterns remain unclear. Dimensional accounts rooted in Darwin's conception of antithesis explain many defensive behaviors. Across species, it is also clear that defensive reactions depend on specific situational factors, a feature long emphasized by psychological appraisal theories. Our study sought to extend prior investigations of human judgments of threat to a broader set of threats, including natural disasters, threats from animals, and psychological (as opposed to physical) threats. Our goal was to test whether dimensional and specific patterns of threat evaluation replicate across different threat classes. 85 healthy adult participants selected descriptions of defensive behaviors that indicated how they would react to 29 threatening scenarios. Scenarios differed with respect to ten factors, e.g., perceived dangerousness or escapability. Across scenarios, we correlated these factor ratings with the pattern of defensive behaviors endorsed. A decision tree hierarchically organized these correlation patterns to successfully predict each scenario's most common reaction, both for the original sample and a separate replication group (n=22). At the top of the decision tree, degree of dangerousness interacted with threat type (physical or psychological) to *predict dimensional approach/avoidance behavior. Subordinate nodes represented specific defensive responses evoked by particular contexts. Our ecological approach emphasizes the interplay of situational factors in evoking a broad range of threat reactions. Future studies could test predictions made by our results to help understand pathological threat processing, such as seen in anxiety disorders, and could begin to test underlying neural mechanisms.*

Introduction

Darwin famously noted the striking phylogenetic continuity of emotional behaviors, including responses to threat (Darwin, 1872/1965). Defensive behaviors, ranging from flight to attack, have evolved to deal with environmental challenges that show a common structure across all animals: the need to attack an aggressor, to flee a predator, or to hide from an inescapable threat, to name only a few prototypical situations. Over the years, several empirical and theoretical studies, largely rooted in biology and ethology, have supported the idea of common structure in defensive behaviors across species, ranging from rodents to humans (Blanchard, Hynd, Minke, Minemoto, & Blanchard, 2001). Various schemes have been proposed for how these are organized, ranging from ethologically-identified (Blanchard, Blanchard, & Hori, 1989) factors like risk assessment (Blanchard et al., 2011) to dimensional accounts including threat imminence (Fanselow & Lester, 1988) and a classic approach/avoidance account whereby all motivated/emotional behaviors are organized along an appetitive and defensive system (Lang, Bradley, & Cuthbert, 1998).

On the other hand, the literature in affective psychology has rarely incorporated specific details of the data from nonhuman animals, although this literature clearly does acknowledge the biological roots of human defensive behaviors (Lazarus, 1991; McNaughton & Corr, 2004, 2009; Adam M. Perkins, Cooper, Abdelall, Smillie, & Corr, 2010). Here, we asked people to select hypothetical defensive behaviors to descriptions of a range of physically threatening situations, as well as to situations of social psychological threat. It is important to emphasize at the outset that we rely on verbal report and ratings, as is common in many psychological studies in humans (e.g., (Cottrell & Neuberg, 2005)), rather than on actual observed defensive behavior. Verbal report to hypothetical scenarios by humans has been found in previous studies to correlate with actual rodent behavior patterns across three laboratories (Blanchard et al., 2011), and we used it here as a first approach to assess responses for which live exposure would be ethically difficult to obtain. Specifically, in the current experiment, threatening situations include situations of social psychological threat, (e.g., blackmail), social physical threat (e.g., stalking), as well as physical threat from other species and natural disasters. Inclusion of these different threat categories underlies an attempt to bridge our understanding of basic approach-avoidance reactions to predators and other physical threats on the one hand, with a characterization of defensive reactions to less physical but more psychological intra-species threats that relate to issues of social inclusion, social hierarchies, and social dominance on the other hand. It is worth noting that socially modulated threat reactions have been observed across diverse phylogenetic classes, including fish (Fernald, 2012), and mammals (Tamashiro, Nguyen, & Sakai, 2005) ranging from rodents (Scheibler, Weinandy, & Gattermann, 2004) to primates (Abbott et al., 2003; Dewaal, 1986).

Defensive behaviors in rodents and primates have been extensively studied, and related to human behavior, such as in the case of humans physically freezing in response to threatening stimuli (Hagenaars, Oitzl, & Roelofs, 2014). Innate patterns of defensive behavior have been identified in some detail in rats: e.g., high magnitude threats elicit a flight response, only if an escape route is available; if an escape route is not available, rodents will freeze, show a defensive threat (e.g., vocalization), or launch an explosive defensive attack depending on the distance of the threat (Blanchard & Blanchard, 1989). Very specific releasing-stimulus like cues can be sufficient to trigger the behavior: for instance, a predator-like visual looming stimulus (just an expanding black circle on the ceiling) is sufficient to produce robust freeze or flight (Yilmaz & Meister, 2013), with the likelihood of each behavior dependent upon the presence of a hiding place in the arena. The size of an enclosure also seems to affect the use of flight or freeze behavior (Kim et al., 2013). The validity of the use of rodent defensive behaviors as a model for human defensive reactions remains an open question, partially addressed by a study that attempted to make direct comparisons between the two species (Blanchard et al., 2001). In that study, written descriptions of physically threatening scenarios were manipulated in terms of factors known to alter rodent behavior, such as the magnitude of threat, escapability of the situation, ambiguity of the threat stimulus, distance between the threat and the subject, and the presence of a hiding place. Strikingly, most of the human participants' choices of what they would do when faced with these scenarios paralleled the rodent behavior observed when a rat faced the same real situational factors. Moreover, the human choices of defensive behaviors paralleled the specific animal defensive behaviors (e.g., defensive attack for near threats; risk assessment for ambiguous threats; hiding when there is a hiding place) across different cultural settings (see Table 2.3.6, Discussion), e.g., in Brazil (Shuhama, Del-Ben, Loureiro, & Graeff, 2008) and Wales (Perkins & Corr, 2006) with "minor or potentially easily explained differences" (Blanchard et al., 2011) compared to the original patterns observed in Hawaiian participants (Blanchard et al., 2001), suggesting cross-cultural generality at least for the physically threatening scenarios investigated in those studies.

These prior studies that built upon rodent behaviors fit well with dimensional accounts of emotion. Although Darwin is often cited in support of discrete emotion theories, Darwin's early principle of antithesis (Darwin, 1872/1965) in fact set the framework for conceiving of emotional behaviors as having a dimensional structure:

> When actions of one kind have become firmly associated with any sensation or emotion, it appears natural that actions of a directly opposite kind…should be unconsciously performed…under the influence of a directly opposite sensation or emotion. (p. 67)

Darwin's notion of antithesis roughly maps onto the modern dimension of "valence". However, the main point that he made, of course, was that emotions, including defensive behaviors, in humans would look similar and have a similar structure to that of other mammals. According to one theory, evolutionary selection can give rise to what have been called "rules of thumb" that advantageously guide behavior under typical ecological conditions (McNaughton & Corr, 2009). These rules of thumb can be conserved across species that have evolved in similar environments, such that emotional behaviors evoked by certain circumstances in one species will evoke similar emotional behaviors in another species faced with the same challenges. If the species are not too phylogenetically distant, one would even expect these shared emotional responses to be mirrored in conserved neural structures (Gross & Canteras, 2012; LeDoux, 2012). It is unknown precisely which features of a shared environment would come into play in this picture, but there are some good candidate dimensions, such as predator imminence (the physical distance and time to discovery between predator and prey) (Fanselow & Lester, 1988) and uncertainty. Notably, these dimensions are broad, can be observed across many species and provide important context for many situations.

However, in addition to such broad dimensional structure, it is clear that emotions also exhibit patterns of response tailored to specific situations that evoke them. For instance, Gray and McNaughton (2000) have proposed that two clusters of defensive behaviors identified also in rodent studies (Blanchard, Blanchard, Rodgers, & Weiss, 1990; Blanchard, Flannelly, & Blanchard, 1986) represent the action of two brain systems, one controlling anxiety, the other fears, and that differences in the reactivity of these systems give rise to personality differences and ultimately could explain psychopathology (Gray & McNaughton, 2000; McNaughton & Corr, 2004). It has been proposed that different circuits involving the amygdala and the bed nucleus of the stria terminalis (BNST) mediate phasic fear versus more sustained anxiety-like fear (Davis, Walker, Miles, & Grillon, 2010). The distinction between anxiety and fear is important, mapping onto those defensive situations where engagement and the acquisition of further information is adaptive (in the former case), and those where disengagement and survival are most important (in the latter case).

Social fear is yet another category, linked to a possibly domain-specific class of eliciting stimuli. There is evidence that social fear is processed differently from other types of fear: in mice, independent hypothalamic circuits for social (intra-species) and predator (inter-species) fear have been identified (Silva et al., 2013). Do patterns of threat response observed in other species extend to the social domain in humans, especially to more psychological as opposed to physical social threat scenarios? Although the Blanchard study (2001) and its replications (Perkins & Corr, 2006; Shuhama et al., 2008) investigated physical threat between humans, psychological threat has rarely been directly compared to physical social threat. Social sources of threat have been studied experimentally in humans with paradigms such as the Trier Social Stress Test (Kirschbaum, Pirke, & Hellhammer, 1993), or the cyberball game (Williams & Jarvis, 2006), which relates to ostracism and social hierarchy, issues that have been explored since Milgram's famous obedience studies (Milgram & Van den Haag, 1978). Nonhuman primates have also been shown to be have mental representations of social hierarchy (Dahl & Adachi, 2013) (a capacity even demonstrated in fish (Grosenick, Clement, & Fernald, 2007)) and are sensitive to social inequality (Brosnan, Schiff, & De Waal, 2005).

Testing the category of social psychological threat in the present experiment is pertinent to open questions remaining from the three prior physical threat scenario studies (Blanchard et al., 2001; A. M. Perkins & Corr, 2006; Shuhama et al., 2008). For instance, Blanchard (2001) argues that risk assessment can play a crucial role in detecting and analyzing threat stimuli. Risk assessment is a highly adaptive process that takes into account the type and location of the threat, as well as the escapability of the situation to predict the most optimal defense mechanism. In fact, risk assessment becomes more important when there is some degree of ambiguity in the situation (Blanchard et al., 2011), as is more often the case in situations of psychological social threat — a threat category we investigate here in our extension of the original Blanchard study (2001). In the psychological domain, most complexity arises from the situational context (with: peers, inferiors, or superiors; or location: work, novel setting, recreational location). We would thus expect that each of these situations creates a unique hierarchy of threat characteristics to be evaluated. Behavioral hierarchies are a prominent ethological concept: according to Tinbergen (1951), an animal will enter one of a handful of broad behavioral hierarches, e.g., defense or reproduction, that then dictate further subordinate behavioral repertoires, all depending on an animal's evaluation of the environmental context.

The situational evaluation emphasized by ethologists offers a point of contact with the human psychology literature, notably appraisal theory as articulated by Arnold (1960), Lazarus (1991), and Scherer (2001, 2009). Appraisal theory postulates so-called "stimulus-evaluation checks" — specific dimensions upon which stimuli are sequentially or hierarchically assessed — that are used to appropriately assess context across points in time (Scherer, 2001, 2009). For instance, first, a stimulus would be checked for relevance; if it were novel and/or (un)pleasant, it will be attended and possibly prompt initial approach or avoidance responses (e.g., pupil dilation, heart rate changes, locomotion). Once attended, the implications of the stimuli would be checked – whether they were likely to produce a consequence for the organism, reflecting the urgency with which they require a reaction. Subsequent checks relate to the organism's coping potential for likely consequences as well as how those consequences relate to issues of normative significance such as ideas of self and social norms. Each of these hierarchical evaluations or stimulus-evaluation-checks relate to patterns of bodily, neural, and behavioral response, and can be conceptualized as a temporal unfolding of emotion (Sander, Grandjean, & Scherer, 2005).

The first goal of our study was to test the generalizability of dimensional factors and specific situational appraisal in guiding defensive responses across a broad range of threats. Recently, a Survival Optimization System (SOS) model has been theoretically proposed to account for cross-species threat responses (Mobbs, Hagan, Dalgleish, Silston, & Prévost, 2015). A notable feature of the SOS model is that it integrates dimensional (imminent threats elicit reflexive responses) and appraisal-like accounts of threat responses. We predict that our empirical account of the ecology of human threat reactions will also highlight the relative strength of dimensional accounts in accounting for basic behavior (specifically, approach-avoidance), while situational appraisal will predict specific instantiations of approach and avoidance behaviors.

In addition to extending the range of threat scenarios, and hence the anticipated range and specificity of defensive behaviors, a second goal of our study was to then use this more comprehensive inventory of threat responses to create a generalized model for characterizing human defensive behavior toward threat. Inspired by both the appraisal theory models discussed above, Tinbergen’s behavioral hierarchies (1951), and the recent SOS model (Mobbs et al., 2015), we aimed to build a hierarchical decision-tree that would accurately predict a participant’s threat response based on features of the threat stimulus. To build such a general decision tree, we aimed to sample different sources of threat, although each type of threat was only sparsely sampled by a few specific scenarios. We hypothesized that many types of defensive states — anxiety, fear, panic — could be mapped to a proximity factor similar to that in predator imminence theory

(Fanselow & Lester, 1988). While basic approach/avoidance processes might remain the same across threat domains (e.g., psychological and physical), we also expected to find differences linked to the specific demands required by certain contextually dependent types of threat (Silva et al., 2013). We achieved our two aims of (1) contrasting ecological patterns of threat response across a broad class of threats, including psychological threats, as well as (2) organizing those patterns of threat response into a decision tree incorporating dimensional, approach-avoidance and hierarchical, appraisal-like features to eventually predict specific defensive responses.

Materials and Methods

Participants

We tested five nonoverlapping groups of participants over the internet as described below. The dependent measures they provided are summarized in Table 2.3.1.

Table 2.3.1. Dependent Measures in Experiments. All dependent measures were given for all 29 threat scenarios (cf. Table 2.3.2).

Experiment	Measures	
Main; Replication (n=85; 22)	Physical Scenarios Response Options	Psychological Scenarios Response Options
	1. Hide	1. Hurt the other person physically
	2. Freeze, become immobilized	2. Hurt the other person verbally or yell
	3. Run away, try to escape, remove self (flight)	3. Verbal confrontation
	4. Threaten to scream or call for help	4. Avoidance or ignore the situation
	5. Yell, scream, or call for help	5. Hide or remove self from the situation
	6. Threaten to attack	6. Freeze up
	7. Attack or struggle	7. Ask for advice and/or plan a course of action
	8. Check out, approach, or investigate (risk assessment)	8. Negotiation
	9. Look for something to use as a weapon	9. Report to a higher authority
	10. Beg, plead for mercy, or negotiate	
Approach-Avoid (n=31)	Approach-Freeze-Avoid ratings on a 9 point slider	
Factor Ratings (n=33)	Slider: Low (1) to High (5)	
	1. Dangerousness	6. Immediacy
	2. Escapability	7. Ability to communicate with the threat
	3. Ambiguity	8. Ability to mitigate or change the threat
	4. Distance to threat	9. Ability to harm the threat
	5. Presence of a hiding place	10. Ability of others to help

Ethics statement. All participants provided informed assent to participate in research under a protocol (RA-392: “Anonymous Online Surveys of Threat Assessment”) that was approved by the Caltech Committee for the Protection of Human Subjects as Institutional Review Board exempt under Part 46.101(b)(2), “Protection of Human Subjects” of Title 45 of the U.S. Code of Federal Regulations. Instead of providing formal written consent, in our assent procedure, at the beginning of the online experiment, anonymous participants read a description of the experiment in which they were told they were free to cease participation at any point.

Main experiment. 88 English-speaking participants living in the United States were recruited through Amazon's Mechanical Turk. Participants were paid approximately $8-10 upon the completion of the survey, and were given a maximum of 5 hours to complete the survey online. Responses from 85 (44 female) participants (age = 33±9 years, mean±SD) were analyzed. Data from two participants were excluded since the subjects had a diagnosis of PTSD, and a third participant was excluded because of an anxiety diagnosis and high state anxiety as measured by the State-Trait Anxiety Inventory (Spielberger, 1983). High state or trait anxiety cutoff scores were defined as 1.5 standard deviations greater than the mean score off all participants across all 4 experiments (all but the factor rating task); cutoff scores were 58 for state anxiety and 60 for trait anxiety. Forty-four percent of participants had a college degree or higher.

Scenario factor ratings. An additional independent set of 33 (17 female) American raters (age = 34±11), were recruited through Mechanical Turk to quantitatively characterize the scenarios, on a scale from 1 (low) to 5 (high), with respect to 10 pre-defined factors. The scenarios were designed in advance to vary along these dimensions; external ratings allowed us to validate and quantify variation in pre-assigned low/moderate/high ratings.

Replication experiment. Results from the main experiment were used to build a decision tree that predicted people's responses to threat scenarios. To test the reliability of that decision tree, an additional set of 25 American participants were recruited through Mechanical Turk to replicate the original threat scenario experiment. Responses from 22 (13 female) participants (age = 33±11) were analyzed; 3 participants were excluded for anxiety diagnoses and high trait anxiety.

Approach-avoidance experiment. To directly relate responses for psychological and physical threat scenarios, whose specific response options differed and thus made them impossible to compare directly in the main experiment, approach-avoidance responses to all scenarios were collected from 35 participants. Responses from 31 (18 female) participants (age = 33±9) were analyzed; 2 were excluded because of a diagnosis of PTSD, and a further 2 were excluded because of diagnoses of anxiety and high state and trait anxiety.

Materials

Participants were presented with twenty-nine scenarios in total (Tables 2.3.2). We designed the scenario descriptions to be relatively concise, simple, and clear. Each scenario contained an instance of one of four categories of potentially threatening situations: one that involved a natural disaster (N; 4 scenarios), an animal (A; 5 scenarios), a physical interaction with another person (P; 11 scenarios), or, in opposition to these three physically threatening categories (20 scenarios total), an interaction with another person that was more psychologically threatening (S; 9 scenarios). All scenarios included in the human physical category were directly taken from Blanchard et al. (2001).

Table 2.3.2. Threat Scenarios Presented to Subjects. Each scenario is assigned a brief descriptor and label, used throughout the paper. N = Natural; A = Animal; P = Physical; S = Psychological. All Physical scenarios taken from (Blanchard et al., 2001).

Descriptor	Scenario	Label
Hurricane 10 min	Imagine you are living in New York City, and you hear on the news that a new hurricane is arriving in 10 minutes. It is going to hit the city any moment now. This one is going to be even bigger than Hurricane Sandy, and no one knows what to make of it.	N1
Hurricane 24 hr	Imagine you are living in New York City, and you hear on the news that a new hurricane is coming tonight. This one is going to be even bigger than Hurricane Sandy, and no one knows what to make of it.	N2
Tornado 10 min	Imagine you are living in Kansas, and you hear on the news that a tornado is approaching your town in the next 10 minutes. You couldn't have anticipated this, and this is an emergency. You live on the countryside, and it takes at least 1 hour to reach another house or any form of help.	N3
Tornado 24 hr	Imagine you are living in Kansas, and you hear on the news that a tornado is approaching your town tonight. You live on the countryside, and it takes at least 1 hour to reach another house or any form of help.	N4
Bear 50 yds	You are camping in the mountains. You go out by yourself to take a walk, and you suddenly see a bear approaching from 50 yards away.	A1
Bear 1 yd	You are camping in the mountains. You go out by yourself to take a walk, and you suddenly see a bear approaching from 1 yard away.	A2
Bear 10 yds	You are camping in the mountains. You go out by yourself to take a walk, and you suddenly see a bear approaching from 10 yards away.	A3
Shark 100 yds	You are swimming near the beach, and you suddenly realize there is some animal that is approaching you from 100 yards away. It may be a shark but you are unsure.	A4
Shark 10 yds	You are swimming near the beach, and you suddenly realize there is some animal that is approaching you from 10 yards away. It may be a shark but you are unsure.	A5
Whisper	Alone at home one night, you have settled down to read a book when you hear some movement right outside of your window. You cannot see anything, but when you listen more closely, it sounds like people whispering.	P1
Elevator	You are alone in an elevator late at night. As it stops and the doors open, a menacing stranger rushes in to attack you, blocking the door.	P2
Stoplight	You are alone in a car on your way home. While stopped at a traffic signal, an angry stranger begins banging on your car window and yelling threatening things at you.	P3

Tailgating	Driving along a two-lane road, you see in your rear-view mirror that a car is dangerously tailgating you. They cannot pass and begin honking their horn aggressively at you while continuing to follow too closely.	P4
Corner	It is past midnight and you are walking through an unfamiliar part of town. As you round a corner, you accidentally run into a man. He becomes angry and shoves you.	P5
Acquaintance	You and someone you do not really know that well are standing around and talking in an empty parking lot. The acquaintance begins to shove and push you. You are unsure whether s/he (same sex as you,) is serious or just kidding around.	P6
Park	You are outside in a park area at night when you see a menacing stranger with a knife about 30 feet away directly approaching you. It is obvious the person is planning to attack you.	P7
Grab	You are alone as you exit an empty campus building late one night. Just as you get outside you feel a hand grab your arm.	P8
Noise	You are sleeping in bed during the night, but suddenly wake up thinking you have heard a suspicious noise. It is dark and you are alone.	P9
Phone	You are alone at home one night about to go to bed when the phone rings. You answer it, and there is an unfamiliar voice on the other end. It tells you that he or she is right outside of your house and hangs up.	P10
Bomb	Coming home one day, you find an unexpected shoe-box-sized package waiting for you by the mailbox. As you sit down to open it, you notice a faint ticking sound that appears to come from inside the package.	P11
Blackmail Email	One of your colleagues has blackmailed you via e-mail, saying that he or she will spread rumors about you. The rumor is an embarrassing one that deals with your personal issues.	S1
Blackmail Face	You are taking a break at work and one of your co-workers approaches you. He/she is blackmailing you face-to-face. He or she threatens to spread rumors about your personal issues.	S2
Boss	It is almost the holiday season and you are planning to take a few days off from work. You bring this up at your weekly meeting, and your boss disapproves. Your immediate boss threatens to fire you if you take a break.	S3
Rumor	Recently, you have noticed that one of your co-workers have been talking behind your back at work. He/she has been spreading rumors, and seems to drop negative remarks about you to your immediate boss as well.	S4
Cellphone	You are working for a cellphone company, and you have had a new design for a phone in mind. You decide to share it with a colleague. You tell him/her that this may be revolutionary and that he/she should keep it confidential. Sooner or later, you realize that the colleague has already spread the idea and has taken credit for it.	S5
Party	Your close friend is having a holiday party, and has not invited you to this party.	S6

Bar	A group of your colleagues go out to a bar after work, but no one asks you to join them.	S7
Homophobic	Imagine you are secretly gay and a close family member is homophobic. During a family party, they make a derogatory comment toward gay people.	S8
Political	You realize that you and your boss hold very different political views. After work, you and your boss get together and he/she asks about your political views.	S9

These scenarios were designed to vary along 10 different factors (Table 2.3.1). The first five factors were derived from Blanchard et al. (2001), and we included additional factors (e.g., ability to communicate to capture human vs. animal scenarios; ability to mitigate to capture elements of social support and social hierarchy) to reflect our expanded set of scenarios.

In order to rate each of these scenarios along the above dimensions, independent raters were each presented with 10 randomly selected scenarios from the set of 29 and asked to rate each of those scenarios for each factor (dangerousness, escapability, etc.) on a scale of 1 (low) to 5 (high) for all 10 factors. Through random assignment, each scenario was rated by at least 8 and up to 15 individuals (mean = 11.4).

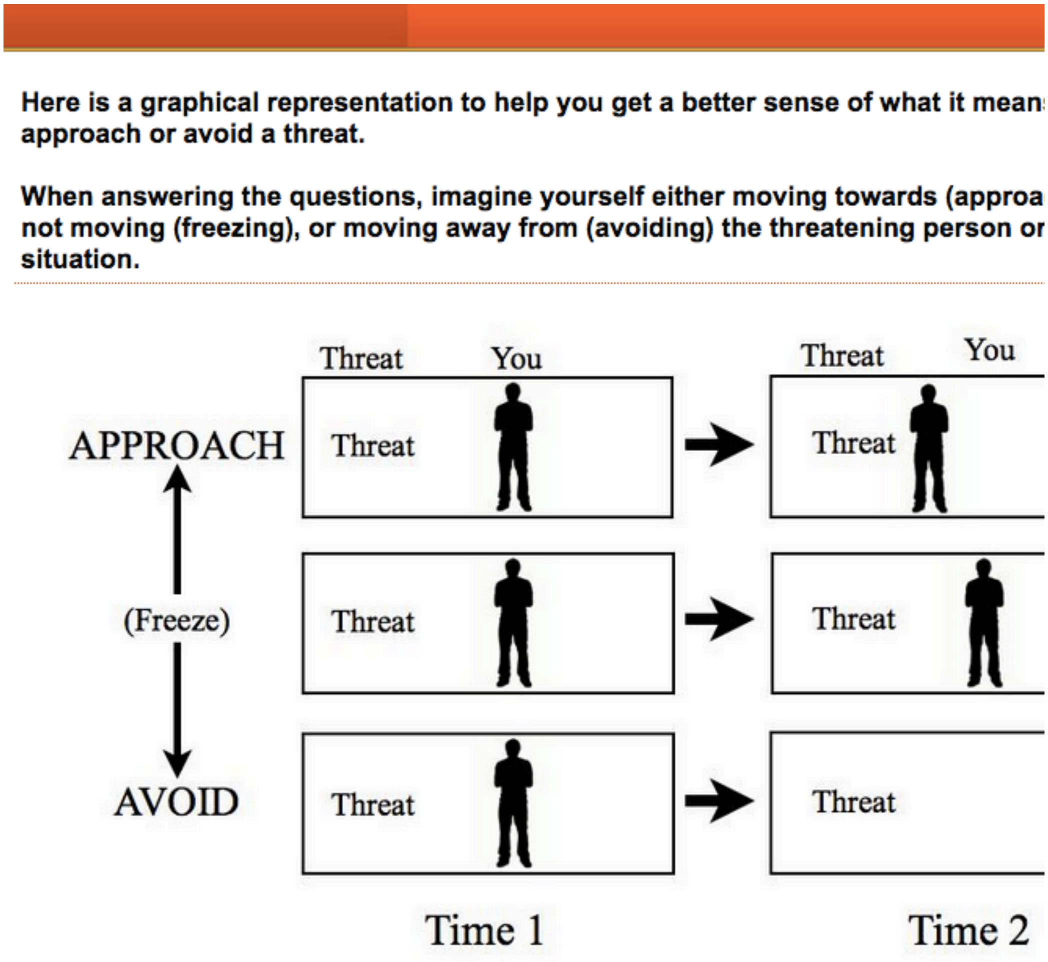
In order to rate each of these scenarios along the above dimensions, independent raters were each presented with 10 randomly selected scenarios from the set of 29 and asked to rate each of those scenarios for each factor (dangerousness, escapability, etc.) on a scale of 1 (low) to 5 (high) for all 10 factors. Through random assignment, each scenario was rated by at least 8 and up to 15 individuals (mean = 11.4).

Procedure

Main and Replication Experiment. Participants were asked to read each scenario and indicate their most likely first-responses. Participants also had the option to choose up to two additional options. If they had selected multiple options, participants were asked to rank their responses from 1 (most likely) to 3 (least likely); here we analyze only the data from the top response option. The psychological scenarios were given a separate category of response options. There were 10 response options for the physical scenarios (natural, animal, and human) and 9 for the psychological scenarios (Table 2.3.1).

Scenario Factor Ratings. Participants were randomly presented 10 of the 29 scenarios, and asked to use a sliding scale to provide factor ratings for each scenario on a scale from 1 (low) to 5 (high), with respect to all 10 of our pre-defined factors (cf. Table 2.3.1). The starting position of the sliders was randomized. Because participants were only presented 10 of the 29 scenarios, factor ratings for each scenario were provided by a subset of the 33 participants. A minimum of 8 and maximum of 15 participants rated each scenario, with 21 of the 29 (72%) scenarios being rated by at least 10 participants.

Approach-Avoidance Experiment. Participants were instructed to read each of the 29 scenarios and imagine how they would respond to the threat in terms of approach/avoidance. An illustration (Figure 2.3.1) explained the concept of approaching/freezing/avoiding a threat. Participants indicated their response on a 9-point slider, which began in the middle of the range. Participants were asked to imagine themselves as the slider moving either toward (left) or away (right) from the threat.



Here is a graphical representation to help you get a better sense of what it mean
approach or avoid a threat.

When answering the questions, imagine yourself either moving towards (approa
not moving (freezing), or moving away from (avoiding) the threatening person or
situation.

Figure. 2.3.1. Approach/avoid instructions. Participants (n=31) in the approach-avoid experiment viewed these instructions, which made explicit that approach-avoidance ratings related to taxis relative to the source of threat.

Analyses

Factor ratings. Means and standard deviations are reported for the independent ratings along 10 factors, and were used to characterize each scenario. The averaged independent ratings ranging from 1.00 to 5.00 were normalized across all scenarios (Normalized Scenario Score = (Scenario Score - Minimum Factor

Score)/(Maximum Factor Score - Minimum Factor Score)) such that the lowest rated scenario for a factor received a score of 0, and the highest rated scenario received a score of 1.

Factor-specific response correlations. To quantify relationships between each of these rated factors and the set of defensive behaviors, we calculated Pearson product-moment correlations for every possible factor-response pairing. The first vector in this correlation consisted of the mean factor rating given to each scenario; the second vector consisted of the proportion of participants who chose a response option as their first choice for each scenario. Because response options differed between physical and psychological scenarios, all analyses were conducted separately for those scenario categories. To visualize patterns of correlations, correlation coefficients were reported in heatmaps with factors and response options organized along rough imminence and approach-avoidance continua, respectively.

Factor-approach/avoid response correlations. To directly relate responses to psychological and physical threat scenarios, the same correlational analysis completed for specific response options was completed according to the proportion of subjects who chose approach-avoid ratings corresponding to categorical approach/freeze/avoid responses.

Gender differences. The above analyses were completed for males and females separately, as well as together. While minor differences were found between male and females' first choice defensive behaviors for some scenarios, as these differences mirrored prior findings (Blanchard et al., 2001) (Table 2.3.3) and did not reflect our primary interest, males and females are consequently pooled in the results with specific differences noted only as they arise.

Table 2.3.3. Gender Differences. Comparison of male and female top response options in scenarios for which their first responses differed. When applicable, these differences are compared to prior results (Blanchard et al., 2001) in the comments column. While the top response option in scenario P10 did not differ between males and females, the scenario is reported since Blanchard (2001) had observed a gender effect. * Denotes a tie between first-choice response options.

Scenario	Male Top Responses (n=41)	Female Top Responses (n=44)	Comments
P2 Elevator	1. Attack or struggle 2. Yell or scream	1. Yell or scream 2. Attack or struggle	Blanchard found the same first choices. First and second choices switched by gender.
P8 Grab	1. Attack or struggle* 1. Risk assessment* 2. Yell or scream	1. Yell or scream 2. Risk-assessment 3. Attack or struggle	Blanchard found the same first choices. Top three choices the same across genders.
P10 Phone	1. Look for a weapon* 1. Risk-assessment* 2. Yell or scream	1. Look for a weapon 2. Yell or scream 3. Risk-assessment	Observed no difference in first response. Top three choices the same across genders. Blanchard's first female response was hide.
N2 Hurricane	1. Flight 2. Risk-assessment	1. Risk-assessment 2. Flight	Comparison to Blanchard not applicable. First and second choices switched by gender.
S1 Blackmail	1. Report to an authority 2. Verbal confrontation	1. Verbal confrontation 2. Report to an authority	Comparison to Blanchard not applicable. First and second choices switched by gender.

Single approach-avoid score. A single approach-avoidance score across participants' ratings was derived for each scenario, with more positive scores indicating approach, more negative avoidance, and those close to zero either indifference or freezing. To construct this score, first, approach, freeze, and avoidance scores were calculated for each scenario. These scores were the proportion of subjects choosing categorical approach (ratings of 1, 2, 3, 4)/freeze (ratings of 5)/avoid (ratings of 6, 7, 8, 9) for each scenario, with the proportion of subjects choosing an approach or avoidance rating weighted by subjects' median approach or avoidance score. All categorical approach/freeze/avoid scores were rescaled on a 0 to 1 interval. Then, a single approach-avoidance score took the signed absolute value of the difference between the rescaled approach and avoidance scores, and penalized it by subtracting the magnitude of the rescaled freeze score for that scenario, such that the single approach-avoid score for scenarios that had larger freeze scores were closer to zero.

Decision tree.

A descriptive decision tree that predicted responses to threat scenarios based on features of those scenarios was created through a multi-step process.

The first major step was describing convergence or divergence between physical and psychological scenarios. Two analyses guided this step. First, to test our hypothesis that psychological and physical threats are characteristically distinct, we calculated the dissimilarity between all pairs of scenarios based on the factor ratings of those scenarios, using the correlation distance measure in Matlab's pdist function. These pairwise scenario dissimilarities were then visualized both (1) as a dissimilarity matrix heatmap, organized by scenario type — animal, natural, human physical, and psychological — and (2) according to multidimensional scaling of the dissimilarity distances between each scenario, to determine whether physical and psychological threats cluster separately. The results of this analysis partially guided early splitting of physical from psychological threat.

Secondly, we sought to determine whether basic approach/avoidance behavior to psychological and physical threats diverged according to any factors. To do this, the single approach-avoidance scores were correlated with factor scores to guide construction of the beginning of the decision tree.

After forming the top branches of the tree, which predicted primary approach/avoidance responses, we extended the tree to predict appraisal-related specific responses (e.g., risk assessment, attack, verbal confrontation, etc.). This portion of the tree was constructed by summarizing how scenarios with the same most popular response options varied with respect to factor ratings. In organizing these nodes of the decision tree, priority was given to explanatory factors that (1) clustered consistently to yield a common top response option (close ties were allowed), and that (2) made ecological sense or adhered to a priori hypotheses (e.g., about the importance of communication, threat imminence, etcetera). While there were no overall gender differences in basic approach-avoidance behavior, specific responses to scenarios sometimes varied with gender; therefore, when appropriate, gender was used as a late node in the tree.

After construction, we tested the predictive success of the tree by calculating the proportion of each individual participants' responses that were correctly predicted for both the original and replication groups and comparing this prediction to chance performance of around 10% (10 vs. 9 specific response options for physical and psychological threat, respectively). Importantly, the decision tree was derived based only on the original group's data, and thus the replication group was an independent set of data on which to test.

Results

Independent Factor Ratings

The scenarios were designed to vary along the 10 specified factors. Externally validating our construction, low (0)/moderate (0.5)/high (1) pre-determined factor ratings (assigned by the experimenters when constructing the scenarios) correlated significantly ($p<0.001$) with all measured factor ratings (scale 1 to 5) (measured in the factor-rating experiment (n=33)).

Across all four threat categories (physical, natural, animal, psychological), scenarios spanned the range of participants' (n=33) raw factor ratings well (Table 2.3.4). The physical scenarios had the highest ratings for escapability, ambiguity, and ability to harm (Table 2.3.5); these ratings were similar to the previously reported ratings for these scenarios (Blanchard et al., 2001). The natural scenarios were rated as especially dangerous and low in ability to communicate. The animal scenarios were rated high in immediacy, and, like the natural scenarios, low in ability to communicate. The psychological scenarios were rated high in ability to mitigate and ability of others to help as well as the ability to communicate.

Table 2.3.4. Scenario Factor Ratings. Lefthand columns demonstrate the range of raw ratings by factor, with lowest and highest rated scenarios listed. Righthand histograms show the number of scenarios that received an average rating corresponding to a score of 1 (low) to 5 (high).

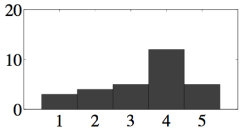
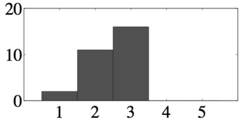
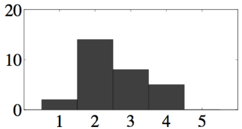
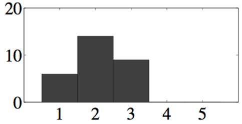
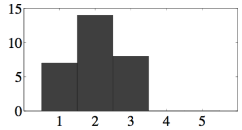
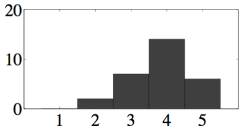
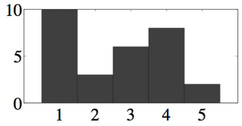
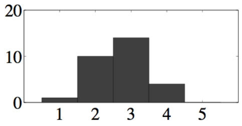
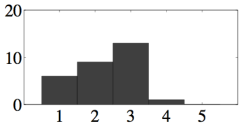
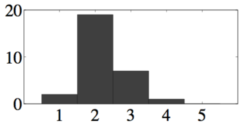
Factor	Low	Low scenario	High	High Scenario	Histogram of Ratings Across Scenarios
Dangerousness	1.00	Party; bar	4.94	Bear 1yd	
Escapability	1.70	Hurricane 10 min	3.44	Acquaintance	
Ambiguity	1.25	Boss	4.5	Acquaintance	
Distance	1.08	Grab	3.14	Tornado 24hr	
Availability of a hiding place	1.25	Bear 1yd	3.46	Whisper	
Immediacy	1.95	Party	4.92	Elevator	
Ability to communicate	1.00	Bomb; Hurricane; Tornado	4.53	Blackmail face	
Ability to mitigate	1.42	Hurricane 24hr	4.15	Political	
Ability to harm	1.00	Hurricane	3.89	Acquaintance	
Ability of others to help	1.60	Noise	3.58	Homophobic	

Table 2.3.5. Category Factor Ratings. Normalized factor ratings (Mean ± SE) by scenario category**.**

Factor	Natural (n=4)	Animal (n=5)	Physical (n=11)	Psychological (n=9)
Dangerousness	0.89 ± 0.05	0.84 ± 0.05	0.72 ± 0.04	0.26 ± 0.08
Escapability	0.45 ± 0.13	0.53 ± 0.14	0.69 ± 0.07	0.51 ± 0.07
Ambiguity	0.41 ± 0.04	0.42 ± 0.14	0.56 ± 0.08	0.28 ± 0.06
Distance	0.62 ± 0.16	0.51 ± 0.14	0.36 ± 0.08	0.60 ± 0.09
Availability of a hiding place	0.69 ± 0.08	0.19 ± 0.08	0.45 ± 0.11	0.28 ± 0.09
Immediacy	0.68 ± 0.10	0.76 ± 0.07	0.75 ± 0.05	0.38 ± 0.08
Ability to communicate	0.00 ± 0.00	0.04 ± 0.01	0.51 ± 0.08	0.84 ± 0.05
Ability to mitigate	0.22 ± 0.11	0.16 ± 0.04	0.53 ± 0.06	0.65 ± 0.07
Ability to harm	0.00 ± 0.00	0.23 ± 0.03	0.65 ± 0.06	0.47 ± 0.08
Ability of others to help	0.40 ± 0.11	0.22 ± 0.10	0.37 ± 0.05	0.57 ± 0.09

Specific situational factors elicit specific behaviors

Correlations between the mean factor ratings and the proportion of subjects endorsing defensive behaviors were calculated to determine whether human defensive behavior could be predicted by certain situational factors across categories of threat. All non-psychological scenarios (natural, animal, and human physical) were combined in the correlations (Figure. 2.3.2a), with results for psychological scenarios reported separately (Figure. 2.3.2b).

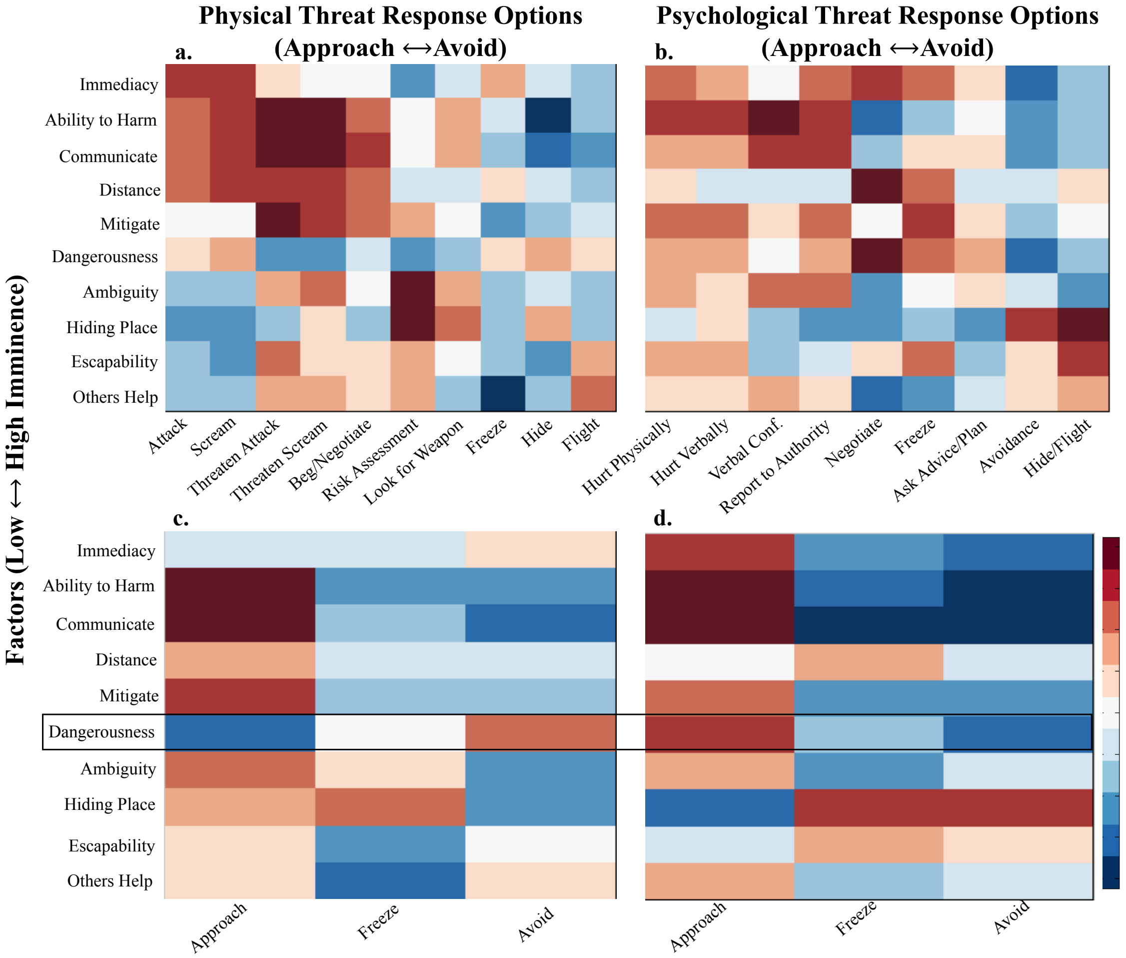


Figure 2.3.2. Factor-Response Option Correlations. Heatmap of correlation coefficients from Pearson's correlations between mean factor ratings and specific defensive behaviors (**a,b**) or approach/freeze/avoid (**c,d**) for physical (left) and psychological (right) threat scenarios. Row-wise factors organized along an approximate low to high imminence continuum. Column-wise response options organized along an approximate approach-avoidance continuum. Original distance scores were reversed to far-to-near to in accordance with the imminence continuum.

Many of the highest correlations were predictable: in the physical scenarios, the ability to communicate with the source of threat was strongly correlated with threatening to attack or threatening to scream; in the psychological scenarios, the presence of a hiding place was strongly correlated with hiding. A global pattern existed, such that imminent threats tended to be approached, and less imminent threats were avoided. For example, threats that were high in immediacy, dangerousness (all scenarios were dangerous, but some more so than others), ability to harm, or proximity were positively correlated with responses that required approaching the source of threat, like attacking, screaming, or threatening to do so. Similarly, threatening scenarios that were escapable, ambiguous, or had a hiding place available were negatively correlated with those approaching actions and positively correlated with avoidant actions including risk-assessment and hiding. This pattern was strongest for the most imminent threats, suggesting a more rigid and restricted set of response patterns to these scenarios.

Comparison of Psychological and Physical Threats with Direct Approach-Avoid Ratings

Direct comparisons of responses to psychological and physical threats as a function of factor ratings were made using approach-avoidance ratings. Correlations between the proportion of subjects choosing to approach, freeze, or avoid a threat and the factor rating of a threat elicited a similar pattern observed for specific responses, whereby more imminent threats were approached and less imminent threats avoided (Figure 2.3.2c,d). This pattern held for both physical and psychological threats, with a notable exception for the factor of dangerousness (the magnitude of the threat): dangerous physical threats were avoided, and dangerous psychological threats approached.

Decision Tree

A data-driven decision tree (Figure 2.3.3) that predicted a person's choice of defensive behavior for all scenarios summarized the most deterministic relationships between scenario factors and defensive behaviors. The construction of this tree was based entirely on data from the original sample (n=85), and was later tested on data from the replication sample (n=22). Partially supported from a clustering of scenarios based on patterns of factor ratings (Figure 2.3.4), the beginning of the tree separated psychological from physical threats.

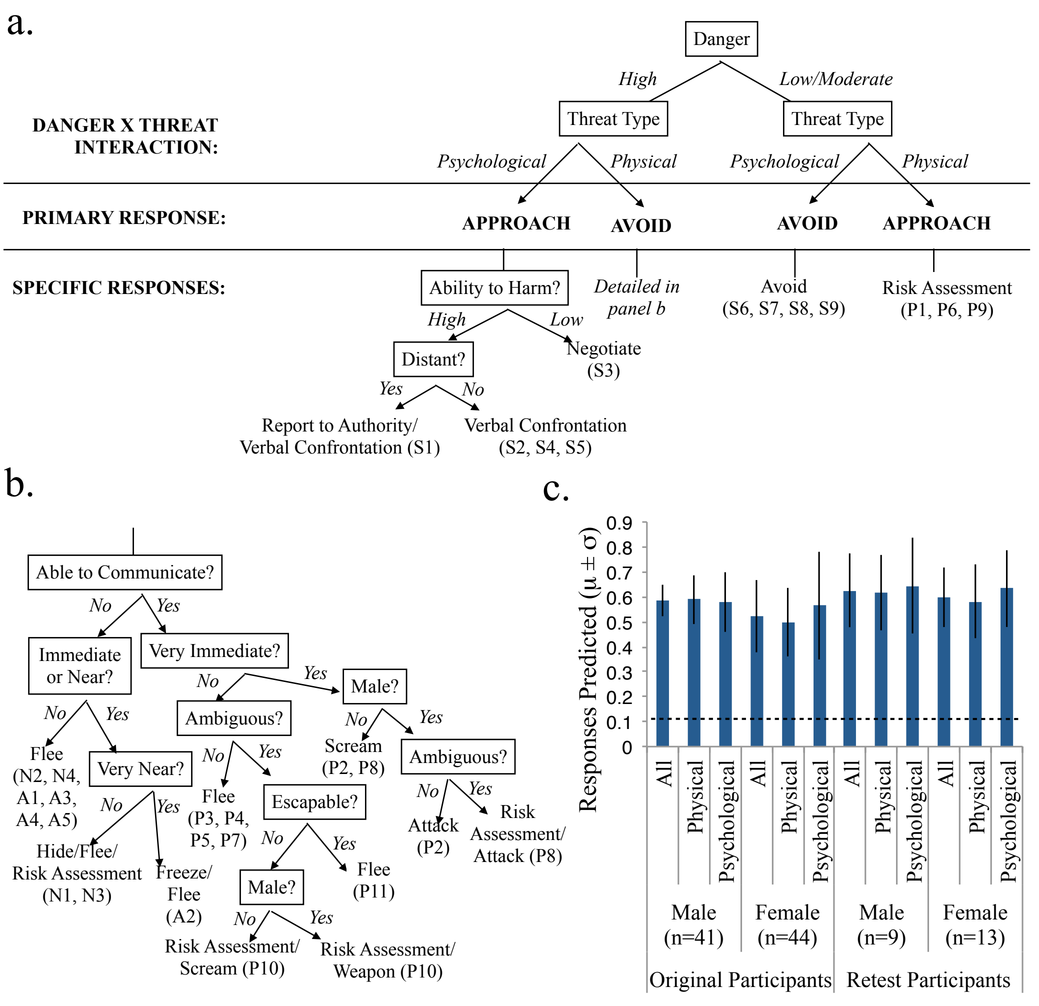


Figure 2.3.3. Decision Tree for Defensive Behaviors to Threatening Scenarios. **Panels a, b.** Decision tree predicting the defensive behavior chosen by the majority of participants based on characteristics of that threat scenario. The tree consists of four main branches, with primary approach/avoid responses predicted by the 2x2 interaction of danger (high, moderate) and threat type (psychological, physical threat). Appraisal of factors along further nodes predicts specific defensive responses for each scenario, denoted by the scenario labels used in Table 2.3.2. Where appropriate, gender differences are noted. The tree successfully predicts the group majority decisions of both original participants (n=85) and replication study participants (n=22) for all scenarios. **Panel c.** The average proportion of original and replication study participants' first responses correctly predicted for all scenarios (n=29), physical scenarios (n=20), and psychological scenarios (n=9). Male and female performance reported separately. Dashed line around 0.12 (All: 0.128; Physical: 0.130; Psychological: 0.123) represents chance performance.

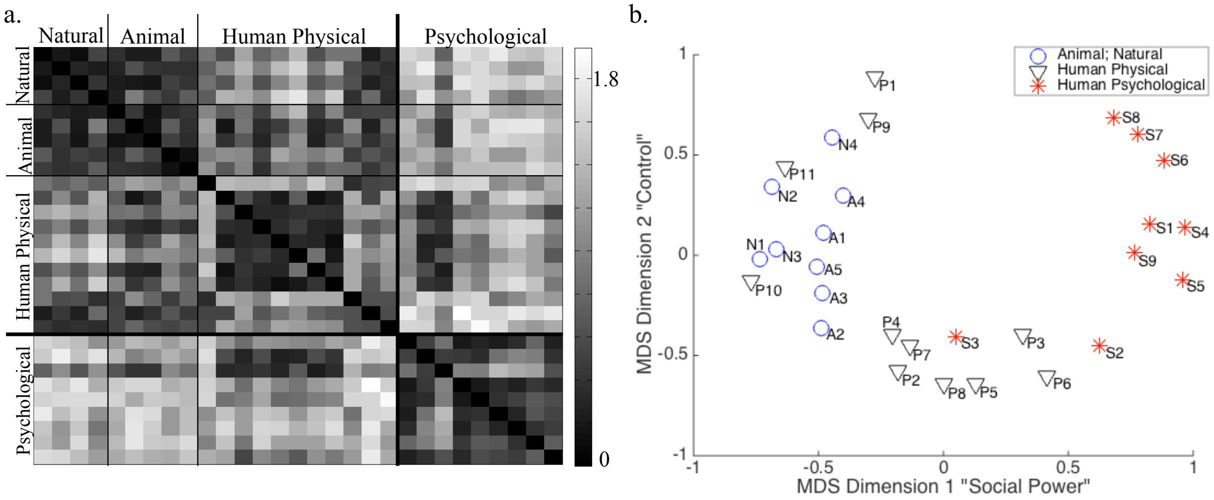


Figure 2.3.4. **Panel a.** Heatmap of scenario factor ratings dissimilarity matrix. Dissimilarity scores (legend right of heatmap) represent the distance between pairs of

scenarios, calculated as one minus the sample correlation between the ten factor ratings for each pairwise scenario comparison. Scenario labels indicated on top and left of heatmap, with individual scenarios denoted by each row/column (i.e., 4 “Natural” scenarios: N1, N2, N3, N4). Black diagonal indicates scenarios are minimally dissimilar to themselves; dark clusters indicate within-category scenarios are most similar according to factor ratings. Psychological scenarios are most distinct from the other categories. Within category similarity exceptions exist, e.g., P1, P9, P11 and S3. **Panel b.** Two-dimensional multidimensional scaling (MDS) of Euclidean distance between scenarios based on factor-rating dissimilarity scores. Human psychological threat scenarios (red stars) mostly clustered separately from physical threats (blue circles: animal and natural threats; black triangles: human physical). A scree plot of stress by MDS dimensions justified the use of 2 dimensions, which had a stress of 0.140. While the primary value of our MDS analysis is as a visualization of the similarity space of scenario factor ratings, we cautiously suggest that the first MDS dimension, positively related to the ability to mitigate (r=0.68) and to communicate (r=0.90) with the threat, while inversely related to dangerousness (r=-0.88) and immediacy (r=-0.63), relates to “social power” or the threatened individual’s ability to communicate with and influence the threat. Meanwhile, the second MDS dimension, inversely related to immediacy (r=-0.68), and positively related to distance (r=0.74) and the presence of a hiding place (r=0.75), captures the threatened individual’s ability to thwart the threat and “control” the situation.

The primary difference in reactions to psychological and physical threat is concisely summarized by differences in the interaction between dangerousness and basic approach/avoidance responses. Correlations between the single approach-avoidance scores and factor ratings for psychological dangerousness and physical dangerousness separately showed a significant positive correlation with psychological dangerousness (r=0.68, p=0.000) and a significant negative correlation with physical dangerousness (r=-0.57, p=0.001). In other words, dangerous psychological scenarios were approached, and dangerous physical scenarios were avoided. This pattern held for both females (psychological: r=0.50, p=0.006; physical: r=-0.52, p=0.004) and males (psychological: r=0.49, p=0.007; physical: r=-0.39, p=0.037).

The pattern of this interaction separated the four major branches of the decision tree: (1) high and (2) moderate danger physical threats, and (3) high and (4) moderate danger psychological threats. The primary response tendency for (1) and (3) was avoidance; types (2) and (4) were primarily approached.

The latter portion of the tree predicted specific responses. Traversing down the tree, an assessment was made at each node, related to a specific factor, to ultimately predict an action. The dominant response to low-threat psychological scenarios was simply avoidance; low-threat physical scenarios were predominately assessed for risk.

The path to specific responses for high danger scenarios was more complicated. The ability to communicate with the source of threat split human from non-human scenarios early in the tree. Animals and natural disasters, which could not be communicated with, and which were also physically difficult or impossible to fight, prompted flight, unless they were quite close, in which case risk assessment or hiding, and freezing (for especially imminent threats) were employed. For human physical threats, immediacy, ambiguity, escapability, and gender interacted to predict responses.

For psychological threats, verbal confrontation was a popular approach option, with negotiation employed in a scenario involving a boss where there was low ability to harm the source of threat, and reporting to an authority a popular option for more distant threats.

This decision tree correctly predicted the most popular response for all 29 scenarios in both the original and replication groups and performed much better than chance (about 12% since some scenarios allowed more than one valid response option) at predicting individual participants' responses for both the original and replication groups (Figure 2.2.3c).

Discussion

Main Findings

Correlations. Whilst using a new methodology (internet collection of scenario responses), we replicated prior findings that human reactions to human physical threats mirror patterns of defensive responses observed in rodents (Blanchard et al., 2001; A. M. Perkins & Corr, 2006; Shuhama et al., 2008) (Table 2.3.6). We extended this to show that similar patterns exist for defensive responses to non-human physical threats (natural disasters and animals), as well as social psychological threat, with some notable differences.

Table 2.3.6. Comparison of correlations coefficients between defensive behaviors and scenario characteristics obtained in 4 studies. Comparison between Blanchard (2001), Perkins and Corr (2006), and Shuhama (2008) reproduced from Blanchard (2011). The first 3 studies (left of line) used Blanchard's original 12 physically threatening scenarios and report male (top) and female (bottom) correlation values separately. The present

study's (right of line) 20 physical scenarios included 11 of the original physically threatening scenarios, along with 4 natural disaster and 5 animal scenarios. For the 9 psychological scenarios, comparable defensive response options are reported. V.C. = verbal confrontation. **p*<0.05, ***p*<0.01, n.s. = not significant with *p*<0.05; *p*-values not reported in Shuhama et al. (2008).

Defensive behavior/factor	Blanchard (Hawaii)	Perkins & Corr (Wales)	Shuhama (Brazil)	Original (USA)	Animal; Natural (USA)	Physical (USA)	Psychological (USA)
Risk assessment/ ambiguity	0.89** 0.86**	0.89** 0.85**	0.91 0.88	0.93***	0.08	0.62**	0.13 (plan)
Flight/ ambiguity	-0.50 -0.63	-0.56 -0.59*	-0.69 -0.61	-0.50	0.73*	-0.20	-0.48 (hide) -0.13 (avoid)
Defensive attack/ ambiguity	-0.53 -0.23	-0.54 -0.44	n.s. n.s.	-0.42	-0.50	-0.23	0.41 (V.C.)
Flight/ escapability	0.10 0.04	0.12 0.10	n.s. n.s.	0.35	0.81**	0.33	0.66 (hide) 0.23 (avoid)
Defensive attack/ escapability	-0.76* -0.65*	-0.87** -0.89**	-0.76 n.s.	-0.67*	-0.60*	-0.30	-0.38 (V.C.)
Defensive attack/distance	-0.59* -0.64*	-0.62* -0.69*	n.s. -0.69	-0.47*	-0.71*	-0.43	-0.72* (negotiate)
Hiding/hiding place	0.59* 0.63*	0.33 0.30	0.61 0.59	0.81**	0.44	0.29	0.84** (hide)

Decision Tree. Features of the threat scenarios determined behavioral responses; these patterns were summarized in a decision tree that successfully predicted scenario responses for the original participant group as well as generalized to a replication sample. This tree demonstrated two processes at play in threat reactions: (1) basic approach/avoidance behavior and (2) situational appraisal. The first set of processes distinguished psychological from physical threats based on a single factor: the magnitude of the threat; more dangerous physical threats were avoided while more dangerous psychological ones were approached. Subsequently, appraisal of further factors determined the best specific response for a particular threat scenario.

Imminence Framework. While our decision tree splits psychological from physical threat as a function of dangerousness, a general dimensional framework across both threat types emerges from the pattern of correlations between situational factors and favored defensive behaviors. Looking across the heatmap columns in Figure 2.3.2, behavioral responses can be organized along an approach-avoidance continuum, with freezing in the middle, while, going down the rows, situational factors exist on an imminence continuum. As a threatening situation becomes more imminent — immediate, close, and dangerous — attack responses are chosen; as the immediate threat wanes, avoidant behaviors, which are less costly to the organism, are adopted. Approaching actions (e.g., attack, negotiate) are only taken when an organism is pressed by imminent threat, with the exception of imminent but escapable threats, which are avoided. This structure mirrors a previously described pattern (Chang et al., 2013) whereby regardless of whether a predator or social conspecific posed a threat, imminent threats (e.g., a cat to a rodent or a dominant rhesus macaque to another rhesus macaque) evoked fast reflexive behaviors (escape/freeze/defensive aggression) while more distant threats (e.g., cat odor or a photo of a dominant rhesus macaque) are cautiously explored; this pattern of imminent threats evoking fast, reflexive responses was recently emphasized in the theoretical Survival Optimization System (SOS) model (Mobbs et al., 2015). Likewise, in conditioned fear paradigms, a conditioned stimulus is an imminent predictor of an aversive stimulus and elicits a prompt response. Our findings agree with the pattern of imminent responses eliciting rapid responses: across domains, imminent threats provoke fast, reflexive actions while more distal threats permit exploration; these generic patterns observed across specific threats likely reflect conserved adaptive mechanisms that evolved to cope with physical predator threat and that were subsequently co-opted also for coping with social/psychological threats.

The importance of threat imminence to explaining defensive behaviors in this two-dimensional manner is in line with prior work on predator imminence (Fanselow & Lester, 1988; McNaughton & Corr, 2004). Indeed, the brain is in fact sensitive to the literal distance to a threat: brain activity shifts from the ventromedial prefrontal cortex to the periaqueductal grey as the imminence of a virtual predator in an fMRI experiment is increased (Mobbs et al., 2007). It is even the case that activation in the amygdala discriminates the directionality of a threat — a tarantula — either towards or away from a subject, regardless of actual distance. This ability of the brain to monitor many dimensions of threat provides direct neurobiological evidence that we “fractionate” basic fear into component mechanisms (Mobbs et al., 2010). A major challenge for the future will be to map such neural components, as revealed with fMRI (Mobbs et al., 2010) or cellular techniques (Silva et al., 2013), onto the appraisal-like components we identified in our decision tree (Figure 2.3.3). While our stimuli here were designed to be concise for this behavioral experiment, it would be useful to design future stimuli that could also be used in fMRI tasks; such stimuli would need to control more stringently for a host of lexical and semantic confounds including length, word frequency, readability, concreteness, and arousal, all of which were not controlled for in our small sample of stimuli.

Interaction Caveat. An important caveat to all these dimensional analyses is that situational factors interact; therefore, it is important to be mindful of the entire context when assessing a behavior in any species. For example, it is the interaction between imminence and perceived magnitude of the danger that explains why imminent psychological threats are only reflexively approached (defensive aggression), while escape is a popular reflexive response for (escapable) imminent physical threats. Notably, while flight was a common behavioral response and was predicted well by our decision tree (cf. Figure 2.3.3b), it did not strongly correlate with any individual physical threat features (cf. Figure 2.3.2a), likely because of the interaction between factors. The concern about careful contextual analysis extends to comparative animal research: different tests can differentiate diverse anxiety phenotypes in non-human primate models for clinical comparisons to humans; however, these tests often do not correlate well with observed diagnostic behaviors, likely exhibiting context dependency which is not generalizable across tests (Coleman & Pierre, 2014).

Study Limitations. This study does not directly measure real threat behavior in humans, but rather ratings about hypothetical scenarios. Therefore, one could argue that the behavior captured here relates to an intuitive, culturally learned “folk” knowledge of how one ought to respond to a threatening situation. However, it is compelling that the patterns described here in humans from reactions to hypothetical threat scenarios in fact relate well to patterns observed in the actual behavior of rodents, as also seen in three other studies using the same methods (Blanchard et al., 2011). As such, it is our assumption that our data reflect the actual behavioral structure of threat. However, it is clear that future experiments should attempt to (1) observe humans' responses in actual threatening encounters in an observational (but non-experimental) context, (2) observe behavioral responses in an experimentally-controlled virtual-reality type of experiment in which participants "experience" a threat but are not placed in danger, and (3) record implicit measures including changes in autonomic arousal, body sway/freeze, and emotional expression in response to these stimuli.

Additionally, it should be noted that additional features could be added to our model, for example nodes determining whether both the potential attacker and threatened individual are aware of the threatening situation (Lima & Dill, 1990). All our scenarios involved an established threat situation, but future work should also incorporate evaluations antecedent to this point in the threat evaluation process.

What's Different for Non-Human Physical and Human Psychological Threats?

Non-Human Threats. In contrast to the results from human physical threat, in our correlation analysis, there was a weaker positive relationship between risk assessment and ambiguity for animals and natural disasters (cf. Table 2.3.6). Additionally, for these two physical threat types, the relationship between flight and ambiguity reversed from negative to positive, and the positive relationship between flight and escapability was stronger. In the decision tree, these threats were never approached**.** These changes are likely a function of the increased danger and decreased ability to communicate with or mitigate the source of threat in animal and natural scenarios as compared with human physical scenarios — in such situations escape is prioritized.

Psychological Threats. While similar to human and non-human physical threats in our dimensional analysis, psychological threats were the most distinctive threat category, requiring unique specific response options. Empirically supporting our hypothesis that while defensive reactions to all threat types draw on similar processes, psychological threats are qualitatively different. Factor ratings (Figure 2.3.4) and approach-danger tendencies distinguished psychological from physical threats. Specific differences for psychological threat arise from two main differences: the timescale of the threat and the specific type of harm inflicted.

Psychological and physical threats might be continuously mapped onto a dimension of temporal immediacy: psychological threat decouples the immediate physical threat from cues that signal it. On an interesting side note, while unconditioned fear stimuli are directly linked to physical threat, conditioned fear stimuli are separable from actual physical threat and therefore similar in nature to our psychological threats. Tautologically, in our study, imminent psychological threats were dangerous *because* they were imminent, could not be avoided, and required a rapid response. Typically, psychological threats and reactions to them unfold more slowly over time, allowing individuals to gather information and plan an optimal response, often drawing on the advice and help of others. In our most immediate/dangerous psychological scenarios, these options were not available.

In humans, there can then be yet another layer to psychological threat that is something like "symbolic threat", e.g., blackmail, where there is no physical threat at all, but instead relevance to factors such as social reputation. Other species show such “psychological" threat to some degree, related to social rank and social ostracism, one of the most potent social threats (Williams, 2007). In other species, these concerns relate more directly to physical concerns (access to food, protection, etc. with strong social bonds even increasing longevity in baboons (Silk et al., 2010)), while humans' worries about social reputation and social exclusion have less immediate physical ramifications (although admittedly, status relates to physical outcomes). Nevertheless, human fMRI studies suggest that aspects of social threat (social exclusion) activate regions that overlap with those activated by physical threats (physical pain) (Eisenberger, Lieberman, & Williams, 2003). The two types of pain share common pathways across several species; this evolutionary overlap has been attributed to physical pain mechanisms being used to prompt appropriate defensive reactions for social threats to inclusion (MacDonald & Leary, 2005). Indeed, some social psychological threats (angry faces) prime defensive bodily reactions, including freezing (Roelofs, Hagenaars, & Stins, 2010).

One notable exception to the observation that behavioral patterns to psychological threat mirror those to physical threat occurred in the case of ambiguous scenarios: defensive attack was not chosen in ambiguous physical scenarios, while verbal confrontation, which is analogous to attacking, was chosen in ambiguous psychological threat scenarios. In the psychological case, it seems that the cost of confrontation is not as high as in the physical case, where attack could likely result in bodily harm. Instead, the psychological form of attack — verbal confrontation — might even garner clarification of the ambiguous situation.

Future Directions

Three observations about our decision tree relate to future directions. Moving beyond Blanchard's (2001) correlational approach, the decision tree allowed us to recapture the complex interaction of situational factors in guiding threat responses.

Hierarchically organized, our decision tree emphasized (1) the importance of stimulus category (early branches separate psychological from physical threat, and then physical threat in which another human is or is not present); (2) that certain factors are more relevant and processed earlier (e.g., danger and immediacy are assessed early because highly imminent situations require immediate action); (3) appraisal occurs at each node, and must integrate information from the prior path traveled to reach that node. The length of a path traversed is relevant: throughout the tree, the general principle of attacking/actively responding to imminent risk and retreating/avoiding in less imminent situations holds constant across situational factors and categories of threat and fewer appraisal nodes are traversed for imminent than less-imminent threats. Each of these three observations relate to other findings and future directions.

Neural support. First, the importance of stimulus category raises the key question of what neural support exists for psychological theories. While common pathways have been discussed, and are evolutionarily efficient, some separate processing of social threat is supported by the finding that different hypothalamic circuits exist for predatory and social fear (Silva et al., 2013). A neural approach may also answer the open question of to what degree appraisals are automatic or controlled (deliberative); different appraisals may participate in different circuits, with varying degrees of automaticity (Pessoa & Adolphs, 2010).

Appraisal theory and relevant factors. Non-human primates appear capable of behavior similar to situational appraisal: woolly monkeys learned to react differently to three types of human intruders; learning when it is appropriate to launch a defensive response to humans' presence saves energy, leaving time for foraging (Papworth, Milner-Gulland, & Slocombe, 2013).

That certain factors have priority for appraisals can be related to the sequential nature of stimulus-evaluation checks postulated by some appraisal theories (Scherer, 1984, 2001). Interestingly, the enhancement of visual and olfactory sensory acquisition by fear expression (Susskind et al., 2008) relates to the cumulative nature of the appraisal process: being afraid involves gathering and assessing information about the source of threat. Appraisal theory identifies pertinent stimulus attributes. For instance, Scherer (2009) proposes four broad sets of such stimulus-evaluation checks that assess, in sequence: personal relevance of the stimulus; evaluation of how it affects well-being; coping potential; and normative significance. While our limited and in general psychologically simple set of scenarios was not designed to probe psychological appraisal theories, they share with such theories the need for a prioritized and integrative structure in how their threat is evaluated. Hierarchical assessment (cf. Tinbergen, 1951) and the contextual nature of that assessment are shared features of Scherer's component-process model and our decision tree. Appraisal theory is concerned with emotional states (Sander et al., 2005), which may elicit a behavior, but need not do so. Meanwhile, our decision tree focuses on the outcome of an emotional state, while remaining relatively agnostic about that state (neither we nor Scherer restrict these states to basic emotions). In the future, emotional evaluations/reactions as well as psychophysiological responses to different threatening scenarios should be empirically assessed. A second nuanced difference concerns the timescale considered in our model and appraisal theories: an "end point" is reached in our model when a first behavioral response is made. This occurs relatively early, especially compared to the complex psychological processes most appraisal theories describe. It is important to keep in mind that all of our data relies solely on descriptions of threat, and on verbal report of what people would do, which may further simplify and truncate the decision process.

Individual differences and psychopathology. Individual differences extending to impairment offer insight into the relationship between nodes. Each node in the tree will be given different weights, according to individual differences, including trait and personality differences (Perkins & Corr, 2006) and personal experience (Gawronski & Cesario, 2013), including prior exposure to or knowledge about "appropriate" responses to a specific scenario, e.g., the "correct" response to a hurricane. However, a node can also be broken. Psychiatric illnesses may be "linked to aberrant processing of environmental uncertainty" (Bach & Dolan, 2012) and amygdala lesions in rhesus macaques (Raper et al., 2013) affect contextual modulation to certain social threat cues, like eye gaze, and approach behaviors in humans (Kennedy et al., 2009). In healthy adults, there is individual variation in peripersonal space around the face according to variations in trait anxiety scores (Sambo & Iannetti, 2013), suggesting that individuals will respond variably to cues like distance in our decision tree model.

Anxiety disorders are of special interest for this model. It is known that individuals with anxiety attend to threats differently: a metanalysis (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & Van Ijzendoorn, 2007) showed that threat-related attentional bias is a robust phenomenon across many types of anxious individuals but not in non-anxious individuals. The type of threat that is overly attended seems to be affected by type of anxiety: individuals with panic disorder are sensitive to physical threats

(Asmundson, Sandler, Wilson, & Walker, 1992) while those with social anxiety are selectively sensitive to social threats (Asmundson & Stein, 1994; Goldin, Manber, Hakimi, Canli, & Gross, 2009). There is also individual variation in sensitivity to threat in other primates: a cognitive bias to social threat develops between 3 and 9 months of age in rhesus macaques, and is sensitive to the social rank of and protectiveness of their mothers, with infants of high status and more protective mothers being more vigilant towards social threats (Mandalaywala, Parker, & Maestripieri, 2014). On the other hand, oxytocin, a neuropeptide known to mediate pro-social behaviors, decreases social vigilance in adult male macaques (Ebitz, Watson, & Platt, 2013). Male and female rhesus macaques show differential response profiles to social threat, including greater high-risk aggression and gregariousness/boldness in males than females (Schwandt et al., 2010), mimicking gender differences observed in our decision three. In macaques, this gender effect interacted with the expression of the serotonin transporter-linked polymorphism (5-HTTLPR) and early environmental exposure to adversity in development (Schwandt et al., 2010). Together, these findings support future investigation of individual differences in threat assessment, including gender, environmental, personality, and other individual differences, which we hope to ultimate relate to neural and genotype differences.

Moving forward. Finally, it is worth returning to Darwin's belief that emotional behaviors could be classified across species. Although there are of course well known problems in anthropomorphizing the subjective feelings of emotions in nonhuman animals, characterizing the structure of context-dependent stimuli and emotional behaviors across species is a high priority in animal models of psychiatric illness, and methods for the behavioral phenotyping of rodents (Crawley, 1999, 2003) are well established. There are now a wealth of genetic and optogenetic manipulations in rodents that all inform mood and anxiety disorders in humans. To utilize the data from these animal models, it is essential to be able to map particular types of emotional behaviors from rodents to humans (Anderson & Adolphs, 2014; Holmes, Murphy, & Crawley, 2003), which may well require a shift towards focusing on both physiological and behavioral changes across species (Björkqvist, 2001). We would hope that characterizations such as the decision tree derived in the present study (Figure 2.3.3) could be developed for such comparisons, linking components of emotional behaviors and their possible pathology across species.

Acknowledgments

We dedicate this section to Robert J. Blanchard.

Chapter 3

FACE-TO-FACE SOCIAL COGNITION

In the first chapter, we proposed that distinct social cognition might occur when a person is interacting with, rather than observing a representation of another person, no matter how many perceptual contextual cues that representation might contain. In this chapter, we test whether social cognition changes in interaction. Specifically, we test whether neural processing of gaze, a potent social cue, is modulated by social interaction.

This last chapter complements the previous chapters in several respects. First, it uses fMRI rather than lesions to investigate neural function. As such, it takes advantage of a whole-brain field-of-view, allowing us to investigate several cortical networks in addition to the amygdala. Second, it extends our investigation of real-world social cognition to a psychiatric population that features profound difficulties with real-world social interaction: individuals with autism spectrum disorder.

The overlap with the preceding chapters in also apparent in three respects: (1) this chapter also investigates ecological validity, in this case focusing on the theme of interaction; and, (2) it also uses faces, and, potentially, threat signaled by faces (through sustained direct gaze). Also relevant to the previous chapter, (3) individuals with amygdala lesions do not use information from the eyes normally.

Unlike the previous chapter's three studies which represent completed, nearly-published work, the project featured in this final chapter is not published at the time of this thesis submission, and is intended to be so submitted over the coming summer.

Chapter 3.1

USING FACE-TO-FACE FUNCTIONAL MAGNETIC RESONANCE IMAGING TO INVESTIGATE THE SOCIAL BRAIN IN AUTISM

In 1943, Leo Kanner observed children with autism having "a far better relationship with pictures of people than with people themselves." While a large literature documents abnormal processing of faces and eye gaze in people with autism spectrum disorder, the vast majority of stimuli used in those studies are pictures or videos, rather than people in the flesh. Yet, as Kanner's quoted observation, our own experience, and the growing literature on interactive social cognition argue, pictures and real people are processed in quite different ways, which may be dissociatively impaired in autism. As autism is partially characterized by a persistent deficit in social interaction, it is reasonable to presume deficits, including in gaze, may be accentuated in actual interaction with another person. Findings from eye tracking studies indicate gaze behavior is influenced by the direct presence of another person. It is unknown whether direct live gaze is differently represented in the brains of individuals with autism.

This study comprised two main objectives. The first objective was to determine whether controls and individuals with autism are sensitive to the difference between live and recorded gaze, namely, whether the neural response to gaze changes as a function of the presence of another person. The second objective of the study was to characterize the differential neural response to live and recorded gaze as a function of autism severity.

We explored these aims with a novel fMRI paradigm that used a live person as a stimulus. Participants viewed a real actor sitting behind the scanner in a Live condition; this was contrasted with a Video condition.

Participants included 14 males with autism and 14 age, gender, and IQ-matched controls. In a Live condition, participants monitored the gaze of a live actor sitting behind the bore of the magnet. Audio instructions delivered to the actor's headphones cued three trial conditions: direct gaze, averted gaze, or eyes closed. In the Video condition, equivalent video recordings of the Live condition were used.

Distinct patterns of neural activation in putative Mirror Neuron System, Theory of Mind, face-processing, and control regions were seen in individuals with autism compared to controls for Live but not Recorded gaze. Activity in some of these areas correlated with autism severity, as well as performance on the Eyes in the Mind Task. Mirroring Kanner's original observation, these findings provide insight into the difficulty individuals with autism face in directly interacting with other people in everyday life.

Introduction

Scientifically, social neuroscientists are interested in whether interaction modulates social cognition (Kingstone, 2009; Schilbach et al., 2013; Schilbach et al., 2013). More than a basic scientific question, differences in reactivity to real, live social interaction and representational approximations thereof may have profound clinical importance. To that end, this study investigates the neural response to a Live and Video person in individuals with autism. We chose to investigate social gaze in this first Face-to-Face neuroimaging study. Gaze is a potent social cue that is both amenable to investigating in the scanning environment and is implicated in autism.

Autism Spectrum Disorder

Autism spectrum disorders (ASD) are developmental disorders partially characterized by atypical development of social and communication skills. In high-functioning adults with an ASD, these impairments can be debilitating in everyday life: they may have difficulty maintaining friendships and jobs, and difficulty understanding and communicating effectively with other people. Further, because of their potentially unusual or awkward social behaviors, high-functioning people with ASD often stand out, leading to discrimination and negative social interactions. Understanding how people with ASD actually process other people is important to understanding their difficulties, and therefore aiding therapy targeted at improving their social interaction capabilities. Nevertheless, because all neuroimaging studies to date have used photographs or videos of people as stimuli, rather than real people, our knowledge of their specific deficits in processing social interaction is limited. Admittedly, evidence is beginning to be accumulated in this area of research: more interactive, dynamic stimuli including eyetracking-controlled virtual characters or avatars (Wilms et al., 2010), a video-feed interaction with a live actor (Redcay et al., 2013; Redcay et al., 2010), and a real-time interaction with a "person" played by a computer (Tanabe et al., 2012) have all been used. However, to our knowledge, no neuroimaging study to date has examined the response to gaze from a real live person not mediated by a computer (let alone investigated this in ASD). Thus, the current study is highly novel; as such our approach and analyses are admittedly relatively exploratory at this stage. While our approach is novel, there is substantial rationale for hypothesizing differences in how the brain processes gaze from a real person as compared to a video of them: actual eye-contact feels qualitatively very different from a photograph or video; we look at real people differently during interactions, increasing the amount of reciprocated direct gaze during face-to-face interactions (cf. Figure 3.1.1); and skin conductance responses are greater for live than recorded gaze stimuli (Hietanen, Leppänen, Peltola, Linna-aho, & Ruuhiala, 2008).

Gaze and Autism. Gaze is a highly salient social cue and is differently processed in autism (Pitskel et al., 2011). Many developmental disorders exhibit prominently abnormal levels of eye contact, including fragile X syndrome (Farzin, Rivera, & Hessl, 2009), Williams Syndrome (Riby & Hancock, 2008), and, most notably, ASD. Both clinical and anecdotal evidence suggests that reduced or absent eye contact is a pervasive feature of ASD (Kanner, 1943; Lord et al., 2000; Lord, Rutter, & Le Couteur, 1994). Clinically, impaired eye gaze is part of the set of impairments by which ASD is diagnosed (Lord et al., 2000; Lord et al., 1994; Wetherby et al., 2004).

Several prominent studies have documented abnormal eye gaze to static images of faces or to films in people with ASD (Kliemann, Dziobek, Hatri, Baudewig, & Heekeren, 2012; Kliemann, Dziobek, Hatri, Steimke, & Heekeren, 2010; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Pelphrey et al., 2002). It has also been shown that there is a failure to make use of information from the eye region of faces in order to make social judgments about faces, in both ASD (Spezio, Adolphs, Hurley, & Piven, 2007) and in parents of people with autism who are characterized as having a broad autism phenotype (Adolphs, Spezio, Parlier, & Piven, 2008). One recent hypothesis has been that perceiving the direct gaze of another person triggers exaggerated emotional responses in people with autism, possibly mediated through a neural system for gaze processing that includes the amygdala (Dalton et al., 2005). While neuroimaging studies have examined the neural correlates of gaze perception (Carlin, Calder, Kriegeskorte, Nili, & Rowe, 2011; Nummenmaa, Passamonti, Rowe, Engell, & Calder, 2009; Pelphrey, Viola, & McCarthy, 2004), as well as abnormalities in gaze processing in ASD (Dalton et al., 2005; Kliemann et al., 2012; Pelphrey, Morris, & McCarthy, 2005), none have compared this with gaze from a real person, i.e., the type of stimulus faced in the real world.

Face-to-Face Eyetracking in ASD. Experimental evidence of gaze behavior during a real conversation indicates that individuals with ASD do not appropriately socially modulate their gaze behavior according to the presence of another individual (Figure 3.1.1) (Harrison, Spezio, Tyszka, Elison, & Adolphs, 2014). In that experiment, we had individuals with autism (*N*=10, 2 female, age=27±4) and controls matched for age and approximate IQ (*N*=27, 11 female, age=28±2) conduct a conversation with an actor in two conditions: Face-to-Face, and Live Video. Data from that experiment supports the assertion that gaze behavior, and presumably gaze processing, are differently influenced in individuals with ASD and controls by social presence (not just interaction, as has been manipulated in other fMRI studies).

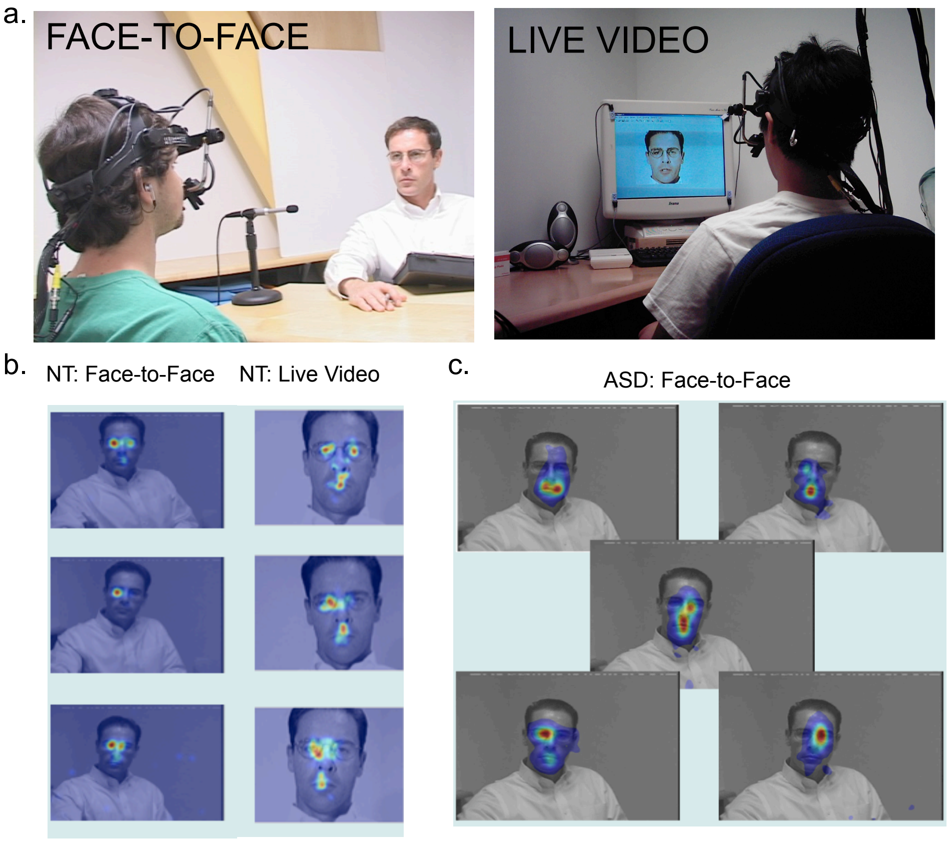


Figure 3.1.1. Face-to-Face Eyetracking in ASD. Wearing a head-mounted eyetracker (Eyelink II, SR Research), participants took part in a live conversation with an actor in two conditions: (a) a Face-to-Face condition in which they sat directly across from the actor, and a Live Video condition in which the actor sat in another room, and the conversation was mediated by a live video feed that allowed the participant to hear and see the actor in real time, but the actor could only hear the participant, removing any necessity for the participant to appropriately socially-modulate their gaze toward the actor. This condition manipulation affected gaze performance in controls, with heatmaps of gaze behavior in three sample participants (b) showing increased gaze to the mouth in the Live Video condition (n.b., only time points when the actor was not speaking were analyzed to remove lip-reading confounds). ASD participants did not appropriately modulate their gaze by social condition, (c) showing an abnormally high amount of gaze to the mouth in the Face-to-Face condition. Quantitatively, control participants showed significant mouth gaze in the Live Video compared to Face-to-Face condition ($\chi^2(1, N=26)=9.02, p<0.01$). Research reported from (Harrison et al., 2014).

Neural Systems for Gaze Processing

Gaze is a core social cue in primates and essential for human communication (Emery, 2000). Our ability to use gaze to infer social attention plays a crucial role in social interactions in humans and other primates, starting from infancy (Emery, 2000; Farroni, Csibra, Simion, & Johnson, 2002; Johnson, 2005; Keating & Keating, 1982).

Based on existing literature from human fMRI and primate single-unit recordings, social gaze is primarily processed by three neural systems. The first is a perceptual network that includes the inferior parietal lobule (IPL, also implicated in the putative Mirror Neuron System), the fusiform gyrus, and a head-view invariant region in the anterior superior temporal sulcus (aSTS, also implicated as a theory of mind (TOM) region) (Calder et al., 2007; Calder, Jenkins, Cassel, & Clifford, 2008; Carlin et al., 2011). Together, these regions rapidly detect dynamic gaze shifts and gaze direction (Conty, N'Diaye, Tijus, & George, 2007; Sato, Kochiyama, Uono, & Yoshikawa, 2008). The second system is a system for associating perceived gaze with its social significance, comprised of the right posterior superior temporal sulcus (pSTS, also implicated as a TOM region), right anterior insula (Ethofer, Gschwind, & Vuilleumier, 2011), and amygdala (Adolphs et al., 2005). The third system is a system for inferring mental states and social attention of another person from their gaze, comprised of the anterior insula (bordering the right inferior frontal gyrus, a putative Mirror Neuron System region), as well as TOM-associated regions, including the temporal parietal junction (TPJ), medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), and temporal poles (Amodio & Frith, 2006; Frith, 2007; Frith & Frith, 2007; Saxe, 2006; Spunt, Satpute, & Lieberman, 2011). These three systems correspond to three components of social gaze processing: gaze detection, extracting social significance/value, and higher-order mentalizing.

In addition to these systems involved in face and gaze processing, which include regions involved in social processing (putative Mirror Neuron System and Theory of Mind regions), in a live gaze monitoring study, we would expect recruitment of brain areas involved in cognitive control and engagement, such as the anterior cingulate cortex (involved in the Stroop task, for example (Pardo, Pardo, Janer, & Raichle, 1990)), as well as the precuneus, a “social” region implicated as a default mode network hub (Utevsky, Smith, & Huettel, 2014).

Experimental Aims and Hypotheses

In this novel, and admittedly somewhat exploratory, experiment, I test three main hypotheses: (1) that across gaze directions (Direct, Averted, Closed), the brain responds differently to the presence of a real person than to videos of them, (2) that there are differential neural responses to direct than to averted gaze (joint versus non-joint gaze); and, most importantly, (3) that the differences between people with ASD and matched controls will be most pronounced for the real person stimuli than for the videos. The precise pattern of results will give us insight into what cognitive processes that are engaged when we have a real person looking at us, and which of these is abnormal in ASD. The first aim of the study is to address these three hypotheses concerning neural differences in response gaze from a live and recorded person. The second exploratory aim of this study is to explore whether individual differences, such as autism severity, correlate with the neural response to face-to-face gaze.

To achieve these aims, in a blocked-design fMRI experiment, we ask individuals with autism and matched controls to observe gaze shifts in an actor, both in a Live, face-to-face condition and a visually matched recorded Video condition.

Materials and Methods

Participants

30 individuals participated in our experiment: 15 high-functioning adult males with autism spectrum disorder (ASD) diagnoses, and 15 typically-developed controls matched for age, gender, and IQ (Table 3.1. 1).

All participants with autism were diagnosed according to DSM-IV criteria in a clinical psychologist's interview and administration of the ADOS.

One individual was excluded from the ASD group for an incidental finding in his anatomical MRI scan; one individual was excluded from the control group for markedly anxious behavior during both this experiment and another laboratory member's experiment.

All participants had normal or corrected-to-normal vision. 13 of the 14 included participants in each group were right handed. All participants provided informed consent to participate in research under a protocol approved by the Caltech Institutional Review Board.

Table 3.1.1. Characterization of Participants. Group means and standard deviations reported.

Group	*N*	Gender	Age (Years)	FSIQ	AQ
ASD	14	Male	33±12	111±13	29.8±7.6
Controls	14	Male	29±9	112±9	14.5±3.8

Experimental Set-Up

Participants completed a gaze-monitoring task while exposed to two experimental conditions: (1) Live and (2) Video gaze (Figure 3.1.2). In the Live condition, participants monitored the actor's gaze through a mirror while lying in the scanner. The actor sat behind the bore of the scanner. A large white screen made of white fiberboard and supported by a PVC pipe structure was placed behind the actor to provide a clean and consistent visual background. The actor wore headphones, through which auditory commands from Matlab were relayed, instructing the actor how to direct their gaze on each trial. The actor's eyes were either closed, directed forward, or averted left or right (cf. Figure 3.1.3). In the averted gaze trials, the actor used specific fixation points in the scenery to ensure consistent gaze behavior across trials and participants. Prior to the participant's entering the scanner, they were introduced to both actors to further impress upon the participants that the actors were indeed real; participants also watched the actors seat themselves behind the bore of the magnet. In the Video condition, participants once again monitored the same actor's gaze, but, instead of viewing a Live person behind the bore of the magnet, a video recording of the same actor was projected onto a display screen at the end of the bore. In the Video condition, trial type was controlled Matlab's Psychophysics toolbox.

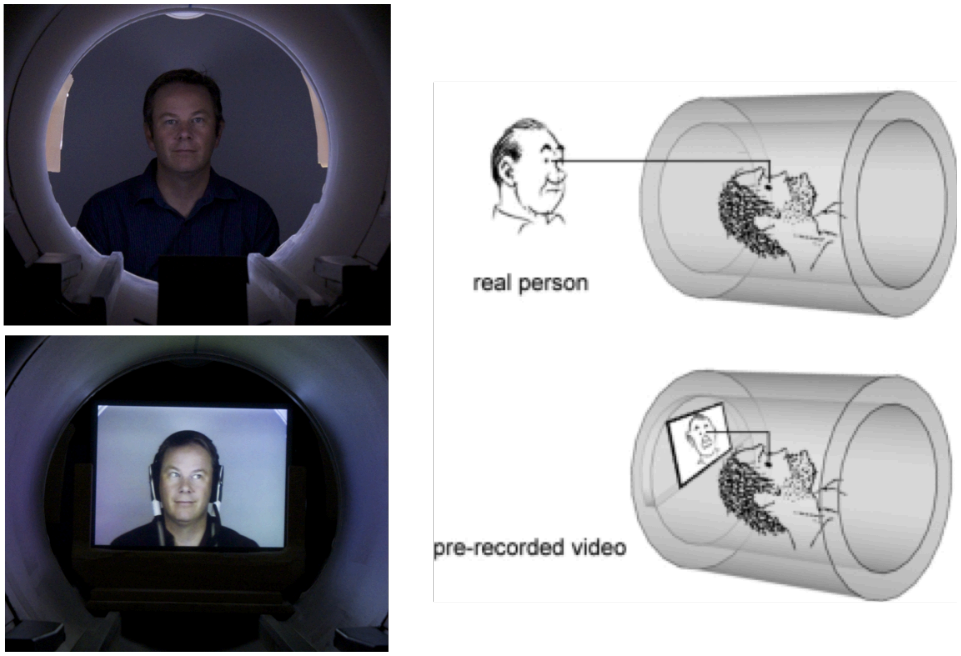
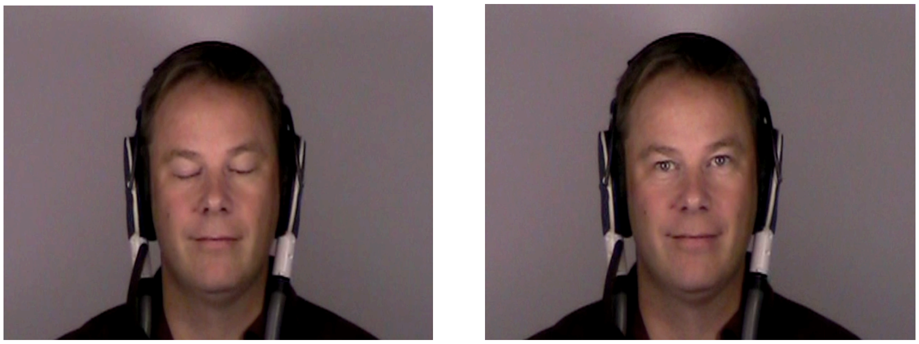


Figure 3.1.2. Live and Video Experimental Conditions. Participants monitored the same actor's gaze in both a Live (top) and Video (bottom) condition.

Experimental Design

Overall, we employed a 2(Subjects: NT, ASD) x 2(Condition: Live, Video) x 3(Gaze: Direct, Averted, Closed) design. Within each condition (Live, Video), two actors were used; results across them were combined for each subject in a mixed effects model (see fMRI analysis methods below).

Each participant completed 4 sessions in the scanner (2 actors x 2 conditions). Both Live sessions were completed sequentially, and both Recorded sessions were completed sequentially, with condition and actor order counterbalanced across participants. Within each session, participants were instructed to look at the actor, monitor their gaze, and press a button to indicate when the actor changed their gaze position. Each session consisted of 36 gaze trials, with a third of trials devoted to each gaze condition (Figure 3.1.3). The presentation order of the 36 trials was randomized for each subject. Each trial lasted approximately 10 seconds. Trial durations were randomly assigned and were drawn from a flat distribution on the interval between 8 and 12 seconds. Because the live actor could not be temporarily removed from the field of view, no interstimulus interval was used (Figure 3.1.4).



Closed

Direct gaze

Averted gaze

Figure 3.1.3. Gaze Trial Conditions. The actor's eyes were closed in a third of trials, directed forward in a third of trials, and averted (half left; half right) in a third of trials.

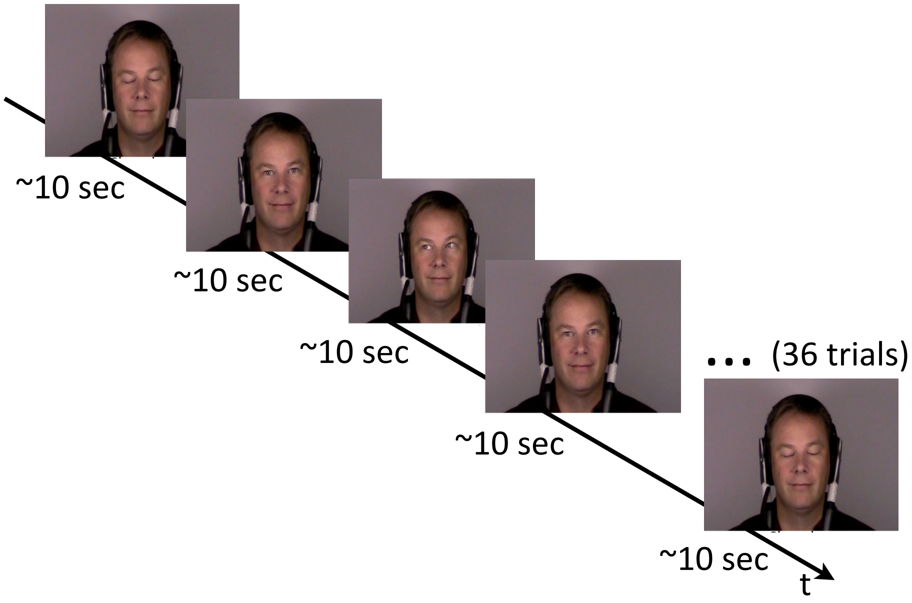


Figure 3.1.4. Session Design. Each of the four sessions completed by the participants (2 actors x 2 conditions, Live and Video) consisted of 36 trials continuous trials of between 8 and 12 seconds.

Data Acquisition

Imaging data were collected in a 3.0 Tesla Siemens TIM Trio scanner using a 32-channel head coil. Gradient echo, T2*-weighted EPI images with BOLD contrast were used to collect functional imaging data. Slices were acquired in an oblique 30-degree axial to coronal tilt orientation relative to the main magnetic field to minimize signal dropout in ventral frontal regions. The imaging parameters were as follows: TR=3000ms; TE=30ms; flip angle=90°; FOV=200x200 mm^2. EPI voxels were 2.5 mm isotropic in-plane, with 54 2.5mm slices being acquired per volume. While in the scanner, participants completed four sessions of the gaze-monitoring task, with 150 volumes acquired per session. For use in geometric correction of the 4 EPI volumes, a minimum of two B_0 fieldmaps with the same geometry as the EPI data were collected for each participant, with additional fieldmaps being collected either when a subject moved or as time permitted. Additionally, one to two high-resolution, whole-brain T1-weighted structural images (TR=1500ms; TE=2.9ms; flip angle=10°; 1 mm isotropic voxel resolution) was collected for each participant.

Eye Tracking

To ensure task compliance, participants' eye movements were tracked throughout the experiment. A small video camera was fixed to the head coil, and recorded pupil movements were analyzed off-line. Prior to the first session, participants completed a nine-point calibration routine (cf. Figure 3.1.5), which was repeated both after subject movement and prior to subsequent sessions as time permitted.

Analysis

fMRI Pre-Processing. All fMRI analyses were conducted with FEAT, part of FSL (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). Pre-processing included standard motion correction and dvars motion scrubbing (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012); fieldmap unwarping of EPI data; slice timing correction; ICA cleanup using MELODIC and an in-house script that removed likely physiological artifacts based on power spectrum (Kelly et al., 2010) as well as removed any temporal fluctuations in the expected region of a Nyquist ghost artifact; MNI registration; 6mm spatial smoothing and co-registration of EPI images to T1-weighted structural scans. When two T1 anatomical scans were collected for a participant, they were averaged prior to co-registration. In group-level analyses, individual subjects' T1 anatomical scans were co-registered to an MNI template brain, to which all subjects' EPI data was co-registered to permit anatomical localization of functional activations at the group level.

fMRI Analysis. Statistical models for each session were based on standard univariate analyses that explored contrasting activation in response to the various gaze conditions. Higher-level analyses were carried out with FLAME in FSL. In all models, sessions across actors within a participant were combined using a fixed effects model before being passed to a mixed effects group analysis model. We ran several models; our hierarchy of models was as follows:

Two broad classes of models were tested. (1) To test main effects of group x condition, a finite impulse response (FIR) plus a temporal lag of 5 seconds was modeled after each shift in the actor's gaze. (2) For comparisons of gaze conditions, boxcar functions for each gaze epoch (Direct/Closed/Averted) were used. In both classes of models, a standard double-gamma plus temporal derivative hemodynamic response function was convolved with each event-related impulse or boxcar function.

For both classes of models, main effects (across both groups and conditions), effects between groups, between condition, and the interaction of group and condition were tested. In the first class of models (FIR), these comparisons were made across all gaze shifts.

In the boxcar class of models, these comparisons were made for all possible contrasts of gaze condition (although only individual Direct/Averted/Closed and Direct>Averted & Closed and Direct>Averted gaze contrasts were investigated).

Finally, all models were tested in the full sample of participants (*n*=14 in each group), as well as a subset of participants (*n*=10 in each group) who self-reported feeling a sense of interaction in the Live condition.

Region of Interest (ROI) Analysis. Group (ASD, NT) x Condition (Live, Video) comparisons of individuals' responses to each gaze condition (Direct, Closed, Averted) as well as contrasts of Direct>Averted gaze (isolating Joint gaze) were performed in independently-defined ROIs related to the so-called Mirror Neuron System, Theory of Mind network, face-processing regions, and regions related to engagement and control (cf. Table 3.1. 2).

Eye Tracking Data. Eye tracking video files were analyzed offline with MR Gaze software (Tyszka, 2015) (Figure 3.1.5). Heatmaps of the fixation locations across all sessions for each group were spatially correlated to ensure that both groups similarly complied with instructions to monitor the actors' gaze.

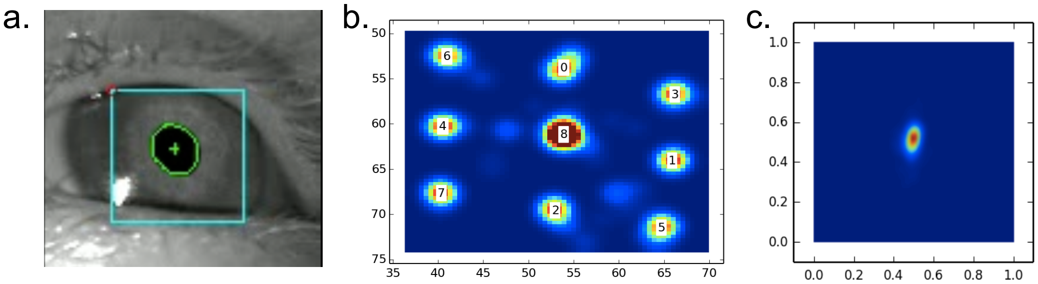


Figure 3.1.5. Eye Tracking. MR Gaze software (Tyszka, 2015) identified the participant's pupil (a) to identify (x,y) fixation coordinates. These coordinates from eyetracking during the functional scans were calibrated with against coordinates generated during a pre-session nine-point calibration routine (b). In the sample calibration grid (b), fixation point number seven would become (0,0) in calibrated (x,y) coordinates. Panel (c) shows a heatmap of sample fixation points from one subject in this calibrated space.

Results

Quality Assurance

To both check the general quality of our data and to ensure that group differences were not reducible to experimental confounds, the following three quality assurances were performed.

Motion. A concern in neuroimaging research involving ASD participants is that group level differences (especially in functional connectivity analyses) maybe artifactually driven by group motion differences (Deen & Pelphrey, 2012). While the effects of motion are less of a concern for our analyses, we took steps to mitigate its effects. First, we checked for group differences in motion. While there were no group differences in the average change in rotational motion across each session (*p*=0.15), there was a group difference in the average change in translational displacement over each session (*p*=0.022) (Figure 3.1.6). We corrected for this group differences in displacement by using dvars to introduce a motion scrubbing variable (an approach recommended by (Power et al., 2012)) to the nusiance regressor files incorporated in our FEAT model.

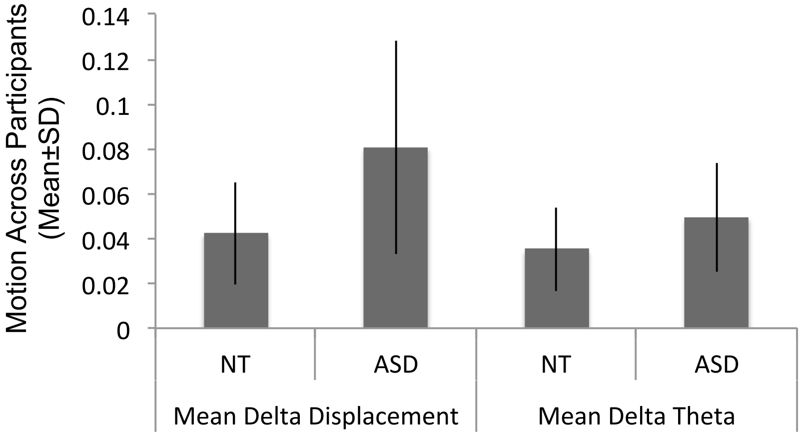


Figure 3.1.6. Group Differences in Motion. Group differences in individual members' mean change in translational displacement (left) and in rotational motion (right) across sessions showed that while the NT and ASD groups had similar amounts of rotational motion (measured in millidegrees), the ASD group had more change in translational displacement (measured in mm).

Behavior and Eye Tracking. Due to a technical error, reliable button press data was not collected. The purpose of the button press was to provide participants with a simple attentional monitoring task. The experimenters can anecdotally confirm that they observed all participants making button presses throughout each session and that participants seemed to perform at ceiling (i.e., they could and did detect changes in the actors' gaze position). Additionally, based on visual inspection of the eye tracking videos, all participants were awake and alert throughout the duration of the experiment.

In a post-task questionnaire, participants were asked to indicate whether they felt differently in one condition than the other. Ten of fourteen participants in each group spontaneously reported feelings of interaction (i.e., that the felt the presence of the actor) in the Live but not the Video condition. In some subsequent analyses, this subset of participants who seemed particularly sensitive to the condition manipulation ("Felt Interaction or FI Subset") was analyzed separately.

In preliminary analysis of eyetracking data from seven participants in each group (Figure 3.1.7), participants from both groups similarly complied with the task to monitor the actors' eye region throughout the duration of the task (spatial correlation between two groups heatmaps: r=0.94).

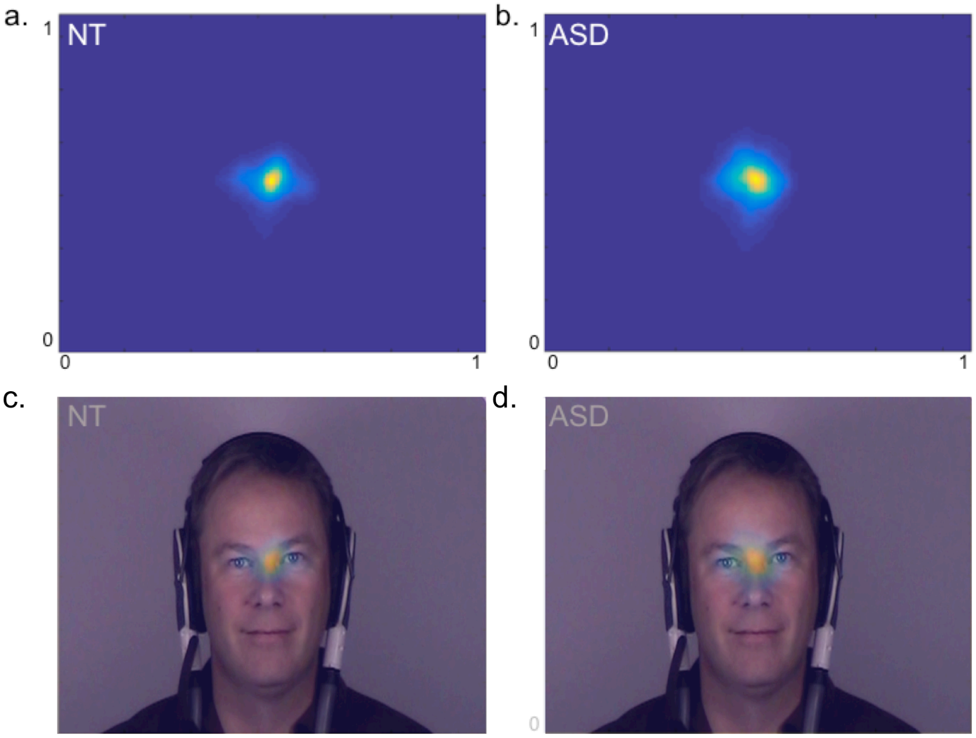


Figure 3.1.7. Group Eye Tracking Heatmaps. Preliminary eye tracking analysis (*n*=7 participants per group) of gaze data from (a) NT and (b) ASD individuals shows that both groups tended to fixate a small area throughout the duration of the task. For illustrative purposes those heatmaps for NT (c) and ASD (d) subjects are overlaid on a frame from the Video condition. In both, the focus of fixation was in a parafovial region that included the eyes.

Pre-processing check: "Button-press" response. To check that our preprocessing routine affected both groups similarly, we ran a model designed to detect motor cortex activation following the participants' presumed button-press activity. A finite-response function was modeled for each change in the actor's gaze position, plus a lag of 5 seconds. In both groups, right-handed participants (n=13 in each group) showed robust left motor cortex activity following time points when they should have pressed a button, supporting both (1) our pre-processing routine and (2) the validity of the assumption that both groups completed the task similarly.

Whole Brain Analyses

Main effects. Before analyzing group (NT, ASD) x condition (Live, Video) differences in response to specific gaze conditions (e.g., Direct versus Averted gaze), we first looked at the main effects of group and condition. This analysis focused on the response to any gaze change regardless of gaze type, and, as in the "button-press" analysis described above, modeled a finite impulse response 5 seconds after each gaze shift made by either the Live or Video actor. Unless otherwise stated, all analyses in this section report whole-brain thresholding of $z>1.96$ (corresponding to $p<0.05$) and cluster-level thresholding at $p=0.05$ (indicating a 5% chance that a cluster of that size would be formed by chance).

The main effect of condition across participant groups showed greater activation in visual and gaze processing regions (fusiform, occipital cortex, and temporal cortex) in the Video than Live condition (Figure 3.1.8). Meanwhile, no voxels survived the Live greater than Video comparison, hypothetically due to greater heterogeneity of response across the two participant groups in the Live condition, which future analyses will investigate.

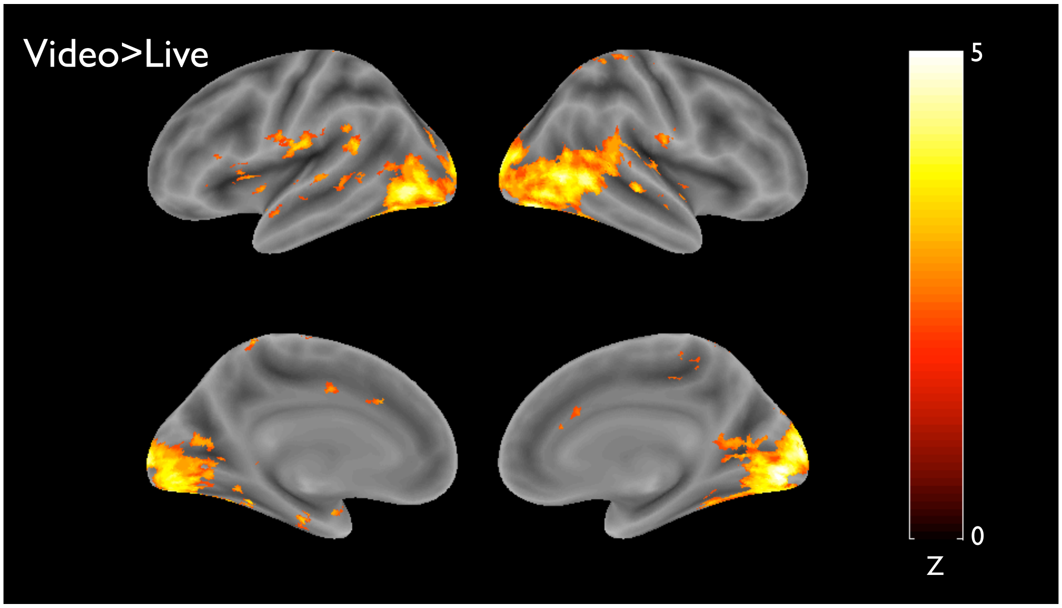


Figure 3.1.8. Main Effect of Video>Live Condition. Across all participants (*N*=28) and gaze types (Direct, Closed, Averted), the main effect of Video compared to Live gaze after each gaze shift (FIR at time of shift plus 5 seconds). Whole-brain thresholding of z>1.96, with cluster-level thresholding at *p*=0.05.

The main effect of group across conditions showed activation in the right inferior frontal gyrus, extending into the insula in the NT group (Figure 3.1.9). Meanwhile, ASD participants had more activation in a small region of the medial prefrontal cortex, as well as a large area in the posterior cerebellum.

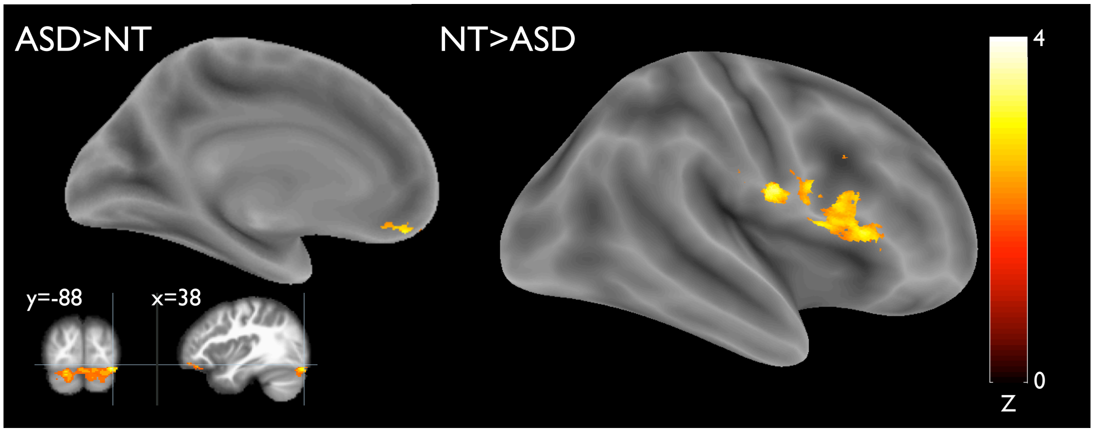


Figure 3.1.9. Main Effect of Group after Gaze Shift. Across gaze types (Direct, Closed, Averted), the main effect of group (left: ASD>NT, right: NT>ASD) after each gaze shift (FIR at time of shift plus 5 seconds). Whole-brain thresholding of $z>1.96$, with cluster-level thresholding at $p=0.05$. $N=14$ in each group.

Looking at the interaction of group and condition, in the Live condition, no voxels survived the NT greater than ASD contrast. However, there was greater medial PFC activation in the ASD than NT group in the Live condition in a nearly identical region to that shown in Figure 3.1.9. In the Video condition, voxels in the right IFG, and temporal lobe regions survived the NT>ASD contrast, while a posterior region of the cerebellar survived the ASD>NT contrast.

Effects of gaze condition. After investigating main effects of group and condition, a boxcar-design analysis was used to assess responses to specific gaze conditions. The average response to each gaze condition (Closed, Averted, and Direct) across subjects and conditions elicited strong visual cortex activity, with Direct and Averted gaze also activating temporal and frontal regions (Figure 3.1.10).

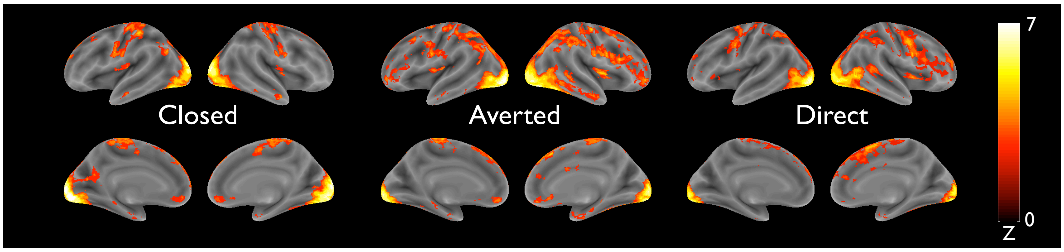


Figure 3.1.10. Average Response to Closed, Averted, and Direct Gaze. Across all participants ($N=28$) and conditions (Live, Closed), the main effect of each gaze type (left to right: Closed, Averted, Direct), modeled in a blocked design covering the duration of

each gaze trial. Whole-brain thresholding of z>1.96, with cluster-level thresholding at *p*=0.05.

Direct gaze was a special condition in our experiment – when the actor looked directly ahead, they formed joint gaze with the participant. Focusing on isolating the effects of Direct gaze, the contrast of Direct greater than Averted and Closed gaze (Figure 3.1.11) shows robust activation in a region of the inferior frontal gyrus (IFG), pars triangularis extending into the pars opercularis (Broca's area) and the anterior insula; the superior temporal sulcus; and the inferior parietal lobule; medially, the supplementary motor area, extending into the anterior cingulate was also activated. Subcortically, the right thalamus was activated in areas known to have structural connectivity with the prefrontal and temporal cortices (Figure 3.1.12).

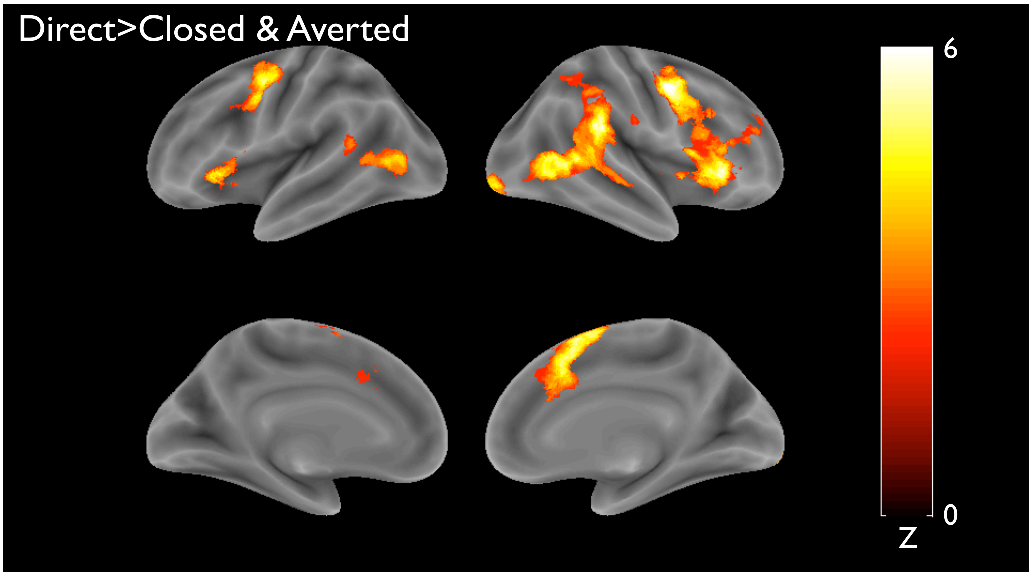


Figure 3.1.11. Main Effect of Specific Direct Gaze. Across all participants (*N*=28) and conditions (Live, Closed), the main effect of Direct greater than Closed and Averted

gaze, modeled in a blocked design covering the duration of each gaze trial. Whole-brain thresholding of z>1.96, with cluster-level thresholding at *p*=0.05.

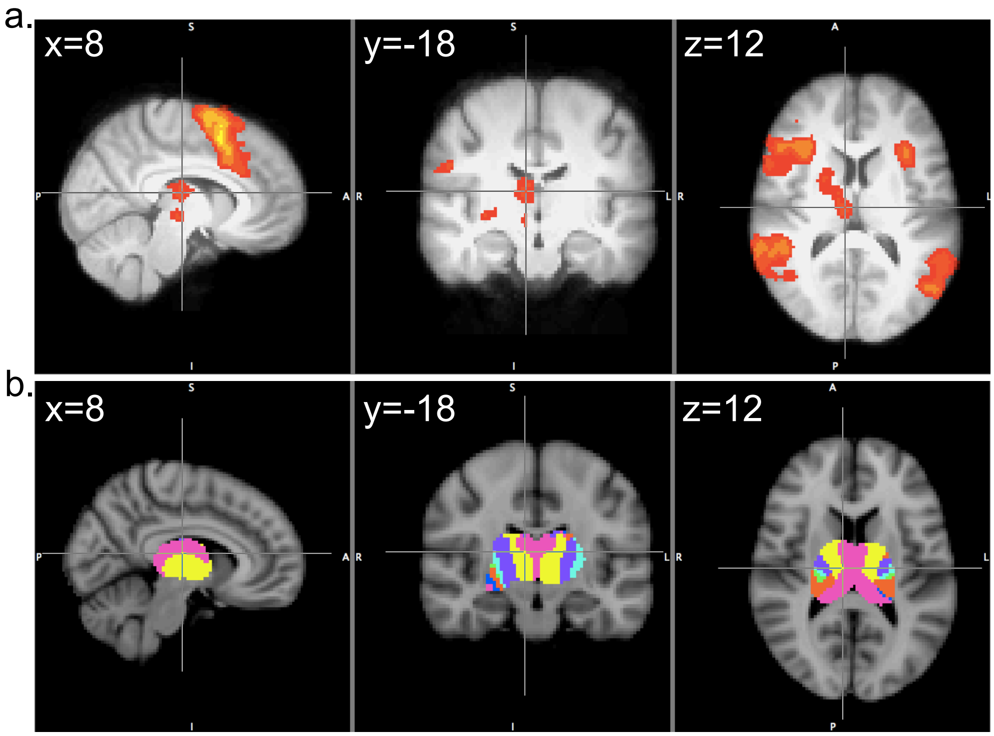


Figure 3.1.12. Thalamic Activation to Specific Direct Gaze. (a) Subcortical activation in the right thalamus from the same Direct>Averted and Closed contrast described in Figure 3.1.13. Thalamic activation overlapped with portions of the thalamus (b) described by the Oxford Thalamic Connectivity Probability atlas as being structurally connected with pre-frontal (yellow) and temporal (pink) cortices.

A contrast of Direct>Averted gaze in the Live condition in the subset of participants who were most sensitive to the experimental manipulation (FI participants) qualitatively indicates that many of the regions specific to Direct gaze in the main effect contrast (Figure 3.1.11) were modulated by group (NT>ASD) in the Live condition (Figure 3.1.13).

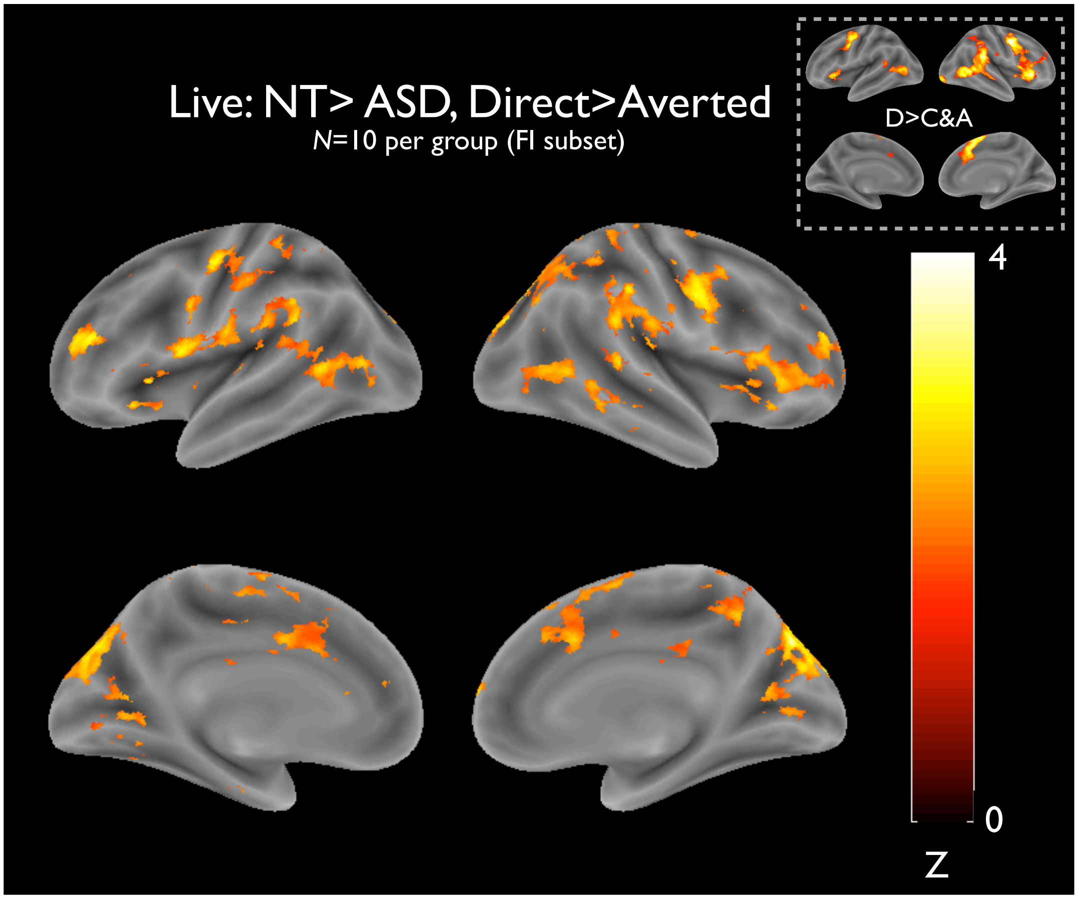


Figure 3.1.13. Qualitative Comparison of (1) NT vs. ASD Response to Direct Gaze in Live Condition and (2) Main Effect of Direct Gaze. Comparisons of the main effect of Direct gaze (inset, copied from Figure 3.1.11) to group contrast of NT >ASD responses to Direct > Averted gaze in the Live condition (whole-brain thresholding of z>1.65, with cluster-level thresholding at *p*=0.05) show that the main effect of Direct gaze was largely driven by the group contrast of Joint gaze, i.e., Direct>Averted gaze in the Live condition

Activity in the three gaze conditions (Direct, Closed, Averted), as well as a direct comparison of Direct versus Averted gaze (i.e., Joint versus non-Joint gaze) are explored in detail in the proceeding region of interest (ROI) analyses.

Group by Condition Analysis of Independently Defined ROIs

Four groups of independently defined Regions of Interest (ROIs) were identified. Relevant to social cognition, (1) putative Mirror Neuron System (MNS) regions (inferior frontal gyrus and inferior parietal lobule) and (2) theory of mind (TOM) regions (dorsal medial prefrontal cortex, ventral medial prefrontal cortex, anterior superior temporal sulcus, and temporoparietal junction) were investigated. Relevant to (3) face processing, the fusiform face area and amygdala (whose involvement we also expected for reasons beyond face processing) were investigated. And, (4) relevant to cognitive control and task engagement, a region of the anterior cingulate cortex involved in the Stroop task as well as the precuneus, a default mode network hub (Utevsky et al., 2014) and also a theory of mind region (Spunt & Adolphs, 2014), was investigated. A list of the ROIs investigated is detailed in Table 3.1.2.

Table 3.1.2. Regions of Interest (ROIs) Tested. ROIs were roughly grouped by four categories. Their names used throughout the paper, anatomical description, how they were defined (either through clusters from Neurosynth (Yarkoni et al., 2011) reverse inference masks or anatomical labels), and voxel size[5] are described.

Group	Name	ROI	How Defined	Voxels
MNS	rIFG	Right inferior frontal gyrus, pars triangularis and some IFG pars opercularis; extending into anterior insula & frontal operculum	Neurosynth “Social”	392
	rIPL	Right inferior parietal lobule	Juelich Atlas	3084
TOM	dmPFC	Dorsal medial prefrontal cortex	Neurosynth “TOM”	1537
	vmPFC	Ventral medial prefrontal cortex	Neurosynth “TOM”	252
	rTPJ	Right temporoparietal junction	Neurosynth “TOM”	1048
	raSTS	Right anterior superior temporal sulcus	Neurosynth “TOM”	555

---

[5] Future analyses will include analysis of larger ROIs, for example TOM ROIs directly from published papers, as well as some subject-specific ROIs from other tasks in the Adolphs lab that feature an overlapping subject pool.

Face	lFFA	Left fusiform face area	Neurosynth “Face”	3182
	rFFA	Right fusiform face area	Neurosynth “Face”	2693
	lAmy	Left amygdala	Neurosynth “Face”	691
	rAmy	Right amygdala	Harvard-Oxford Atlas	2083
Control/	Precuneus	Precuneus	Neurosynth “TOM”	868
Engagement	ACC	Anterior cingulate cortex	Neurosynth “Stroop”	91

Putative Mirror Neuron System Regions. Two ROIs associated with the so-called Mirror Neuron System were explored. The first was the right Inferior Parietal Lobule (rIPL), defined anatomically by the Juelich atlas. The second was a “social” cluster in the right Inferior Frontal Gyrus (rIFG), identified by the meta-analysis tool Neurosynth (Yarkoni et al., 2011). In the rIFG, in the Live condition only, the pattern of the response to Direct gaze was distinct from the other gaze conditions in the NT group, but not the ASD group (Figure 3.1.14a). The same pattern was observed in the rIPL (Figure 3.1.14b). Comparisons of Direct and Averted gaze allowed us to directly compare responses to Joint and Non-Joint gaze (especially in the Live condition). In the FI subset of participants who self-reported being sensitive to our condition manipulation (n=10 of 14 per group, orange bars in Figure 3.1.14), in the rIFG and rIPL, the contrast of Direct to Averted gaze was amplified in the Live compared to Recorded condition, but only in controls (rIFG: $t(9)=2.16$, $p=0.059$; rIPL: $t(9)=2.63$, $p=0.028$) (Figure 3.1.14c,d).

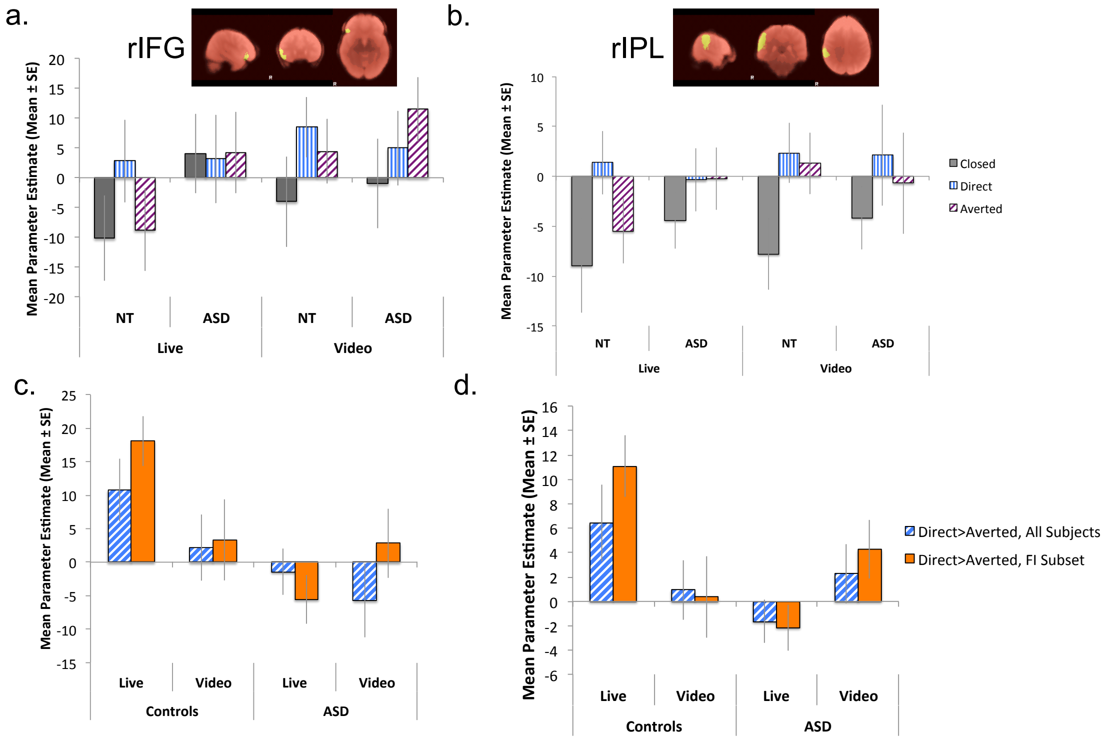


Figure 3.1.14. ROI Analysis in pMNS Regions. (Top) Main effect of Direct>Averted and Closed gaze and (bottom) contrast of Direct>Averted gaze in the Live condition (Joint vs. Non-Joint gaze). Analyses conducted in the rIFG (left) and rIPL (right).

Potentially, group differences in neural response to a Live person could be used as a robust neural marker of ASD diagnosis. In a preliminary analysis, individual responses to Joint gaze (Direct>Averted gaze in the Live condition) in just two regions (our rIFG and rIPL ROIs) mostly separated our FI subset of participants into individuals with autism and matched controls (Figure 3.1.15).

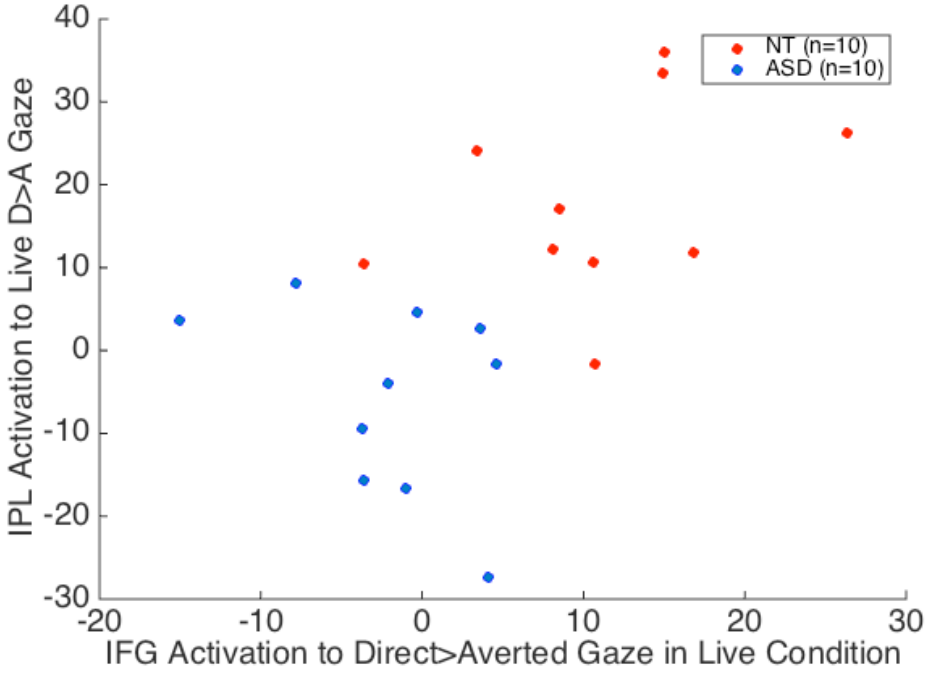


Figure 3.1.15. Mirror Region Activation to Joint Gaze Separates NT and ASD Participants. Among the FI subgroup of participants, individuals' responses to Direct>Averted Live gaze in the rIPL and rIFG mostly separated into two groups according to autism diagnosis. Red dots: NT; blue dots: ASD.

Putative TOM Regions. Activity in the dmPFC, vmPFC, and rTPJ all exhibited patterns of deactivation across gaze conditions in the Live relative to Video condition. In rTPJ (Figure 3.1.16d), this is true for both NT ($F(1,13)=3.97$, $p=0.0497$) and ASD ($F(1,13)=13.3$, $p=0.0005$) participants, but only true for controls in dmPFC ($F(1,13)=5.5$, $p=0.022$) (Figure 3.1.16a) and vmPFC ($F(1,13)=20.7$ $p<0.000$) (Figure 3.1.16b). A positive response to Direct, Averted, and Closed gaze was observed in the raSTS for Controls in both the Live and Video conditions, with no difference observed between those two conditions. Unlike controls, ASD subjects did not have a positive response in the raSTS to Video gaze; as a group, their response to Live gaze was highly variable and any visual trend for it to be higher than in the Video condition was statistically insignificant (*F*(1,13)=1.98 *p*=0.16) (Figure 3.1.16c). There was no significant condition modulation of specific processing of Direct (relative to Averted) gaze in these TOM ROIs.

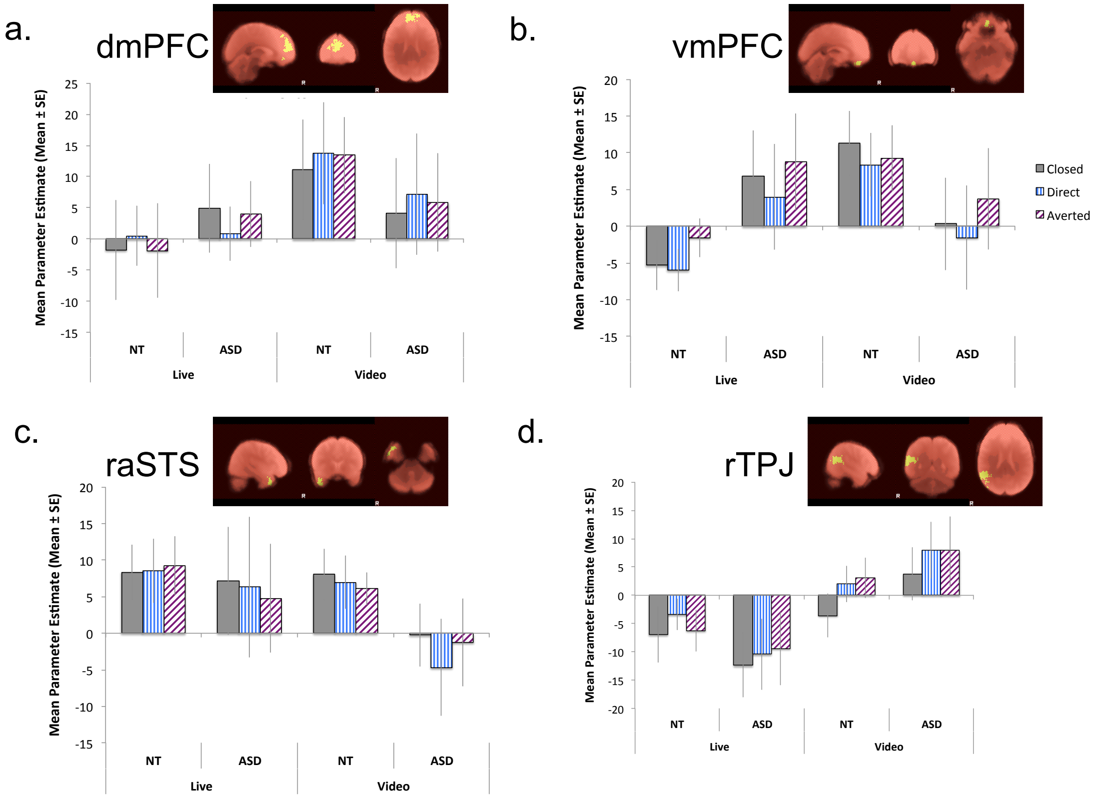


Figure 3.1.16. ROI Analysis in TOM Regions, Main Effect of Condition. Main effect of Direct>Averted and Closed gaze in the (a) dmPFC, (b) vmPFC, (c) raSTS, and (d) rTPJ.

Face-Responsive Regions: In the left FFA, there was decreased activity in the Live relative to Video condition in each gaze condition for both groups (NT: *F*(1,13)=5.69, *p*=0.0195; ASD: *F*(1,13)=14.45, *p*=0.0003) (Figure 3.1.17a). This was also true in the right FFA for controls (*F*(1,13)=9.19, *p*=0.003), but the pattern did not hold for the ASD group (Figure 3.1.17b). In the right FFA, ASD subjects (but not NT subjects) tended to respond differently in the Video and Live conditions to Direct relative to Averted gaze (*t*(13)=2.09, *p*=0.057, Figure 3.1.17d).

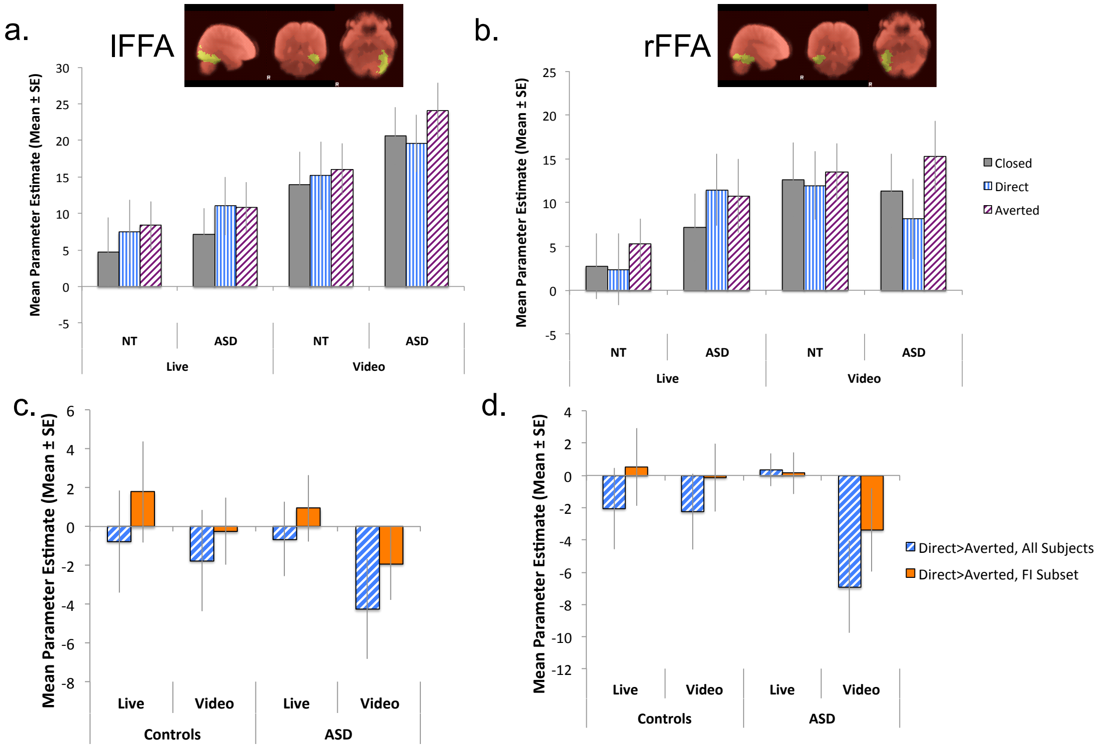


Figure 3.1.17. ROI Analysis in FFA. (Top) Main effect of Direct>Averted and Closed gaze and (bottom) contrast of Direct>Averted gaze in the Live condition (Joint vs. Non-Joint gaze). Analyses conducted in the left FFA (left) and right FFA (right).

In the amygdala, ASD subjects differed more from controls in the Video than Live condition. While initially counterintuitive, this result was driven by the unsurprising finding that ASD subjects had a greater amygdala response to potentially more arousing Direct>Averted gaze in the Live than Video condition.

Specifically, in the left amygdala there was no difference between the two groups in the Live condition, but the groups were different in the Video condition (*F*(1,27)=17.43, *p*=0.0001) (Figure 3.1.18a). Across gaze types, there was a trend in NT subjects for decreased responses in the Live relative to Video condition (*F*(1,13)=3.37, *p*=0.070). In ASD participants, the visual trend for the opposite effect (increased negative response in the Video condition) was insignificant.

In the right amygdala, as in the left amygdala, there was a difference between the groups in the Video *(F*(1,27)=7.89, *p*=0.0063) but not the Live condition (Figure 3.1.18b). There was no effect of condition within group.

In the amygdala, only the ASD group showed a difference in response to Direct>Averted gaze as a function of experimental condition. In both the left and right amygdala, they showed an increase in response to Direct relative to Averted gaze in the Live condition (lAMY: *t*(13)=2.11, *p*=0.055; rAMY: *t*(13)=2.25, *p*=0.042) (Figure 3.1.18c,d).

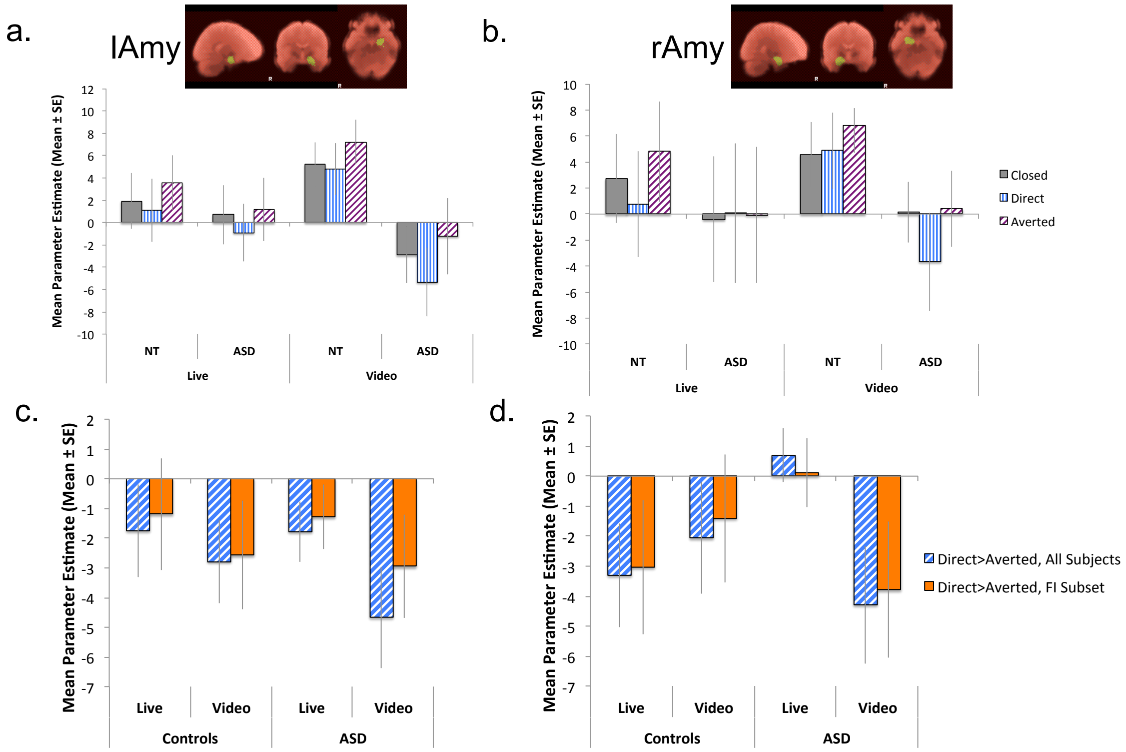


Figure 3.1.18. ROI Analysis in Amygdala. Main effect of Direct>Averted and Closed gaze and (bottom) contrast of Direct>Averted gaze in the Live condition (Joint vs. Non-Joint gaze). Analyses conducted in the left amygdala (left) and right amygdala (right).

Control/Engagement. In addition to social and face processing, we were interested in differences in cognitive control and engagement in the two conditions. Our initial investigation of these processes focused on two ROIs: (2) precuneus voxels extracted from a Neurosynth "TOM" mask (note the precuneus is implicated both in TOM and default mode networks) and (2) an ACC region defined by the Neurosynth "stroop" reverse inference map.

In the precuneus ROI, both groups actually responded similarly within each condition. Both groups had a strong decrease in activity when presented with Live compared to Video gaze for all three gaze types (NT: $F(1,13)=15.32$, $p=0.0002$; ASD:

F(1,13)=5.54, *p*=0.021) (Figure 3.1.19a). In the NT group, the effect was especially pronounced for direct and averted gaze, while in the ASD group, the effect was most pronounced for direct gaze specifically, resulting in trend towards a conditional effect on Direct>Averted activation being observed in ASD participants who reported being sensitive to the experimental condition manipulation (Trend: *t*(9)=-1.77, *p*=0.11) (Figure 3.1.19c).

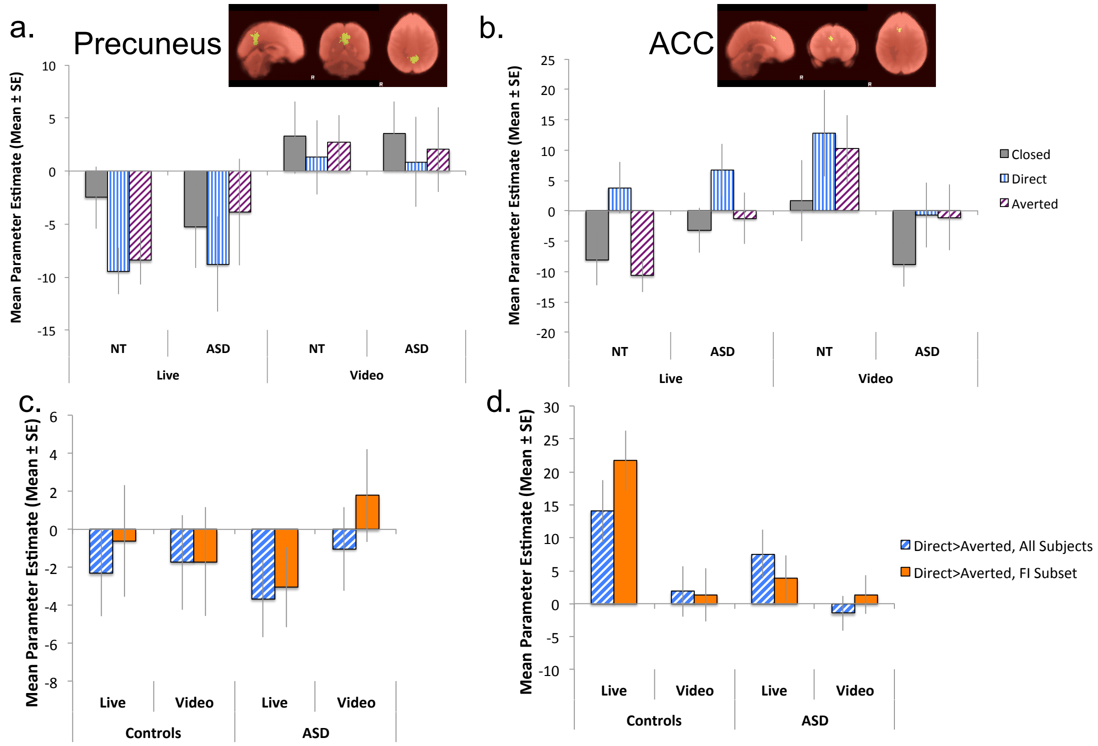


Figure 3.1.19. ROI Analysis in Precuneus and ACC Regions. Main effect of Direct>Averted and Closed gaze and (bottom) contrast of Direct>Averted gaze in the Live condition (Joint vs. Non-Joint gaze). Analyses conducted in the precuneus (left) and ACC (right).

Meanwhile, in the ACC region associated with the Stroop task and cognitive control, there was an interaction between gaze type and experimental condition.

In the Live condition, sustained Direct gaze elicited more ACC activity in both groups than either Averted or Closed gaze (Figure 3.1.23b). This contrast between Direct and Averted gaze in the Live condition was stronger in controls (NT: *t*(13)=2.23, *p*=0.044; ASD: (*t*(13)=1.82, *p*=0.092), especially the subset who reported feeling a sense of interaction in the Live condition (*t*(9)=4.86, *p*=0.0008) (Figure 3.1.19d). In the Video condition, controls actually had elevated activity in the ACC ROI for both the Direct and Averted relative to Closed conditions, while no gaze type elicited positive ACC activation in Video condition for ASD individuals (Figure 3.1.19b).

Preliminary Analysis of Individual Differences in Response to Joint Gaze

It is important to note that these correlational analyses involve a small sample size and are therefore preliminary.

Inspired by the finding in children that autisms symptom severity was inversely correlated with activity in the IFG during an emotional expression imitation and observation task (Dapretto et al., 2006), we tested whether activity in response to Joint gaze (i.e., Direct>Averted gaze in the Live condition) in our social IFG region was also related to autism severity. While ADOS scores were designed for classification rather than a severity metric, the ADOS Calibrated Severity Score (CSS) is appropriate for use as a severity metric (Hus, Gotham, and Lord, 2015). While ideal, this metric was only available for 12 of our 14 ASD participants (for whom we had copies of their original ADOS score sheets). Therefore, we decided to use a different measure, the ADOS Social Interaction Module 4 algorithm scores (ADOS-SI), as a metric of severity. This was justified because those 12 individuals' available CSS social-affective sub-scores (CSS-

SA) correlated highly with those same individual's ADOS Social Interaction Module 4 algorithm scores (*n*=12, *r*=0.88, 95% Confidence Interval (C.I.) [0.62, 0.97], *p*=0.0001). This high correlation made sense as the CSS was designed to remove confounds like behavioral problems that were not prevalent in our high-functioning adult population. In all 14 ASD participants, that ADOS Social Interaction score was weakly inversely correlated with rIFG responses (*n*=14, *r*=-0.50, 95% C.I. [-0.81, 0.04], *p*=0.07) to Joint gaze. The inverse correlation relationship was stronger in the FI subset of ASD participants who were sensitive to our condition manipulation (*n*=10, *r*=-0.70, 95% C.I. [-0.92, -0.12], *p*=0.02).

While responses in the right IFG ROI, associated with the putative Mirror Neuron System, were inversely correlated with autism severity, responses in the perceptual right FFA to Live Joint gaze were positively correlated with ADOS severity (*n*=14, *r*=0.62, 95% C.I. [0.13, 0.87], *p*=0.02), in line with the decrease in activity seen in this region in the Live condition in controls (cf. Figure 3.1.17b).

In both individuals with autism and controls, performance on the Reading the Mind in the Eyes task (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) was related to responses to Joint gaze in various regions.

In the ASD group, scores on the Eyes task were weakly inversely correlated with activity in the rTPJ (which in controls saw decreased activity in the Live condition, cf. Figure 3.1.16d) (*n*=14, *r*=-0.51, 95% C.I. [-0.82, 0.03], *p*=0.06).

The two groups tended to have opposite relationships between right amygdala responses to Joint gaze and performance on the Eyes task (NT: *n*=14, *r*=0.66, 95% C.I.

[0.08, 0.92], *p*=0.03); ASD: *n*=14, *r*=-0.50, 95% C.I. [-0.82, 0.04], *p*=0.066), with stronger correlations seen in the FI subgroups (NT: *n*=14, *r*=0.88; ASD: *n*=14*, r*=-0.55).

Engagement with the Live condition modulated responses to Joint gaze in the ACC: ACC responses to Joint gaze was positively correlated with Felt-Interaction (FI) categorization from our post-task questionnaire in the NT group (*n*=14, *r*=0.72, 95% C.I. [0.29, 0.90], *p*=0.004) but possibly weakly inversely related with FI categorization in the ASD group (*n*=14*, r*=-0.44, 95% C.I. [-0.79, 0.12], *p*=0.12).

Discussion

Summary of Findings

Relevant to our first aim to address three hypotheses concerning group differences in response to Live and Video gaze, we (1) showed neural sensitivity to the Live versus Video presence of another person; we (2) showed that this sensitivity was especially prominent with respect to Direct, joint gaze (i.e., Direct>Averted; Direct>Averted and Closed gaze); and (3) we showed that the Live condition differently affected individuals with autism (ASD) relative to matched controls (NT). In contrasts of Live and Recorded Direct (joint) versus Averted (non-joint) gaze, controls had more activity in various regions, including the inferior frontal gyrus, dorsal medial prefrontal cortex, anterior cingulate, precuneus, anterior insula, and fusiform gyrus. These regions showed considerable overlap with those involved in a “face-to-face” social interaction mediated by a Live video feed (Redcay et al., 2010). In contrast to the controls, we saw no whole-brain difference in response to direct Live versus Recorded gaze in the group of individuals with autism.

Notably, direct gaze elicits activity not only in regions associated with direct gaze (Calder et al., 2002) , but also in the putative Mirror Neuron System. Functional (Dapretto et al., 2006) and structural (Hadjikhani, Joseph, Snyder, & Tager-Flusberg, 2006) differences in the so-called Mirror Neuron System have been associated with autism. Importantly, these “mirroring” regions are best associated with processing *emotional* social cues, pointing to impaired top-down social modulation in autism, as in Hamilton and colleague’s STORM (“social top-down response modulation” model) (Hamilton, 2013). Interestingly, processing in face and TOM regions tended to decreased in the Live condition; one potential hypothesis, which remains to be tested using functional connectivity and mediation analyses is that the Live condition, i.e., the presence of another person, elicits a switch in social processing from theory-based mentalizing about others to a more simulation-based understanding of others; potentially, such a “switch” could be mediated by the precuneus, a DMN hub (Utevsky et al., 2014) often implicated in TOM networks. Direct gaze elicited greater activity in the thalamus, which has been shown to mediate attention and arousal (Portas et al., 1998); to that end, we are especially interested in how future analysis of autonomic arousal (see below) interacts with this region. Finally, in the Live condition, Direct gaze elicited ACC activity, a region implicated in cognitive control – a process our participants may well have needed to maintain sustained direct gaze with another person!

Relevant to our second aim to preliminarily relate differences in the neural response to Joint Live gaze to individual differences, we found that activity in certain putative Mirror Neuron System and Theory of Mind regions as well as the ACC related to autism severity, as well as two social skills measures: the Empathizing Quotient and the Reading the Mind in the Eyes task. A third, potent individual difference related to neural activity across both groups was the degree to which individuals were sensitive to the experimental manipulation, i.e., whether or not they reported feeling interaction with or the presence of the actor during the Live condition. Looking within this subgroup of individuals (*n*=20), the neural response to Joint gaze in just two ROIs, the rIFG and rIPL, both putative Mirror Neuron System regions, was sufficient to separate 19 of the 20 FI participants into two clusters that corresponded with their ASD diagnosis. Experimentally, the emotional engagement observed in this subset of subjects may have been necessary to engender truly interactive, second-person social cognition (Schilbach et al., 2013).

Caveats and Shortcomings

While we are confident that a majority of our participants were sensitive to our manipulation of social presence (i.e., the Live condition did in fact feel subjectively different form he Video condition), one shortcoming to point out is that we do not definitely know (1) whether all participants were sensitive to our social modulation (2) and whether there was heterogeneity in the degree to which participants felt they were being watched (based on our FI categorization of participants, we are fairly confident that at least 71% of our participants felt that they were watching a live person, but we are not sure whether they felt the actor could see them). Some individuals did spontaneously self-report this, but we do not know whether they all felt bi-directional observation, which would be required for true joint gaze (an experience that would distinguish this study from others such as the work by Redcay and colleagues (2013; 2010)). In future work, it will be important to more carefully interview participants to be able to divide them into subgroups based on their social experience within the experiment. Potentially, we could also modify the experimental set up to better facilitate social engagement during the Live condition.

An important caveat for all our findings is that they are exploratory: within a small sample of participants, we employed a fairly open-ended design that permitted different levels of personal engagement with the task.

Future Directions

Future directions fall under two categories: (1) further analyses to be carried out within the current data set and (2) future studies and extensions.

Further Analyses. One area of future analysis is to better explore individual differences. For instance, while this analysis focused on ROI's defined completely independently from this study, looking at subject-specific ROIs (inspired by (Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010)) may allow us to see similarities or differences obscured in our ROI analyses that assumed all participants' functional localization was anatomically defined. In a subject-specific analysis, it would be possible to take externally-defined ROIs (as we did here, but perhaps in broader regions defined by a handful of highly-relevant studies — e.g., (Dapretto et al., 2006; Pitskel et al., 2011; Redcay et al., 2013; Redcay et al., 2010; Spunt & Adolphs, 2014) instead of the meta-analysis tool we used — and then explore volumes around each individual subject's peak activation in those externally-defined regions. In addition to defining ROIs differently, it is important to rigorously test whether outliers exist in either participant group to determine whether those individuals may be driving any observed group effects (Byrge, Dubois, Tyszka, Adolphs, & Kennedy, 2015). Future ROI analyses will also investigate response time courses as well as potential functional connectivity differences between groups and conditions. We can look more at individual differences that correspond with ROI response once ROI analyses are refined.

A future aim would be to train a classifier to predict groups based on activity in a subset of regions. We already started to achieve this in the Felt Interaction subset of participants, but would want better predictive power across the full set of participants. In this analysis, we may find subtypes within each group.

Another exciting possibility is to extract a pupil dilation response (PDR) from our eye tracking data as a measure of autonomic response, which could be analyzed as a function of condition and also added to our fMRI model.

Finally, we were surprised by the variability of response found in the amygdala. In group registration masks, drop out did not seem problematic in this region, but future analyses can look for drop out and field distortion in more detail, also looking for potential physiological artifacts, such as from the middle cerebral artery. Future scanning sequences could be tailored to target specific regions, such as the amygdala.

Future Extensions. In a novel study using real, live people as stimuli in the scanning room, we demonstrated that the social presence of another person modulates the neural response to gaze in a broad number of brain regions associated with social, face and gaze processing and cognitive attention/control. The pattern of response between individuals with autism and controls was more different in the Live than Video condition, suggesting the importance of using truly social stimuli in both research and therapy targeted at ASD. Future work in this area of interactive social neuroscience should (1) examine processing of interaction in developmental and non-human populations (Banaji

& Gelman, 2013) as well as (2) develop more spontaneous, bi-directionally interactive paradigms. It is exciting to imagine what this future work will reveal about the social brain.

CONCLUSIONS AND FUTURE DIRECTIONS

Building on a large literature and theoretical framework, we provide empirical evidence that ecologically valid methods need to be incorporated into the study of social brain function. In this thesis, three themes of ecological validity were proposed: (1) the addition of contextual cues — perceptual or implied — to make an experimental manipulation or stimulus convincingly naturalistic; (2) a special contextual manipulation of inducing an interactive experience — this can be achieved through face-to-face interactions or through convincing subject-contingent stimuli; and (3) a theme that runs in parallel to the first two: believability — subjects must believe in the experimental manipulation.

Contextual Cues and Interaction Influence Social Cognition

In the first two chapters, we showed that (1) contextual cues and (2) face-to-face interaction do matter for social cognition. Further, we showed that particular brain structures mediate assessment of contextual cues.

Specifically, in the second chapter, “The Amygdala and Social Threat”, we showed that the amygdala mediates basic approach-avoidance behavior. Lesions to this structure promote an enhanced default approach bias, operationalized as the difference in trust/threat judgments to low-information, occluded faces relative to unaltered facial images. The contextual factor of ambiguity was one of many that was then used in the formation of a hierarchical decision tree that successfully predicted a replication group of participants’ specific responses to a set of threating scenarios (about 60% success versus about 12% chance performance). The tree comprised a hierarchical model of threat assessment, in which responses were first divided according to two broad factors: perceived degree of dangerousness and threat category (psychological versus physical threat) to guide basic approach-avoidance behavior. Subsequently, sequential processing of various factors along nodes in the decision tree elicited specific defensive responses. Hypothetically, specific brain structures, including the amygdala, perform neural computations associated with each of those nodes present; future work can test this behavioral model neuroscientifically.

In the third chapter, "Face-to-Face Social Cognition", we utilized a novel approach in neuroimaging research to test the response of specific brain structures to the live presence of another person. This experimental manipulation showed that face, gaze, and social regions of our brains respond differently to another person's gaze in the presence of that other person instead of in the presence of a representation, i.e., video, of that person. Specifically, live interaction promoted increased activity to Direct relative to Averted gaze in putative Mirror Neuron System regions relative to a video condition. Meanwhile responses to all gaze conditions (Direct, Closed, and Averted) actually decreased in the live relative to video condition in Theory of Mind regions, where Direct gaze was not distinct from Averted gaze. Together these patterns preliminarily suggesting that different experimental paradigms may modulate the interaction of these two broad-scale social neural networks. Further, modulation of the neural response to direct gaze by the presence of another person was impaired in individuals with autism — a disorder with core impairments in social interaction — relative to controls. In preliminary analyses, it was largely possible to separate the groups by autism diagnosis as a function of individuals' neural response in two putative Mirror Neuron System regions of interest. This separation was not possible in the video condition, suggesting that research using interactive paradigms may be necessary to gain insight into autism and other psychiatric conditions featuring social impairments.

Future Directions in Real-World Social Neuroscience

While this thesis has yielded several observations in support of the importance of ecological validity in social neuroscience research, the most exciting outcome of this thesis is the potential for future findings in "real-world" social cognition. Direct extensions of this thesis can be broken down into near-term directions directly related to the projects of this thesis, as well as future directions for further in the future.

Near-Term Directions. These can be broken down by specific sections of the thesis:

Extending and further validating our discovery of a default approach-related bias in bilateral amygdala lesion patients, we would like to test for an approach bias across many classes of stimuli.

Our threat scenario paradigm relied on self-report to imagined scenarios. While the scenario paradigm has been used by others in psychology, it is desirable to improve realism by immersing a participant in an experience rather than relying on the participant's imagination. We would like to start to achieve this using virtual reality. Additionally, we would like to examine the import of individual differences, including those related to approach-avoidance behavior, such as anxiety and amygdala lesions. Further, we would like to corroborate explicit ratings with implicit, affective response measures.

An obvious near-term goal for our Face-to-Face fMRI experiment would be to establish the statistical reliability of the patterns found by replicating the study in a second group or by increasing our sample size. Future analyses in the experiment include further region of interest analyses, including defining ROIs from other studies, as well as idiosyncratically in subjects. Using these ROIs, we would like to look at individual differences in connectivity between regions as well as the time course of response to gaze conditions, to further characterize individual and group differences, and test whether group differences are driven by outliers or whether different sub-groups exist. Finally, we would like to extract a measure of autonomic response from the pupil in our eyetracking data, to potentially add condition-dependent fluctuations in autonomic processing to our model.

Further Future Directions. Further future directions center around (1) improved ecological validity in measuring individual responses, focusing on implicit and affective response, and (2) continuing to probe social neuroscience in interaction.

There is evidence that humans bodily react to affective stimuli, such as freezing to emotional stimuli (Roelofs et al., 2010). Measuring these implicit measures as well as any affective cues (autonomic response as well as changes in facial expression) will provide insight into individual differences as well as potential embodied processing in social cognition. Such embodied processing may be relevant in interactive paradigms, including ours in which the anterior insula (extending from the MSN inferior frontal gyrus region) was more strongly activated in the live than video condition.

In interaction, it will be important to study more dynamic, truly contingent actions. This has been done in EEG research relating coherence in interacting individuals' brain activity measured at the scalp with bodily synchrony (Liao, Acar, Makeig, & Deak, 2015; Yun, Watanabe, & Shimojo, 2012). Body synchrony, including functions such as pupil mimicry (Kret, Tomonaga, & Matsuzawa, 2014), is an interesting area of research enabled by interactive techniques. Finally, to better understand the role of interaction in social cognition, we should test whether human infants and other species (Banaji & Gelman, 2013) and sensitive to the contingent, social presence of another being.

Future. Social neuroscience is a young, rapidly growing field. By incorporating ecologically valid methods, especially interactive paradigms, the field is poised to exponentially increase our understanding of how our brains facilitate navigation of our incredible social environment.

Bibliography

Abbott, D. H., Keverne, E. B., Bercovitch, F. B., Shively, C. A., Medoza, S. P., Saltzman, W., . . . Sapolsky, R. M. (2003). Are subordinates always stressed? A comparative analysis of rank differences in cortisol levels among primates. *Hormones and Behavior, 43*(1), 67-82. doi: 10.1016/s0018-506x(02)00037-5

Adams, R. B, Gordon, H. L, Baird, A. A, Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science, 300*(5625), 1536-1536.

Adolphs, R. (2010). What does the amygdala contribute to social cognition? In A. Kingstone & M. B. Miller (Eds.), *Year in Cognitive Neuroscience 2010* (Vol. 1191, pp. 42-61). Malden: Wiley-Blackwell.

Adolphs, R., Tranel, D., & Damasio, A.R. (1998). The human amygdala in social judgment. *NATURE-LONDON-*, 470-473.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A, . . . Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia, 37*(10), 1111-1117.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*(7021), 68-72. doi: http://www.nature.com/nature/journal/v433/n7021/suppinfo/nature03086_S1.html

Adolphs, R., Spezio, M. L., Parlier, M., & Piven, J. (2008). Distinct face-processing strategies in parents of autistic children. *Current Biology, 18*(14), 1090-1093.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature, 372*(6507), 669-672.

Aggleton, J.P. (2000). *The Amygdala: A functional analysis* (J. P. Aggleton Ed.). New York: Oxford University Press.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences, 4*(7), 267-278. doi: 10.1016/s1364-6613(00)01501-1

Amaral, D. G., Price, J.L., Pitkänen, A., & Carmichael, S.T. (1992). Anatomical Organization of the Primate Amygdaloid Complex. In J. P. Aggleton (Ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (pp. 1-66). New York: Wiley-Liss, Inc.

Amodio, D. M, & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277.

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., . . . Zilles, K. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anatomy and Embryology, 210*(5-6), 343-352. doi: 10.1007/s00429-005-0025-5

Anderson, A. K, Christoff, K., Stappen, I., Panitz, D., Ghahremani, D.G., Glover, G., . . . Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience, 6*(2), 196-202.

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell, 157*(1), 187-200.

Anderson, D. J., & Perona, P. (2014). Toward a Science of Computational Ethology. *Neuron, 84*(1), 18-31.

Appenzeller, S., Chaloult, E., Velho, P., de Souza, E. M., Araujo, V. Z., Cendes, F., & Li, L. M. (2006). Amygdalae calcifications associated with disease duration in lipoid proteinosis. *Journal of Neuroimaging, 16*(2), 154-156. doi: 10.1111/j.1552-6569.2005.00018.x

Arnold, M.B. (1960). *Emotion and personality.* New York, NY: Columbia University Press.

Asmundson, G.J.G., Sandler, L.S., Wilson, K.G., & Walker, J.R. (1992). Selective attention toward physical threat in patients with panic disorder. *Journal of Anxiety Disorders, 6*(4), 295-303.

Asmundson, G.JG, & Stein, M.B. (1994). Selective processing of social threat in patients with generalized social phobia: Evaluation using a dot-probe paradigm. *Journal of Anxiety Disorders, 8*(2), 107-117.

Bach, D.R., & Dolan, R.J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience, 13*(8), 572-586. doi: 10.1038/nrn3289

Ballew, C.C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences, 104*(46), 17948-17953.

Banaji, M.R., & Gelman, S.A. (2013). *Navigating the social world: What infants, children, and other species can teach us*: Oxford University Press.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., & Van Ijzendoorn, M.H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological Bulletin, 133*(1), 1.

Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., & Williams, S. C. R. (2000). The amygdala theory of autism. *Neuroscience and Biobehavioral Reviews, 24*(3), 355-364. doi: 10.1016/s0149-7634(00)00011-7

Baron‐Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high‐functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Becker, B., Mihov, Y., Scheele, D., Kendrick, K.M., Feinstein, J.S., Matusch, A., . . . Oros-Peusquens, A. (2012). Fear processing and social networking in the absence of a functional amygdala. *Biological Psychiatry, 72*(1), 70-77.

Bickart, K. C., Hollenbeck, M. C., Barrett, L. F., & Dickerson, B. C. (2012). Intrinsic Amygdala-Cortical Functional Connectivity Predicts Social Network Size in Humans. *Journal of Neuroscience, 32*(42), 14729-14741. doi: 10.1523/jneurosci.1599-12.2012

Björkqvist, K. (2001). Social defeat as a stressor in humans. *Physiology & Behavior, 73*(3), 435-442.

Blanchard, D.C., Griebel, G., Pobbe, R., & Blanchard, R.J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience and Biobehavioral Reviews, 35*(4), 991-998. doi: 10.1016/j.neubiorev.2010.10.016

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

Blanchard, R. J., & Blanchard, D. C. (1989). ATTACK AND DEFENSE IN RODENTS AS ETHOEXPERIMENTAL MODELS FOR THE STUDY OF EMOTION. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 13*, S3-S14. doi: 10.1016/0278-5846(89)90105-x

Blanchard, R. J., Blanchard, D. C., & Hori, K. (1989). *AN ETHOEXPERIMENTAL APPROACH TO THE STUDY OF DEFENSE* (Vol. 48).

Blanchard, R. J., Blanchard, D. C., Rodgers, J., & Weiss, S. M. (1990). THE CHARACTERIZATION AND MODELING OF ANTIPREDATOR DEFENSIVE BEHAVIOR. *Neuroscience and Biobehavioral Reviews, 14*(4), 463-472. doi: 10.1016/s0149-7634(05)80069-7

Blanchard, R. J., Flannelly, K. J., & Blanchard, D. C. (1986). DEFENSIVE BEHAVIORS OF LABORATORY AND WILD RATTUS-NORVEGICUS. *Journal of Comparative Psychology, 100*(2), 101-107. doi: 10.1037//0735-7036.100.2.101

Broks, Paul, Young, Andrew W, Maratos, Elizabeth J, Coffey, Peter J, Calder, Andrew J, Isaac, Claire L, . . . Cezayirli, Enis. (1998). Face processing impairments after encephalitis: amygdala damage and recognition of fear. *Neuropsychologia, 36*(1), 59-70.

Brosnan, S.F., Schiff, H.C., & De Waal, R. BM. (2005). Tolerance for inequity may increase with social closeness in chimpanzees. *Proceedings of the Royal Society of London B: Biological Sciences, 272*(1560), 253-258.

Byrge, L., Dubois, J., Tyszka, J.M., Adolphs, R., & Kennedy, D.P. (2015). Idiosyncratic Brain Activation Patterns Are Associated with Poor Social Comprehension in Autism. *The Journal of neuroscience, 35*(14), 5837-5850.

Calder, A.J., Beaver, J.D., Winston, J.S., Dolan, R.J., Jenkins, R., Eger, E., & Henson, R.N.A. (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Current Biology, 17*(1), 20-25.

Calder, A.J., Jenkins, R., Cassel, A., & Clifford, C.W.G. (2008). Visual representation of eye gaze is coded by a nonopponent multichannel system. *Journal of Experimental Psychology: General, 137*(2), 244.

Calder, A.J., Lawrence, A.D, Keane, J., Scott, S.K., Owen, A.M., Christoffels, I., & Young, A.W. (2002). Reading the mind from eye gaze. *Neuropsychologia, 40*(8), 1129-1138.

Carlin, J.D., Calder, A.J, Kriegeskorte, N., Nili, H., & Rowe, J.B. (2011). A head view-invariant representation of gaze direction in anterior superior temporal sulcus. *Current Biology, 21*(21), 1817-1821.

Castelli, F., Frith, C., Happe, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*, 1839-1849. doi: 10.1093/brain/awf189

Chang, S.W.C., Brent, L.J.N., Adams, G.K., Klein, J.T., Pearson, J.M., Watson, K.K., & Platt, M.L. (2013). Neuroethology of primate social behavior. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 10387-10394. doi: 10.1073/pnas.1301213110

Choi, J-S, & Kim, J.J. (2010). Amygdala regulates risk of predation in rats foraging in a dynamic fear environment. *Proceedings of the National Academy of Sciences of the United States of America, 107*(50), 21773-21777. doi: 10.1073/pnas.1010079108

Coleman, K., & Pierre, P.J. (2014). Assessing Anxiety in Nonhuman Primates. *Ilar Journal, 55*(2), 333-346. doi: 10.1093/ilar/ilu019

Conty, L., N'Diaye, K., Tijus, C., & George, N. (2007). When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing. *Neuropsychologia, 45*(13), 3024-3037.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to "prejudice". *Journal of Personality and Social Psychology, 88*(5), 770-789. doi: 10.1037/0022-3514.88.5.770

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science, 304*(5667), 115-117.

Crawley, J. N. (1999). Behavioral phenotyping of transgenic and knockout mice: experimental design and evaluation of general health, sensory functions, motor abilities, and specific behavioral tests. *Brain Research, 835*(1), 18-26. doi: 10.1016/s0006-8993(98)01258-x

Crawley, J. N. (2003). Behavioral phenotyping of rodents. *Comparative Medicine, 53*(2), 140-146.

Cunningham, W.A., & Brosch, T. (2012). Motivational salience amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science, 21*(1), 54-59.

Dahl, C.D., & Adachi, I. (2013). Conceptual metaphorical mapping in chimpanzees (Pan troglodytes). *elife, 2*, e00932.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., . . . Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8*(4), 519-526. doi: 10.1038/nn1421

Dalton, K.M., Nacewicz, B.M, Johnstone, T., Schaefer, H.S., Gernsbacher, M.A., Goldsmith, H.H., . . . Davidson, R.J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8*(4), 519-526.

Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience, 9*(1), 28-30.

Darwin, C. (1872/1965). *The Expression of the Emotions in Man and Animals*. Chicago: University of Chicago Press.

Davis, M. (1992). THE ROLE OF THE AMYGDALA IN FEAR AND ANXIETY. *Annual Review of Neuroscience, 15*, 353-375. doi: 10.1146/annurev.neuro.15.1.353

Davis, M., & Whalen, P. J. (2001). The amygdala: vigilance and emotion. *Molecular Psychiatry, 6*(1), 13-34. doi: 10.1038/sj.mp.4000812

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

Deen, B., & Pelphrey, K. (2012). Perspective: Brain scans need a rethink. *Nature, 491*(7422), S20-S20.

Dewaal, F. B. M. (1986). THE INTEGRATION OF DOMINANCE AND SOCIAL BONDING IN PRIMATES. *Quarterly Review of Biology, 61*(4), 459-479. doi: 10.1086/415144

Doyle, Arthur Conan. (1892/2002). *Project Gutenberg's The Adventures of Sherlock Holmes: The Adventure of the Speckled Band* Retrieved from https://http://www.gutenberg.org/files/1661/1661-h/1661-h.htm

Ebitz, R.B., Watson, K.K., & Platt, M.L. (2013). Oxytocin blunts social vigilance in the rhesus macaque. *Proceedings of the National Academy of Sciences, 110*(28), 11630-11635.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science, 302*(5643), 290-292. doi: 10.1126/science.1089134

Elliot, A. J., Eder, A. B., & Harmon-Jones, E. (2013). Approach-Avoidance Motivation and Emotion: Convergence and Divergence. *Emotion Review, 5*(3), 308-311. doi: 10.1177/1754073913477517

Emery, N.J, Capitanio, J.P, Mason, W.A, Machado, C.J., Mendoza, S.P., & Amaral, D.G. (2001). The effects of bilateral lesions of the amygdala on dyadic social interactions in rhesus monkeys (Macaca mulatta). *Behavioral Neuroscience, 115*(3), 515.

Emery, N.J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581-604.

Ethofer, T., Gschwind, M., & Vuilleumier, P. (2011). Processing social aspects of human gaze: a combined fMRI-DTI study. *Neuroimage, 55*(1), 411-419.

Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American Journal of Psychiatry, 164*(10), 1476-1488. doi: 10.1176/appi.ajp.2007.07030504

Fanselow, M.S., & Lester, L.S. (1988). A functional behavioristic approach to aversively motivated behavior: predatory imminence as a determinant of the topgraphy of defensive behavior. In R. C. Bolles & M. D. Beecher (Eds.), *Evolution and Learning* (pp. 185-211). Hillsdale, NJ: Erlbaum.

Farroni, T., Csibra, G., Simion, F., & Johnson, M.H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*(14), 9602-9605.

Farzin, F., Rivera, S.M., & Hessl, D. (2009). Brief report: Visual processing of faces in individuals with fragile X syndrome: An eye tracking study. *Journal of Autism and Developmental Disorders, 39*(6), 946-952.

Fedorenko, E., Hsieh, P., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology, 104*(2), 1177-1194.

Feinstein, J. S., Buzza, C., Hurlemann, R., Follmer, R. L., Dahdaleh, N. S., Coryell, W. H., . . . Wemmie, J. A. (2013). Fear and panic in humans with bilateral amygdala damage. *Nature Neuroscience, 16*(3), 270-272. doi: 10.1038/nn.3323

Feinstein, J.S., Adolphs, R., Damasio, A., & Tranel, D. (2011). The Human Amygdala and the Induction and Experience of Fear. *Current Biology, 21*(1), 34-38. doi: http://dx.doi.org/10.1016/j.cub.2010.11.042

Fernald, R.D. (2012). Social Control of the Brain. *Annual Review of Neuroscience, Vol 35, 35*, 133-151. doi: 10.1146/annurev-neuro-062111-150520

Fitzgerald, D.A, Angstadt, M., Jelsone, L.M, Nathan, P.J., & Phan, K.L. (2006). Beyond threat: amygdala reactivity across multiple expressions of facial affect. *Neuroimage, 30*(4), 1441-1448.

Freese, J. L., & Amaral, D. G. (2005). The organization of projections from the amygdala to visual cortical areas TE and V1 in the macaque monkey. *Journal of Comparative Neurology, 486*(4), 295-317. doi: 10.1002/cne.20520

Freese, J. L., & Amaral, D. G. (2009). Neuroanatomy of the Primate Amygdala. In P. J. Whalen & E. A. Phelps (Eds.), *The Human Amygdala* (pp. 3-42). New York: Guilford Press.

Frith, C.D. (2007). The social brain? *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1480), 671-678.

Frith, C.D., & Frith, U. (2007). Social cognition in humans. *Current Biology, 17*(16), R724-R732.

Gawronski, B., & Cesario, J. (2013). Of Mice and Men: What Animal Research Can Tell Us About Context Effects on Automatic Responses in Humans. *Personality and Social Psychology Review, 17*(2), 187-215. doi: 10.1177/1088868313480096

Goldin, P.R., Manber, T., Hakimi, S., Canli, T., & Gross, J.J. (2009). Neural bases of social anxiety disorder: emotional reactivity and cognitive regulation during social and physical threat. *Archives of general psychiatry, 66*(2), 170-180.

Gothard, K.M, Battaglia, F.P., Erickson, C.A., Spitler, K.M., & Amaral, D.G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *Journal of Neurophysiology, 97*(2), 1671-1683.

Gray, J.A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system.* (2nd ed.). Oxford: Oxford University Press.

Grosenick, L., Clement, T.S., & Fernald, R.D. (2007). Fish can infer social rank by observation alone. *Nature, 445*(7126), 429-432.

Gross, C.T., & Canteras, N.S. (2012). The many paths to fear. *Nature Reviews Neuroscience, 13*(9), 651-658. doi: 10.1038/nrn3301

Hadjikhani, N., Joseph, R.M., Snyder, J., & Tager-Flusberg, H. (2006). Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral Cortex, 16*(9), 1276-1282.

Hagenaars, M.A., Oitzl, M., & Roelofs, K. (2014). Updating freeze: Aligning animal and human research. *Neuroscience and Biobehavioral Reviews, 47*, 165-176. doi: 10.1016/j.neubiorev.2014.07.021

Hamann, S.B, Ely, T.D, Hoffman, J.M., & Kilts, C.D. (2002). Ecstasy and agony: activation of the human amygdala in positive and negative emotion. *Psychological Science, 13*(2), 135-141.

Hamann, S.B., Stefanacci, L., Squire, L.R., Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1996). Recognizing facial emotion. *Nature*.

Hamilton, AF. (2013). Reflecting on the mirror neuron system in autism: a systematic review of current theories. *Dev Cogn Neurosci, 3*, 91-105.

Harrison, L., & Adolphs, R. (2015). The Amygdala and Social Perception. In A. W. Toga (Ed.), *Brain Mapping* (pp. 91-96). Waltham: Academic Press.

Harrison, L., Spezio, M. L., Tyszka, J.M., Elison, J., & Adolphs, R. (2014). *Eyetracking and Neuroimaging Investigations of Direct Face-to-Face Interactions.* Paper presented at the Social and Affective Neuroscience Society Conference, Denver, Colorado.

Hatfield, T., Han, J., Conley, M., Gallagher, M., & Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *The Journal of neuroscience, 16*(16), 5256-5265.

Herry, C, Bach, D.R., Esposito, F., Di Salle, F., Perrig, W.J., Scheffler, K., . . . Seifritz, E. (2007). Processing of temporal unpredictability in human and animal amygdala. *Journal of Neuroscience, 27*(22), 5958-5966. doi: 10.1523/jneurosci.5218-06.2007

Hietanen, J.K., Leppänen, J.M., Peltola, M.J., Linna-aho, K., & Ruuhiala, H.J. (2008). Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. *Neuropsychologia, 46*(9), 2423-2430.

Hofer, P. A. (1973). Urbach-Wiethe disease (lipoglycoproteinosis; lipoid proteinosis; hyalinosis cutis et mucosae). A review. *Acta dermato-venereologica. Supplementum, 53*, 1-52.

Holmes, A., Murphy, D. L., & Crawley, J. N. (2003). Abnormal behavioral phenotypes of serotonin transporter knockout mice: Parallels with human anxiety and depression. *Biological Psychiatry, 54*(10), 953-959. doi: 10.1016/j.biopsych.2003.09.003

Izquierdo, A., & Murray, E.A. (2007). Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. *The Journal of neuroscience, 27*(5), 1054-1062.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., & Smith, S.M. (2012). Fsl. *Neuroimage, 62*(2), 782-790.

Johansen, J. P., Hamanaka, H., Monfils, M. H., Behnia, R., Deisseroth, K., Blair, H. T., & LeDoux, J. E. (2010). Optical activation of lateral amygdala pyramidal cells instructs associative fear learning. *Proceedings of the National Academy of Sciences of the United States of America, 107*(28), 12692-12697. doi: 10.1073/pnas.1002418107

Johnson, M.H. (2005). Subcortical face processing. *Nature Reviews Neuroscience, 6*(10), 766-774.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Keating, C.F, & Keating, E.G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception, 11*(2), 211-219.

Kelly, R.E., Alexopoulos, G.S., Wang, Z., Gunning, F.M., Murphy, C. F., Morimoto, S. S., . . . Hoptman, M.J. (2010). Visual inspection of independent components: Defining a procedure for artifact removal from fMRI data. *Journal of Neuroscience Methods, 189*(2), 233-245. doi: 10.1016/j.jneumeth.2010.03.028

Kennedy, D.P., Glaescher, J., Tyszka, J.M., & Adolphs, R. (2009). Personal space regulation by the human amygdala. *Nature Neuroscience, 12*(10), 1226-1227. doi: 10.1038/nn.2381

Kim, E.J., Horovitz, O., Pellman, B.A., Tan, L.M., Li, Q., Richter-Levin, G., & Kim, J.J. (2013). Dorsal periaqueductal gray-amygdala pathway conveys both innate and learned fear responses in rats. *Proceedings of the National Academy of Sciences of the United States of America, 110*(36), 14795-14800. doi: 10.1073/pnas.1310845110

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology, 19*(1), 52-56.

Kirschbaum, C., Pirke, K-M, & Hellhammer, D.H. (1993). The ‘Trier Social Stress Test’–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1-2), 76-81.

Kliemann, D., Dziobek, I., Hatri, A., Baudewig, J., & Heekeren, H.R. (2012). The role of the amygdala in atypical gaze on emotional faces in autism spectrum disorders. *The Journal of neuroscience, 32*(28), 9469-9476.

Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., & Heekeren, H.R. (2010). Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. *The Journal of neuroscience, 30*(37), 12281-12287.

Klin, A., Jones, W., Schultz, R., Volkmar, R., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry, 59*(9), 809-816.

Kling, A.S., & Brothers, L.A. (1992). The amygdala and social behavior. In J. P. Aggleton (Ed.), *The Amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction.* . New York: Wiley-Liss.

Klüver, H., & Bucy, P.C. (1939). Preliminary analysis of functions of the temporal lobes in monkeys. *Archives of Neurology & Psychiatry, 42*(6), 979-1000.

Kret, M.E., Tomonaga, M., & Matsuzawa, T. (2014). Chimpanzees and humans mimic pupil-size of conspecifics. *PLOS ONE, 9*(8), e104886.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry, 44*(12), 1248-1263. doi: 10.1016/s0006-3223(98)00275-3

Lazarus, R. S. (1991). COGNITION AND MOTIVATION IN EMOTION. *American Psychologist, 46*(4), 352-367. doi: 10.1037//0003-066x.46.4.352

LeDoux, J. (1996). *The Emotional Brain*. New York: Simon and Schuster.

LeDoux, J.E. (2012). Evolution of human emotion: A view through fear. In M. A. Hofman & D. Falk (Eds.), *Evolution of the Primate Brain: From Neuron to Behavior* (Vol. 195, pp. 431-442).

Liao, Y., Acar, Z.A., Makeig, S., & Deak, G. (2015). EEG imaging of toddlers during dyadic turn-taking: Mu-rhythm modulation while producing or observing social actions. *Neuroimage, 112*, 52-60.

Lima, S.L, & Dill, L.M. (1990). Behavioral decisions made under the risk of predation: a review and prospectus. *Canadian Journal of Zoology, 68*(4), 619-640.

Lord, C., Risi, S., Lambrecht, L., Cook Jr, E.H, Leventhal, B.L., DiLavore, P.C., . . . Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659-685.

MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin, 131*(2), 202-223. doi: 10.1037/0033-2909.131.2.202

Málková, L., Gaffan, D., & Murray, E.A. (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *The Journal of neuroscience, 17*(15), 6011-6020.

Mandalaywala, T.M., Parker, K.J., & Maestripieri, D. (2014). Early Experience Affects the Strength of Vigilance for Threat in Rhesus Monkey Infants. *Psychological Science, 25*(10), 1893-1902.

Mason, W.A., Capitanio, J.P., Machado, C.J., Mendoza, S.P., & Amaral, D.G. (2006). Amygdalectomy and responsiveness to novelty in rhesus monkeys (Macaca mulatta): generality and individual consistency of effects. *Emotion, 6*(1), 73.

Matsuzawa, T. (2013). Evolution of the brain and social behavior in chimpanzees. *Current Opinion in Neurobiology, 23*(3), 443-449.

McGaugh, J.L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci., 27*, 1-28.

McNaughton, N., & Corr, P. J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience and Biobehavioral Reviews, 28*(3), 285-305. doi: 10.1016/j.neubiorev.2004.03.005

McNaughton, N., & Corr, P. J. (2009). Central Theories of Motivation and Emotion. In G. C. Bernston & J. T. Cacioppo (Eds.), *Handbook of Neuroscience for the Behavioral Sciences* (Vol. 2, pp. 710-730). Hoboken, NJ: John Wiley & Sons, Inc.

Milgram, S., & Van den Haag, E. (1978). Obedience to authority: Ziff-Davis Publishing Company.

Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., . . . Frith, C. D. (2007). When fear is near: Threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science, 317*(5841), 1079-1083. doi: 10.1126/science.1144298

Mobbs, D., Hagan, C.C., Dalgleish, T., Silston, B., & Prévost, C. (2015). The Ecology of Human Fear: Survival Optimization and the Nervous System. *Frontiers in Neuroscience, 9*. doi: 10.3389/fnins.2015.00055

Mobbs, D., Yu, R., Rowe, J.B.., Eich, H., FeldmanHall, O., & Dalgleish, T. (2010). Neural activity associated with monitoring the oscillating threat value of a tarantula.

Proceedings of the National Academy of Sciences of the United States of America, 107(47), 20582-20586. doi: 10.1073/pnas.1009076107

Murray, E.A. (2007). The amygdala, reward and emotion. *Trends in Cognitive Sciences, 11*(11), 489-497.

Murray, E.A., Izquierdo, A., & Malkova, L. (2009). Amygdala Function in Positive Reinforcement: Contributions from Studies of Nonhuman Primates. In P. J. Whalen & E. A. Phelps (Eds.), *The Human Amygdala* (pp. 82-104). New York, NY: The Guilford Press.

Norris, C.J., Gollan, J., Berntson, G.G., & Cacioppo, J.T. (2010). The current status of research on the structure of evaluative space. *Biological psychology, 84*(3), 422-436.

Nummenmaa, L., Passamonti, L., Rowe, J., Engell, A.D., & Calder, A.J. (2009). Connectivity analysis reveals a cortical network for eye gaze perception. *Cerebral Cortex*, bhp244.

Papworth, S., Milner-Gulland, E. J., & Slocombe, K. (2013). Hunted Woolly Monkeys (Lagothrix poeppigii) Show Threat-Sensitive Responses to Human Presence. *PLOS ONE, 8*(4). doi: 10.1371/journal.pone.0062000

Pardo, J. V, Pardo, P.J., Janer, K.W., & Raichle, M.E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences, 87*(1), 256-259.

Pelphrey, K.A., Morris, J.P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain, 128*(5), 1038-1048.

Pelphrey, K.A., Sasson, N.J., Reznick, J.S., Paul, G., Goldman, B.D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32*(4), 249-261.

Pelphrey, K.A., Viola, R.J., & McCarthy, G. (2004). When strangers pass processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science, 15*(9), 598-603.

Perkins, A. M., & Corr, P. J. (2006). Reactions to threat and personality: Psychometric differentiation of intensity and direction dimensions of human defensive behaviour. *Behavioural Brain Research, 169*(1), 21-28. doi: 10.1016/j.bbr.2005.11.027

Perkins, Adam M., Cooper, Andrew, Abdelall, Maura, Smillie, Luke D., & Corr, Philip J. (2010). Personality and Defensive Reactions: Fear, Trait Anxiety, and Threat Magnification. *Journal of Personality, 78*(3), 1071-1090. doi: 10.1111/j.1467-6494.2010.00643.x

Perrett, DI, Rolls, ET, & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental brain research, 47*(3), 329-342.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*(11), 773-782. doi: 10.1038/nrn2920

Petrinovich, L. (1989). Representative design and the quality of generalization. In L. Poon (Ed.), *Everyday Cognition in Adulthoood and Late Life* (pp. 11-24). Cambridge, UK: Cambridge University Press.

Pitkänen, A., Savander, V., & LeDoux, J.E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends in neurosciences, 20*(11), 517-523.

Pitskel, N. B., Bolling, D. Z., Hudac, C. M., Lantz, S. D., Minshew, N. J., Vander Wyk, B. C., & Pelphrey, K. A. (2011). Brain Mechanisms for Processing Direct and Averted Gaze in Individuals with Autism. *Journal of Autism and Developmental Disorders, 41*(12), 1686-1693. doi: 10.1007/s10803-011-1197-x

Poon, L. (1989). Preface. In L. Poon (Ed.), *Everyday Cognition in Adulthood and Late Life*. Cambridge, UK: Cambridge University Press.

Portas, C.M., Rees, G., Howseman, A.M., Josephs, O., Turner, R., & Frith, C.D. (1998). A specific role for the thalamus in mediating the interaction of attention and arousal in humans. *The Journal of neuroscience, 18*(21), 8979-8989.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., & Petersen, S.E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage, 59*(3), 2142-2154.

Prather, MD, Lavenex, P, Mauldin-Jourdain, ML, Mason, WA, Capitanio, JP, Mendoza, SP, & Amaral, DG. (2001). Increased social fear and decreased fear of objects in monkeys with neonatal amygdala lesions. *Neuroscience, 106*(4), 653-658.

Raper, J., Wallen, K., Sanchez, M.M., Stephens, S.B.Z., Henry, A., Villareal, T., & Bachevalier, J. (2013). Sex-dependent role of the amygdala in the development of emotional and neuroendocrine reactivity to threatening stimuli in infant and juvenile rhesus monkeys. *Hormones and Behavior, 63*(4), 646-658. doi: 10.1016/j.yhbeh.2013.01.010

Redcay, E., Dodell-Feder, D., Mavros, P. L., Kleiner, M., Pearrow, M. J., Triantafyllou, C., . . . Saxe, R. (2013). Atypical brain activation patterns during a face-to-face joint attention game in adults with autism spectrum disorder. *Human Brain Mapping, 34*(10), 2511-2523. doi: 10.1002/hbm.22086

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *Neuroimage, 50*(4), 1639-1647. doi: 10.1016/j.neuroimage.2010.01.052

Riby, D.M., & Hancock, P.J.B. (2008). Viewing it differently: Social scene perception in Williams syndrome and autism. *Neuropsychologia, 46*(11), 2855-2860.

Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience, 6*. doi: 10.3389/fnhum.2012.00143

Roelofs, K., Hagenaars, M. A., & Stins, J. (2010). Facing Freeze: Social Threat Induces Bodily Freeze in Humans. *Psychological Science, 21*(11), 1575-1581. doi: 10.1177/0956797610384746

Rubin, D. C. (1989). Introduction to Part I: The how, when, and why of studying everyday cognition. In L. Poon (Ed.), *Everyday Cognition in Adulthoood and Late Life* (pp. 3-7). Cambridge, UK: Cambridge University Press.

Rutishauser, U., Tudusciuc, O., Neumann, D., Mamelak, A.N., Heller, A.C., Ross, I.B., . . . Adolphs, R. (2011). Single-Unit Responses Selective for Whole Faces in the

Human Amygdala. *Current Biology, 21*(19), 1654-1660. doi: 10.1016/j.cub.2011.08.035

Sambo, C. F., & Iannetti, G. D. (2013). Better Safe Than Sorry? The Safety Margin Surrounding the Body Is Increased by Anxiety. *Journal of Neuroscience, 33*(35), 14225-14230. doi: 10.1523/jneurosci.0706-13.2013

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences, 14*(4), 303-316.

Sander, D., Grandjean, D., & Scherer, K.R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural networks, 18*(4), 317-352.

Sato, W., Kochiyama, T., Uono, S., & Yoshikawa, S. (2008). Time course of superior temporal sulcus activity in response to eye gaze: a combined fMRI and MEG study. *Social Cognitive and Affective Neuroscience*, nsn016.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*(2), 235-239.

Scheele, D., Mihov, Y., Kendrick, K.M., Feinstein, J.S., Reich, H., Maier, W., & Hurlemann, R. (2012). Amygdala lesion profoundly alters altruistic punishment. *Biological Psychiatry, 72*(3), e5-e7.

Scheibler, E., Weinandy, R., & Gattermann, R. (2004). Social categories in families of Mongolian gerbils. *Physiology & Behavior, 81*(3), 455-464. doi: 10.1016/j.physbeh.2004.02.011

Scherer, K.R. (1984). On the nature and function of emotion: a component process approach. In K. R. Scherer & P. Eckman (Eds.), *Approaches to Emotion* (pp. 293-318). Hillsdale, NJ: Lawrence Erlbaum.

Scherer, K.R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Approasial processes in emotion: theory, methods, and research.* (pp. 92-102). New York, NY: Oxford University Press.

Scherer, K.R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion, 23*(7), 1307-1351. doi: 10.1080/02699930902928969

Schilbach, L., Eickhoff, S. B., Cieslik, E. C., Kuzmanovic, B., & Vogeley, K. (2012). Shall we do this together? Social gaze influences action control in a comparison group, but not in individuals with high-functioning autism. *Autism, 16*(2), 151-162. doi: 10.1177/1362361311409258

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). A second-person neuroscience in interaction. *Behavioral and Brain Sciences, 36*(4), 441-462. doi: 10.1017/s0140525x12002452

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(04), 393-414.

Schumann, C. M., & Amaral, D. G. (2006). Stereological analysis of amygdala neuron number in autism. *Journal of Neuroscience, 26*(29), 7674-7679. doi: 10.1523/jneurosci.1285-06.2006

Schumann, C. M., Hamstra, J., Goodlin-Jones, B. L., Lotspeich, L. J., Kwon, H., Buonocore, M. H., . . . Amaral, D. G. (2004). The amygdala is enlarged in children but not adolescents with autism; The hippocampus is enlarged at all ages. *Journal of Neuroscience, 24*(28), 6392-6401. doi: 10.1523/jneurosci.1297-04.2004

Schwandt, M.L., Lindell, S.G., Sjöberg, R.L., Chisholm, K.L., Higley, J.D., Suomi, S.J,. . . . Barr, C.S. (2010). Gene–environment interactions and response to social intrusion in male and female rhesus macaques. *Biological Psychiatry, 67*(4), 323-330.

Shuhama, R., Del-Ben, C. M., Loureiro, S. R., & Graeff, F. G. (2008). Defensive responses to threat scenarios in Brazilians reproduce the pattern of Hawaiian Americans and non-human mammals. *Brazilian Journal of Medical and Biological Research, 41*(4), 324-332.

Siebert, M., Markowitsch, H.J., & Bartel, P. (2003). Amygdala, affect and cognition: evidence from 10 patients with Urbach–Wiethe disease. *Brain, 126*(12), 2627-2637.

Sierra, K. (2007). Face-to-Face Trumps Twitter, Blogs, Podcasts, Video... Retrieved from http://headrush.typepad.com/creating_passionate_users/2007/03/sxsw_interactiv.html

Silk, J.B., Beehner, J.C., Bergman, T.J., Crockford, C., Engh, A.L., Moscovice, L.R., . . . Cheney, D.L. (2010). Strong and consistent social bonds enhance the longevity of female baboons. *Current Biology, 20*(15), 1359-1361.

Silva, B.A., Mattucci, C., Krzywkowski, P., Murana, E., Illarionova, A., Grinevich, V., . . . Gross, C.T. (2013). Independent hypothalamic circuits for social and predator fear. *Nature Neuroscience, 16*(12), 1731-1733. doi: 10.1038/nn.3573

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE, 94*(11), 1948-1962.

Somerville, L.H., Kim, H., Johnstone, T., Alexander, A.L., & Whalen, P.J. (2004). Human amygdala responses during presentation of happy and neutral faces: correlations with state anxiety. *Biological Psychiatry, 55*(9), 897-903.

Spezio, M. L., Loesch, L., Gosselin, F., Mattes, K., & Alvarez, R. M. (2012). Thin-Slice Decisions Do Not Need Faces to be Predictive of Election Outcomes. *Political Psychology, 33*(3), 331-341. doi: 10.1111/j.1467-9221.2012.00897.x

Spezio, M.L., Adolphs, R., Hurley, R.S.E., & Piven, J. (2007). Analysis of face gaze in autism using "Bubbles". *Neuropsychologia, 45*(1), 144-151.

Spezio, M.L., Huang, P.S., Castelli, F., & Adolphs, R. (2007). Amygdala damage impairs eye contact during conversations with real people. *Journal of Neuroscience, 27*(15), 3994-3997. doi: 10.1523/jneurosci.3789-06.2007

Spunt, R.P., & Adolphs, R. (2014). Validating the Why/How contrast for functional MRI studies of Theory of Mind. *Neuroimage, 99*, 301-311.

Spunt, R.P., Satpute, A.B., & Lieberman, M.D. (2011). Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience, 23*(1), 63-74.

Stanley, D.A., & Adolphs, R. (2013). Toward a neural basis for social behavior. *Neuron, 80*(3), 816-826.

Susskind, J.M., Lee, D.H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A.K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience, 11*(7), 843-850. doi: 10.1038/nn.2138

Tamashiro, K. L. K., Nguyen, M. M. N., & Sakai, R. R. (2005). Social stress: From rodents to primates. *Frontiers in Neuroendocrinology, 26*(1), 27-40. doi: 10.1016/j.yfrne.2005.03.001

Tanabe, H. C., Kosaka, H., Saito, D. N., Koike, T., Hayashi, M. J., Izuma, K., . . . Sadato, N. (2012). Hard to"tune in": neural mechanisms of live face-to-face interaction with high-functioning autistic spectrum disorder. *Frontiers in Human Neuroscience, 6*. doi: 10.3389/fnhum.2012.00268

Terburg, D., Morgan, B. E., Montoya, E. R., Hooge, I. T., Thornton, H. B., Hariri, A. R., . . . van Honk, J. (2012). Hypervigilance for fear after basolateral amygdala damage in humans. *Translational Psychiatry, 2*. doi: 10.1038/tp.2012.46

Tinbergen, N. (1951). The study of instinct.

Todorov, A., Baron, S.G., & Oosterhof, N.N. (2008). Evaluating face trustworthiness: a model based approach. *Social Cognitive and Affective Neuroscience, 3*(2), 119-127. doi: 10.1093/scan/nsn009

Todorov, A., & Uleman, J.S. (2002). Spontaneous trait inferences are bound to actors' faces: evidence from a false recognition paradigm. *Journal of Personality and Social Psychology, 83*(5), 1051.

Tsao, D.Y, Freiwald, W.A, Tootell, R.B.H., & Livingstone, M.S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311*(5761), 670-674.

Tye, K. M., Prakash, R., Kim, S. Y., Fenno, L. E., Grosenick, L., Zarabi, H., . . . Deisseroth, K. (2011). Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature, 471*(7338), 358-362. doi: 10.1038/nature09820

Tyszka, J.M.. (2015). MR Gaze. v0.6.6. 2015, from https://github.com/jmtyszka/mrgaze

Utevsky, A.V., Smith, D.V., & Huettel, S.A. (2014). Precuneus is a functional core of the default-mode network. *The Journal of neuroscience, 34*(3), 932-940.

van der Zwaag, W., Da Costa, S.E., Zürcher, N.R., Adams Jr, R.B., & Hadjikhani, N. (2012). A 7 tesla FMRI study of amygdala responses to fearful faces. *Brain topography, 25*(2), 125-128.

Wetherby, Amy M, Woods, Juliann, Allen, Lori, Cleary, Julie, Dickinson, Holly, & Lord, Catherine. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders, 34*(5), 473-493.

Whalen, P.J., & Phelps, E.A. (2009). *The human amygdala*: Guilford Press.

Whalen, P.J. (2007). The uncertainty of it all. *Trends in Cognitive Sciences, 11*(12), 499-500. doi: 10.1016/j.tics.2007.08.016

Williams, K.D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior research methods, 38*(1), 174-180.

Williams, K.D. (2007). Ostracism *Annual Review of Psychology* (Vol. 58, pp. 425-452).

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598. doi: 10.1111/j.1467-9280.2006.01750.x

Willis, J., & Todorov, A. (2006). First impressions making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G.R., & Vogeley, K. (2010). It's in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, nsq024.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*(3), 277-283. doi: 10.1038/nn816

Yamamoto, S., Humle, T., & Tanaka, M. (2012). Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals. *Proceedings of the National Academy of Sciences, 109*(9), 3588-3592.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods, 8*(8), 665-U695. doi: 10.1038/nmeth.1635

Yilmaz, M., & Meister, M. (2013). Rapid Innate Defensive Responses of Mice to Looming Visual Stimuli. *Current Biology, 23*(20), 2011-2015. doi: 10.1016/j.cub.2013.08.015

Yun, K., Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific reports, 2*.