Title

Validating the Judgement Bias Task as a Reliable Proxy for Affective Processing in Macaca mulatta

Author

Cogo, Patrick Raymond

Publication Date

2022

Validating the Judgement Bias Task as a Reliable Proxy for Affective Processing in *Macaca mulatta*

By

PATRICK RAYMOND COGO
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Animal Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Eliza Bliss-Moreau, Chair

Kristina Horback

Brenda McCowan

Committee in Charge

2022

ACKNOWLEDGMENTS

It is nearly impossible to succinctly thank all those who helped shape me into the scientist I am today. To avoid writing an additional thesis filled with acknowledgments alone, I therefore choose to thank a select few instrumental in my growth over these two years. First and foremost, I want to express all my love and thanks to the 18 juvenile rhesus macaques I worked with over these past two years. I cannot conduct animal behavior research without animals, and these monkeys were the best study subjects a scientist could have asked for! These monkeys gave so much of themselves to science, and I will always keep them close to my heart. I also want to express my deepest gratitude to UC Davis, a school I had always dreamed of attending and a place that molded me from a pre-veterinary student into an applied ethologist through a world-class education. This university ignited my passion for animal behavior and welfare that I will continue to stoke for the rest of my life. To my mother and stepfather, you provided unending support and love over the years, allowing me to pursue my wildest dreams. To my father, you have had to answer far too many stressed-out phone calls to provide a calming presence that allowed me to keep moving forward. I want to express a deep appreciation to Dr. Eliza Bliss-Moreau and Dr. Florent Pittet for seeing my potential and working alongside me to unearth the scientist inside me I never knew I could be. The journey has been filled with ups and downs, but I am glad I had you both to weather the storm with me. I also want to extend my gratitude to my committee members, Dr. Kristina Horback and Dr. Brenda McCowan, for their intellectual contributions in developing this thesis. I hope to follow in both of your footsteps in using science to enhance the health and well-being of animals! Thank you to the Bliss-Moreau Lab members that have kept me sane these past two years. I could not have asked for a better lab family! Lastly, I want to acknowledge (not thank) this global pandemic for showing me that I have the resilience to persevere through any hardship and emerge stronger on the other side.

ABSTRACT

Accurate measures of affective processing in animals are essential to advancing fields ranging from animal welfare science to neuroscience and psychology. However, objective measures of affect are currently unavailable, leading scientists to rely on measurable changes in decision-making (judgement biases) as a proxy for affective processing in various species. Yet, minimal research has been conducted on the connection between decision-making and affect in rhesus macaques, with a lack of focus on validating the judgement bias task as a measure of both affective traits and states. Here, research was conducted to validate whether the judgement bias task captures both stable individual differences in baseline biases, indicative of trait affect, and momentary negative shifts in biases following a negative threat induction, suggestive of shifts in affective states. Additionally, research aimed to determine whether judgement biases were related to affective reactivity, a conventional measure of affective processing obtained from the Human Intruder Test. Results showed temporal stability in individual differences in baseline judgement biases, suggesting the judgement bias task provides a suitable index of trait affect within individuals. Consistently negative baseline judgement biases also predicted a higher overall susceptibility to affective reactivity during the Human Intruder Test. A reduction in response post threat induction compared to baseline measures was also found, reflecting the capturing of categorically negative affective states within individuals. During post threat induction judgment bias testing, shifts in reward valuation were also related to a higher human-directed reactivity during the prior Human Intruder Test. A lack of context stability in individual differences in judgment biases and no relationship between overall affective reactivity and judgement biases demonstrated that baseline trait judgement biases alone could not predict how influential threat induction would be on an individual's affective processing and thus decision-making. Results validated the judgement bias task as a proxy measure of affective traits and states in captive rhesus macaques. However, additional recommendations must be considered to further our understanding of the link between affect and cognition in non-human primates.

TABLE OF CONTENTS

1. Comprehensive Literature Review

1.1 Role of Affective Processing in Measures of Animal Welfare

The foundation of animal welfare science is built upon three pillars (Fraser et al., 1997b): (1) species-typical behavior; focused on natural behaviors and the utilization of natural adaptations and abilities, (2) biological functioning; focused on optimal physiological functioning through the reduction of injury, disease, and deformities, and (3) affective processing; focused on the global states of an animal that arise in response to allostasis perturbations caused by external and internal stimuli. While some scientists argue that measures of animal welfare should rely on a multidimensional approach that utilizes the three pillars of study in conjunction (e.g., Dawkins, 1990; Fraser et al., 1997; Fraser & Duncan, 1998), most studies have relied upon measures of species-typical behavior and biological functioning as metrics of welfare due to the challenge of developing reliable measures of affective processing. This challenge has arisen partly due to a struggle to provide objective measures of subjective experiences in animals, as well as due to disagreements over whether and which animals have the capacity for such experience (e.g., Dawkins, 1993; Griffin, 1998; Macphail, 1998; Panksepp, 2005; Paul et al., 2020). Despite the challenge of developing objective measures of affective processing, an increased emphasis in the non-human animal (hereafter, animal) welfare literature has been placed on understanding how the housing and management conditions of captive animals acutely and/or chronically influences affective processing. Affective processing can be defined as internal conditions characterized by both valence (or hedonics) and arousal that result from the integration of incoming sensory information from the external world with homeostatic information from the body (Barrett & Bliss-Moreau, 2009). An assumption in the animal welfare literature is that the promotion of positive affect (e.g., pleasure) and reduction of negative affect (e.g., displeasure) leads to improved welfare in captive animals (Boissy et al., 2007; Dawkins, 1990; Mendl, 2001). Therefore, to progress the field of animal welfare science in captive animals, it is imperative to develop valid and reliable proxies of affect that allow for a more comprehensive and evidence-based understanding of animal well-being.

1.2 Affective Processing in Humans and Non-Human Animals

Affect is theorized as both a physiological and psychological process, with the physiological process ultimately producing the psychological process (Russell, 2003; Russell & Barrett, 1999). Physiologically, affective processing results from the integration of incoming sensory information from the external world with homeostatic and interoceptive information from the body to create a neurophysiological barometer of an individual’s relation to their environment (Barrett, 2006; Barrett & Bliss-Moreau, 2009). Psychologically, affective processing is structured into two independent dimensions, a degree of pleasantness (i.e., valence) and a degree of activation (i.e., arousal), that combine in an integral fashion to generate a mental state (Barrett & Bliss-Moreau, 2009; Russell, 2003; Russell & Barrett, 1999).

Affect can be conceptualized as both a trait and a state. An affective trait reflects relatively stable and individual-specific patterns of behavior, action, and experience that are not bounded in time and represent the baseline affective processing from which subsequent momentary affective states arise (Bliss-Moreau & Rudebeck, 2021; Strelau, 2001). Unlike traits, affective states are temporally bound and vary across time as a function of the situation the animal encounters, with states reflecting the momentary status of an individual based on changes in the environment and/or in the individual's physiology (Barrett & Bliss-Moreau, 2009; Bliss-Moreau & Rudebeck, 2021; Russell, 2003, 2005; Russell & Barrett, 1999). An animal’s momentary affective state can be psychologically described and mapped by a single point on the two-dimensional space of valence and arousal (Russell, 2003). Momentary shifts in affect occur in response to environmental cues or changes in an animal's physiology, providing the individual with everchanging information about how to respond in a given context (Bliss-Moreau, 2017). Furthermore, the repeated generation of these shifts in affective states can cause shifts in baseline trait affect, allowing an individual to adapt to erratic environmental changes and respond accordingly (Bliss-Moreau & Rudebeck, 2021).

Temporality is also vital in differentiating affect-related phenomena, specifically affective states. As previously discussed, while traits are not bounded in time, affective states are temporally bound. However, *how* affective states are bound in time differs, with some states being momentary while others can be temporally extended (Bliss-Moreau & Rudebeck, 2021). Momentary affective states arise in response to brief presentations of affect-related stimuli to produce ephemeral changes to biology and behavior concurrent with the presentation. Alternatively, temporally extended affective states (TEAS) are drawn out in response to stimuli due to ongoing allostatic processes as an individual attempts to adapt to the presentation (Bliss-Moreau & Rudebeck, 2021). Thus, TEAS in humans have been theorized to provide information about our ability to adapt to environmental challenges and threats (Morris, 1992). However, TEAS need not reflect direct adaptation but instead may occur once the stimulus of initial affective response is no longer present and thus can be thought of as "objectless" or free-floating and seemingly lacking a cause (Gross, 1998; Larsen, 2000; Russell, 2005). Temporally extended affective states (TEAS) are also referred to as moods in humans (Gross, 1998; Larsen, 2000; Parkinson et al., 1996; Russell, 2005). Due to the lack of clarity on the extent to which animals consciously access and asses their internal states, researchers have shifted away from discussing animal mood (which requires conscious experience) to focusing on temporally extended affective states (TEAS), which are moods without conscious experience and therefore can be present in organisms without the capacity for consciousness (Bliss-Moreau & Rudebeck, 2021).

While humans can be consciously aware of their affective states in generating emotional experience (reviewed in Russell, 2005), the extent to which animals consciously access and assess their internal states is unclear. Evidence of metacognitive abilities in rhesus macaques suggests that non-human primates may utilize and act on internal state information (Charbonneau et al., 2022; Kornell et al., 2007; Middlebrooks & Sommer, 2012; Rosati & Santos, 2016; Templer & Hampton, 2012). However, there is uncertainty about whether animals possess secondary reflexive awareness of internal states (see Lambie & Marcel, 2002 for discussion on 'levels' of consciousness). Nevertheless, it is theorized that the capacity for affect is likely conserved across species with nervous systems and almost certainly amongst mammals (Bliss-Moreau, 2017). Even in single-cell organisms with simple nervous systems, such as bacteria, affective value guides behavior, with the microorganisms retracting from harmful chemicals and moving towards nutritionally valuable chemicals (Macnab & Koshland, 1972). The omnipresent nature of affect across the animal kingdom allows for the possibility of studying the internal states of species other than humans. As such, studying a general model of affective processing in animals can occur without making assumptions about the extent to which they have secondary reflexive awareness that would allow for consciousness. For this reason, we refer to indicators of affective processing and affect, but not affective experience, when discussing an animal’s internal characteristics.

1.3 Alternative View of Emotions in Animals

Before discussing the indicators utilized by scientists to assess affective processing in animals, I must first address the classical view of emotion, which has been continually adopted by many animal affect and animal welfare researchers as they strive to develop indicators of affect. The necessity of addressing the classical view of emotion involves the differences in assumptions of the presence of emotions in animals, which dictates how indicators are used to interpret internal states and have caused a theoretical and conceptual divergence between the classical theory and the one adopted by this thesis. The classical view of emotion argues that an "emotion" is an internal state centrally located in the nervous system that is triggered by specific intrinsic and extrinsic stimuli (D. J. Anderson & Adolphs, 2014; Ekman & Cordaro, 2011; Panksepp, 2011; Tracy & Randles, 2011). In this classical view, scientists assume emotions are biologically hardwired and produce distinct and reproducible behavioral and physiological outputs that are specific indicators of evolutionarily conserved, discrete emotions (Tracy & Randles, 2011). This view allows scientists to utilize these distinct signatures of emotion, first proposed in humans, as homologous across species and able to be applied to animals as a measure of emotional experience. However, meta-analytic evidence has demonstrated that the distinct signatures of emotion suggested by the classical view of emotion, such as facial behavior (Barrett et al., 2019; Durán & Fernández-Dols, 2021; Nelson & Russell, 2013) and "fingerprints of emotion" in the autonomic nervous system (Kreibig, 2010; Lindquist et al., 2013; Siegel et al., 2018) and central nervous system (Guillory & Bujarski, 2014; Kober et al., 2008; Lindquist et al., 2012), do not consistently map to emotions in humans. Therefore, given the lack of distinct and reproducible signatures of emotion in humans, this thesis avoids the categorization of animal behavioral and biological outputs into human emotion categories. Instead, this thesis adopts an alternative theoretical framework for conceptualizing an animal's internal states that does not restrict the features of animal emotion to that of human emotion.

The theory of constructed emotion adopted by this thesis was first developed in humans (Barrett, 2006c; Russell, 2003) and translated to animals as a model for emotion (Bliss-Moreau, 2017). The foundation of this theory is that emotions are not hardwired or discrete but instead emerge from a combination of psychological 'building blocks' (Barrett & Bliss-Moreau, 2009; Russell, 2003). While theories of constructed emotion differ in which 'building blocks' allow emotions to emerge, all theories agree that affect forms the foundation on which emotions are 'built' (Barrett, 2006c; Barrett et al., 2007; Barrett, 2017a, 2017b; Barrett & Bliss-Moreau, 2009; Russell, 2003; Russell & Barrett, 1999). Thus, it is theorized that emotions are broken down into fundamental components, with affect serving as a foundational component that is distinct from, yet linked to, emotion (Russell & Barrett, 1999). This theory of constructed emotion shifts the focus on understanding animal emotion away from developing specific measures of discrete emotions toward investigating the fundamental components of emotion using physiological and behavioral measures (Bliss-Moreau, 2017). As such, it shifts the focus away from measures of animal emotion and instead towards measures of animal affective processing. Due to affect being omnipresent in animals and at the core of emotional construction, it is thus imperative to develop reliable indices of affective processing to begin to develop an understanding of the emergence of animal emotions free from anthropocentric influences.

1.4 Indicators of Affective Processing

Despite the ever-growing interest in several fields for developing direct measurements of affective processing in animals, as mentioned previously, there is the ongoing challenge of being unable to develop these direct measures in animals. In humans, the 'gold standard' indicator for subjective experience is linguistic reporting of experienced emotion; however, even this standard lacks precision and is subject to variance in whether a person focuses on the valence- or arousal-based properties of their feelings (Barrett, 2004, 2006b; Barrett & Bliss-Moreau, 2009). Animals are unable to provide this linguistic reporting of affective processing. Nevertheless, physiological and behavioral continuities between humans and animals have been considered in the development of physiological and behavioral measurements in an attempt to provide objective measures through linking particular measurements to affective space (Paul et al., 2005).

1.4.1 Physiological Indicators of Affect

Affective processing represents the continuous stream of homeostatic feedback resulting from somatovisceral, kinesthetic, proprioceptive, and neurochemical fluctuations within the core of an individual's body that delivers affective information about their current relation to the world (Barrett, 2006a; Barrett & Bliss-Moreau, 2009). As such, an animal's affective state is realized through physiological information that researchers can measure as a possible proxy for such states. Physiological indicators of affect have primarily focused on negative affective processing and have overlapped with commonly used measures of stress physiology (Paul et al., 2005). As a result, circulating hormones in the hypothalamic-pituitary-adrenal (HPA) and sympathetic-adrenal-medullary (SAM) axes (Boissy, 1995; Bradley & Lang, 2000; Davis & Whalen, 2001; Ledoux, 1998) have commonly been used as physiological measures of affect (Blanchard et al., 1998; Boissy, 1995; Bradshaw et al., 1996; Landgraf & Wigger, 2002; McCarty et al., 1988). Measures of the cardiac system, specifically the sympathetic pre-ejection period and parasympathetic respiratory sinus arrhythmia, have also been used as physiological indicators of affect in rhesus macaques (Bliss-Moreau et al., 2013). In addition to neuroendocrine and cardiac measures, researchers have also used measures of the skin, specifically conductance and temperature, as an indicator of affect (Parr, 2001; Süer et al., 1998).

Despite their widespread use, physiological indicators are not free of issues in reliability and interpretation. The use of endocrine measures requires that the influence of individual and diurnal variability on baseline hormone levels be taken into account, as well as temporal differences in hormonal response to a stressor (Ruis et al., 1997; Schrader & Ladewig, 1999; von Borell & Ladewig, 1992). Additionally, similar physiological responses can also arise from different affective experiences, providing a challenge for interpretation. For example, cortisol levels may increase in high arousal states of differing valence, like in response to threat or sexual activity (Broom & Johnson, 1993; Rushen, 1986, 1991; Toates, 1995). Furthermore, heart rate elevation may simply indicate a relative increase in activity or result from the anticipation of either a punishing or rewarding event (Baldock et al., 1988; Baldock & Sibly, 1990; Marchant et al., 1995; Paul et al., 2005). Taken together, the interpretative power of physiological indicators may allow for a valid measure of arousal, however it is less clear whether they can discern the affective valence (i.e., positive or negative) of an affective state. Additionally, due to the increased accessibility and greater utility, behavioral indicators will be the focus of this thesis.

1.4.2 Behavioral Indicators of Affect

A variety of behavioral measures have been employed by different fields (animal welfare, psychopharmacology, neuroscience, etc.) as a means of providing indices of affect in animals. While a brief overview of the numerous behavioral measures assumed to accompany specific affective states will be provided for completeness, the focus of this thesis relies on the use of measurements of responses to an unfamiliar human during a Human Intruder Test as an assessment of affective reactivity and proxy for affective processing.

Numerous studies have strived to link patterns of behavior to measures of emotional expression in humans and animals. In humans, patterns of facial behaviors have often been used as a measure of emotional expression (e.g., Ekman et al., 1987; Keltner et al., 2003; Shariff & Tracy, 2011). From these studies in humans, researchers have strived to develop facial "expressions" of emotion in other mammals such as orangutans (Caeiro et al., 2013), chimpanzees (Vick et al., 2007), gibbons (Waller et al., 2012), rhesus macaques (Parr et al., 2010), horses (Wathan et al., 2015), dogs (Waller et al., 2013), and cats (Caeiro et al., 2017). However, evidence has demonstrated a lack of mapping between facial behaviors and emotion categories in humans (Barrett et al., 2019) and non-human primates (Bliss-Moreau & Moadab, 2017), questioning facial behavior as a measurement of emotional expression in animals. However, facial behaviors may serve as indicators of affect. Some facial behaviors in rhesus macaques appear to consistently map with affective valence, such as a relaxed open-mouth face only occurring in positive states while threat faces never occurring during positive states (Bliss-Moreau & Moadab, 2017). Additionally, a large number of studies in mammals, including rats, pigs, cattle, and baboons, have used vocal behaviors as a proposed behavioral index of affective valence (Blumberg & Sokoloff, 2001; Briefer et al., 2022; Knutson et al., 2002; Manteuffel et al., 2004; Panksepp & Burgdorf, 2003; Rendall, 2003; Watts & Stookey, 2000; Weary & Fraser, 1997). While consistent patterns of behavior can indicate underlying affective processing, there is not enough concrete evidence to suggest that patterns of behavior provide a measure of emotional expression in either humans or animals.

Tasks have also been developed in laboratory settings that utilize measurements of unconditioned responses to unfamiliar contexts or specific stimuli as affective processing indices. Studies utilizing the open field test and elevated plus maze in rodents (e.g., Landgraf & Wigger, 2002; Ramos & Mormède, 1997) rely on measures of unconditioned responses to unfamiliar contexts as indicators of negatively valenced affective states. These tests measure general locomotor activity and willingness to explore as an indicator of high arousal, negatively valenced states, and are based on the general aversion of rodents and other animals to open spaces. Laboratory-based tasks have also been developed that utilize unconditioned responses to specific stimuli as a measure of an individual's capacity for affective reactivity to negative contexts, such as the object responsivity test and Human Intruder Test (Bliss-Moreau et al., 2010, 2011; Gottlieb & Capitanio, 2013; Kalin & Shelton, 1998; Mason et al., 2006). The object responsivity test involves the placement of objects of varying complexity to evaluate changes in an animal's responses with variation in a stimulus' affective value (e.g., Bliss-Moreau et al., 2011). The Human Intruder Test involves the presentation of an unfamiliar human in different spatial conditions (near or far) and threat-related contexts (direct eye contact or averted gaze) relative to a caged macaque (Gottlieb & Capitanio, 2013). The Human Intruder test is predicated on the general principle that rhesus monkeys utilize direct eye contact as a threatening challenge to conspecifics (De Waal & Yoshihara, 1983; Perrett & Mistlin, 1990; Sato & Nakamura, 2001), allowing for direct eye contact by an unfamiliar human to be used as a threat induction to analyze affective behaviors produced in response. The summation of affect-related behaviors produced by an animal during these tasks are used to quantify affective reactivity, a valuable indicator of individual variation in affective responsiveness to stimuli (Bliss-Moreau & Moadab, 2016). Measurements from these tasks can then be used to evaluate an individual's affective state, with more robust shifts in affective state being reflected by a high affective reactivity score. As such, affective reactivity can be paired with other measures to allow for a clearer picture of affective states in animals. In this thesis, the Human Intruder Test will not only be used for the generation of negative affective states via threat induction, but also as a measure of affective reactivity to pair with an assumed cognitive proxy of affect as a method to validate this cognitive measure as an indicator of affective processing.

1.5 Cognitive Indicators of Affect

In addition to physiological and behavioral indicators, researchers have incorporated measures of cognition as a proposed indicator of affect. Cognition can be defined as the mechanisms by which animals acquire, process, store, and act upon information from the environment (Shettleworth, 2009). While this definition of cognition encompasses several types of information processing, one focus of cognitive indicators of affect has been on shifts in appraisal and action selection. The link between cognition and affect in humans has been theorized to occur in both directions, with cognitive appraisals influencing consciously felt affect (Mathews & MacLeod, 2002) and affective manipulations generating changes to cognitive processing (Forgas, 2001; Hinde, 1985). Appraisals can be thought of as the continual assessments made by an individual of stimuli and situations as they arise. While some research has focused on cognitive appraisals in humans (Frijda, 1986; Lazarus, 1991; Oatley & Johnson-Laird, 1987; Ortony et al., 1990; Scherer, 1984) as a presumed measure of cognition influencing emotions, appraisal theories have primarily been developed within the human context utilizing linguistic reporting of subjective experience (Paul et al., 2005; Scherer, 1999). Additionally, there is a lack of consensus on whether appraisal processes are strictly cognitive or involve a more complex cognitive-affective interaction (Forgas, 2001; Lane et al., 2000; Lazarus, 1999; Panksepp, 2003; Zajonc, 1980, 2000). Therefore, the focus of this thesis will stem from research demonstrating the influence of reported emotions and affect on action selection in humans as a starting point for animal research. Action selection is conceptualized as the decision-making generated by cognitive processes, including attention, memory, and judgement. Due to affect modulating sensory processing, psychological processes that utilize sensory information will have an affective quality (Duncan & Barrett, 2007). As a result, affect influences cognitive processes, like decision making (e.g., Bechara, 2004; Janis & Mann, 1977; Tversky & Kahneman, 1979), to make external information about the surrounding world personally relevant and allow for the individual to decide how to respond. Thus, affect generates measurable changes in cognition (biases) in humans that have been utilized as an indicator of affective processing and emotional experience.

1.5.1 Use of the Cognitive Bias Task in Humans

The cognitive bias task is inspired by human psychological studies which have demonstrated that the valence of an individual's affective processing influences cognitive processes including attention, memory, and judgement (Bower, 1981; A. K. MacLeod & Byrne, 1996; Mathews & MacLeod, 1994; Paul et al., 2005). Participants in these studies are either clinically diagnosed with an affective disorder, score low or high on a particular affective trait, or have an affective state induced experimentally (Paul et al., 2005). Results from this task demonstrate that individuals in negatively valenced states show an increased tendency to orient visual attention towards threatening stimuli (e.g., Kindt & Van Den Hout, 2001; Mathews & MacLeod, 1994; Mogg & Bradley, 1998), an increased retrieval of accessible memories related to negative events (e.g., Bower, 1981; Burke & Mathews, 1992; Mineka et al., 1998), and a greater propensity to overestimate the likelihood of negative events and judge ambiguous stimuli as predicting a negative outcome (e.g., Eysenck et al., 1991; A. K. MacLeod & Byrne, 1996; Nygren et al., 1996; Wright & Bower, 1992) compared to individuals self-reporting positive states. Taken together, these findings suggest that measurements of biases in cognitive processing closely emulate reported affective processing in humans and therefore may provide a reliable proxy of affective processing in not only humans, but animals as well.

1.5.2 First Use of the Judgement Bias Task in Animals

Studies in humans have illustrated that the influence of affect on cognition can occur in any cognitive domain (e.g., attention, memory, judgment) and lead to measurable biases that can serve to index affective processes. While a few scientists have developed attention bias tasks (e.g., Crump et al., 2018; Howarth et al., 2021) and memory bias tasks (e.g., Burman & Mendl, 2018) for use in animals as indices for affect, the central area of research has focused on the judgement bias task, which evaluates how decision-making in ambiguous contexts changes as a function of presumed changes in affect (Mendl et al., 2009). In striving to elucidate the influence of affect on decision-making in animals, the first published judgement bias study in animals tested rats on a judgement bias task designed to systematically assess the impact of housing conditions on rats' affective processing (Harding et al., 2004). Rats were trained to press a lever when hearing a 4 kHz to receive a food pellet and refrain from pressing the lever when hearing a 2kHz tone to avoid aversive white noise. Rats were then allocated to either 'unpredictable' housing, in which zero to two negative interventions (i.e., tilting of cage, pairing with an unfamiliar individual, reversed light/dark cycle, damp bedding) were made at random on any one day, or 'predictable' housing, where no interventions made, during nine days of training. After nine days of differing housing conditions, rats participated in testing sessions where both training tones and ambiguous tones with frequencies intermediate to training tones (i.e., 2.5, 3, 3.5 kHz) were presented to the animals, and lever-press responses and response latency were recorded. The scientists found that rats in 'unpredictable' housing were slower to press the lever and tended to show fewer lever responses to the near-positive ambiguous stimulus and the rewarded signal stimulus itself compared to control animals. The results linked the 'unpredictable' housing, an experimental modulation presumed to cause a negative shift in affective state, to a measurable reduction in response indicative of reduced anticipation of the reward (food pellet). From this study, the judgement bias task emerged, and the link between shifts in decision-making and changes in affective processing began to be explored in an array of species.

1.5.3 Overview of the Basic Judgement Bias Task in Animals

In animals, the judgement bias task occurs across two distinct phases: a training and a testing phase. During the operant discrimination training phase, animals are presented with a "positive" signal stimulus (S+) and trained to perform a behavioral response (e.g., touch stimuli) in order to experience a positive event (e.g., food reward), and then presented with a "negative" or null signal stimulus (S-) and trained to suppress their behavioral response (e.g., do not touch stimuli) to avoid an adverse event (e.g., white noise, unpalatable food, time out from testing). This type of task, where the animal's options are to either perform a response or suppress it, is frequently referred to as a 'go/no-go' task (Mendl et al., 2009). During the testing phase, both training signal stimuli are presented along with ambiguous probe stimuli that possess characteristics intermediate to both training stimuli. For example, if the "positive" signal stimulus was a black square and the "negative" signal stimulus was a white square, then the ambiguous probe stimuli would be squares of varying degrees of grey color.

During a presentation of an ambiguous stimulus, the animal must compare the stimulus' characteristic to the characteristics of the trained signal stimuli held in memory. The animal either associates the characteristic of the stimulus with a memory of the S+ stimulus and a higher likelihood of a rewarding outcome, or the S- stimulus and a higher likelihood of a punishing outcome. However, the characteristic of the stimulus alone is insufficient in allowing the animal to discriminate between the two outcomes. Instead, an animal's affective processing provides additional exteroceptive and interoceptive information that facilitates decision-making in the ambiguous situation. Affect integrates sensory information from the presented stimulus with information about the animal's internal environment to direct an individual towards decisions that optimize rewards and minimize punishments (Mendl & Paul, 2020). However, during the judgement bias task, an animal's current affective state is not known; therefore, to evaluate whether the judgement bias task is a valid proxy of affective states, scientists must first determine whether the animal is in a relatively positive or negative state to link decision-making results to presumed affective processing. In most judgement bias studies, researchers use experimental modulations presumed, or previously tested using physiological and behavioral measures as an index for affect, to cause momentary shifts in affective state and prime animals for judgement bias testing (Mendl et al., 2009).

During the testing phase, both an animal's expression or suppression of touch behavior and response latency are collected for all stimuli. Data is used to understand whether subjects perform behavioral responses indicative of anticipation of a rewarded outcome (e.g., touch ambiguous stimuli as if similar to positive stimulus and indicative of reward) or of a punished outcome (e.g., do not touch ambiguous stimuli as if similar to negative stimulus and predictive of punishment). These studies aim to determine the behavioral changes that occur in animals in a particular affective state to develop a reliable indicator of such a state using the judgement bias task. For example, this allows scientists to utilize a judgement bias task to test whether an experimental modulation used to induce a presumed negative shift in affective state does indeed cause a decrease in the proportion of responses to ambiguous stimuli compared to control animals.

1.6 Historical Use of the Judgement Bias Task

Since the original report in rats (Harding et al., 2004), the majority of judgement bias studies have utilized the ‘go/no-go’ task (N=57 ‘go/no-go’ tasks; Lagisz et al., 2020) as a proxy measurement for affective states. The task has been implemented in a wide range of taxa, from mammals to birds to insects (Bateson et al., 2011; Lagisz et al., 2020; Mendl et al., 2009; Neville et al., 2020; Paul et al., 2005), to evaluate how affect, and in turn welfare, are impacted by common manipulations inherent to captivity. Typically in the judgement bias task, experimental housing manipulations are used to induce ostensible shifts in underlying affective processing, either negatively through unpredictable housing conditions (e.g., Harding et al., 2004) or positively through the implementation of environmental enrichment (e.g., Bateson & Matheson, 2007). To ensure effects of housing conditions occur, animals will often remain in experimental housing for 1-3 weeks before testing commences (e.g., Brydges et al., 2011; Matheson et al., 2008; Stephenson & Haskell, 2020). Scientists then use the judgement bias task to make predictions about responses following the chronic affect-related manipulations to assess the impacts of husbandry procedures on affective processing and thus welfare. However, while scientists have historically utilized this task as proxy measurements for momentary states, due to the chronic application of treatments, it is unclear whether results may also reflect temporally-extended affective states and/or affective traits. Animals may undergo an immediate shift in affective state only in experimental manipulations; however, the prolonged application of the treatment may cause development within the animal to also be affected and induce long-term changes in decision-making under ambiguity once testing has commenced. Additionally, while longer-term affect manipulations have shown TEAS-congruent shifts in judgement bias (e.g., Bateson & Matheson, 2007; Harding et al., 2004), studies using short-term, acute stressors have shown inconsistent patterns of response (e.g., Doyle et al., 2010; Sanger et al., 2011). Thus, this over-reliance on chronic housing and enrichment manipulations in judgement bias tasks highlights the need for a greater emphasis on the utilization of shorter-term, acute stressors in the task to better understand the effect of modulation on decision-making during the judgement bias task. Through this, we can better understand whether judgement bias results capture momentary changes in affective states and temporally stable individual differences in trait affect, allowing for a more reliable proxy of overall affective processing.

1.6 Use of the Judgement Bias Task as a Proxy for Trait Affect

A question of increasing importance for understanding an animal's affective processing is whether the judgement bias task may provide a suitable index of affective traits within an individual. Affective traits reflect general patterns of an individual's affective processing, and thus can also theoretically reflect temporally stable judgment biases stemming from the long-lasting influence of affective processing on decision-making patterns. Studies in humans have consistently shown that about 80% of the population are optimistic about their judgement of future events (Carver et al., 2010; Sharot, 2011), often rating their expectations higher than reality (Shepperd et al., 2015). However, despite this stable positive judgement bias in most of the population, individuals reporting higher anxiety and/or depression typically showed less optimism than otherwise healthy subjects (Aue & Okon-Singer, 2015; Scheier et al., 1994). Given the link between trait optimism and affect in humans, scientists have begun to conduct longitudinal judgement bias tasks that involve repeated testing over multiple days to understand the temporal stability of individual differences in judgement bias and its link to affective processing. Across three days of judgement bias testing, house mice showed consistent individual differences in the relative latencies to approach and consume an ambiguous reward, a judgement bias index (Verjat et al., 2021). Similarly, captive dolphins were found to have stable individual differences in their judgement of ambiguous cues over three consecutive testing days (Clegg et al., 2017). In rats, after a series of 10 consecutive judgement bias tests performed in 1-week intervals, a temporally stable increase in the proportion of negative responses to ambiguous cues emerged in specific individuals (Rygula et al., 2013). Repeated judgement bias testing in a small sample of captive chimpanzees (n = 3) also found individual differences in judgement bias that remained stable across five test sessions spanning 1 to 2 weeks (Bateson & Nettle, 2015). Throughout a 25-day testing interval, dairy calves demonstrated long-term consistent individual differences in judgement bias, with some individuals consistently displaying a significantly slower response time to the ambiguous stimuli (Lecorps et al., 2018). This growing body of evidence revealing consistent findings of temporally stable individual differences in decision-making after longitudinal judgement bias testing is suggestive that the task may provide a proxy for trait affect. However, no previous studies have investigated temporally stable individual differences in judgement bias in non-human primates as a means to capture the influence of affective traits on decision-making in rhesus macaques. Understanding whether the judgement bias task provides an index for trait affect in rhesus monkeys is of particular importance because these animals are a more widely used model for human psychological and biological studies (Phillips et al., 2014)- due to their close phylogenetical proximity to humans (Perelman et al., 2011)- compared to many of the animals that have been previously tested on the judgement bias task. Thus, it is imperative to explore the link between trait affect and temporally stable judgement biases in rhesus macaques, as it has the potential to further elucidate the evolutionary development of the relationship between cognition and affect in humans.

1.8 Use of the Judgement Bias Task as a Proxy for Affective States

While momentary affective states are constantly being generated in response to environmental changes and influencing decisions made by the animal to generate a judgement bias, scientists cannot merely observe an animal's decision-making in the task and link results to a specific affective state. Instead, as mentioned previously, scientists must first elicit a presumed affective state in the animal through experimental modulation before conducting the judgement bias task. After presumed affect modulation, animals are then run through the task, and decision-making results are often compared to either a control group or another treatment group to map changes in decision-making to presumed affective states. A majority of judgement bias studies have relied on this between-treatment-group experimental design (Lagisz et al., 2020; Neville et al., 2020) to examine relative differences in the presumed affective states between treatment groups. While this design may allow for the judgement bias task to be used as a tool in animal welfare studies to elucidate the relative effects of environmental manipulations on affective states, they do not reveal whether an animal is in a categorically positive or negative affective state after modulation.

To evaluate whether the task can capture categorically positive and negative affective states, a within-individual repeated measures design utilizing baseline measures must be used to investigate categorically positive and negative affective states within an individual in response to differing modulations. In the sole judgement bias study conducted in rhesus macaques, a repeated measures design was used to understand whether judgement biases within an individual were influenced by husbandry procedures (veterinary health check and environmental enrichment) presumed to induce negative and positive shifts in affective state, respectively (Bethell et al., 2012). However, while researchers found a positive judgement bias after a period of enrichment compared to following a health check, the lack of inclusion of baseline judgement bias measures against which to compare changes in judgement bias did not allow for interpretation of categorically positive and negative states, but instead merely reflected a relative difference between modulations. Alternatively, in dairy calves, a baseline judgement bias measure was used to show that individuals demonstrate a negative shift in judgement bias after experiencing both hot-iron disbudding and separation from the dam, capturing a categorically negative affective state in response to modulations (Daros et al., 2014; Neave et al., 2013). Baseline measurements in this within-individual repeated measures design allow for the integration of an individual's distinctive affective processing into the interpretation of results to allow for a more robust and reliable proxy of categorically positive and negative states. Thus, it is not only imperative to validate the previous judgement bias study in rhesus macaques, but also to analyze intraindividual variability in judgment bias by comparing results after experimental modulation to baseline measurements to validate the judgement bias task as a reliable proxy for categorically positive and negative states.

1.9 Influence of the Interaction between Trait Affect and Affective States on Judgement Biases

The judgement bias task may not merely provide an index for affective states and traits, but also biases may reflect the interaction between states and traits. In humans, it has been suggested that both state and trait variables provide an essential contribution to affect-linked decision-making patterns, with both variables exerting influence in an interactive manner (Mathews & MacLeod, 1994a). For example, individuals experiencing elevated state anxiety only demonstrated selective attention to threatening stimuli if they also had high trait vulnerability to anxiety (MacLeod & Mathews, 1988). Additionally, scientists have suggested that judgement bias in animals may display behavioral flexibility, with both within and between-subject variability mechanisms influencing biases (Faustino et al., 2015). Judgement biases may be a relatively stable and individual-specific generalized tendency to respond to ambiguity (affective trait) that could then codetermine the manifestation of affective states in various situations (Roelofs et al., 2016). However, this state-trait interaction theory has yet to be consistently mapped to results from studies. Although a consistent negative judgment bias trait determined vulnerability to stress-induced anhedonia in rats, the stress-induced shift in affective state during the judgement bias task did not, with a substantial decrease in judgement bias index being observed regardless of trait measures (Rygula et al., 2013). Therefore, the current lack of consistent evidence demonstrating the proposed role of the interaction between affective states and traits in modulating judgment biases has garnered further evaluation on whether the task can capture these effects to allow scientists to begin to disentangle judgment bias variability.

Additionally, it must also be noted that trait affect is not static but instead may change gradually over time under the influence of the repeated generation of momentary states in response to environmental factors (Bliss-Moreau & Rudebeck, 2021). Therefore, it is also essential to understand how the repeated generation of momentary affective states interacts with trait affect to lead to an updated temporally stable “baseline” from which subsequent affective states arise and decision-making biases are generated. Research in capuchin monkeys has demonstrated the impacts of sustained individual experience and other behavioral traits on judgement biases. While there were no short-term detectable effects of a single grooming bout on judgement bias in the monkeys, it was found that consistently high rates of grooming translated into longer-term inter-individual differences in judgement biases suggestive of dynamic changes in trait affect in response to repeated grooming (Schino et al., 2016). However, this study did not test the temporal stability of judgement biases during baseline measures and after repeated affect induction, which would have allowed for the linkage of results to a shift in trait affect. In fact, no study to our knowledge has tested for both temporal stability in baseline individual difference in judgement bias and how this temporal stability shifts in response to repeated generation of affective states. Therefore, to better map results to affective traits, a judgement bias task must be designed that tests not only for stability over time in baseline judgement bias but also during prolonged periods of experimental modulation to capture dynamic changes in trait affect in response to repeated momentary affective states.

1.10 Aims

This thesis will investigate the validity of the judgement bias task as a reliable proxy for affective states and trait affect in captive rhesus macaques (*Macaca mulatta*). The aims of this research are to: (*i*) investigate whether the task captures temporally stable individual differences in decision-making suggestive of baseline trait affect, (*ii*) examine if the task captures categorically negative affective states in the overall sample of individuals after threat induction, (*iii*) analyze whether the interaction between trait affect and momentary affective states influences judgment biases after repeated threat inductions, and (*iv)* explore the relationship between affective reactivity and judgement biases to further validate the judgement bias task as a tool to index affective processing in rhesus macaques. If the judgement bias task reliably indexes trait affect, monkeys will show relatively stable individual differences in judgement bias during baseline measurements. Additionally, if the task is a reliable index of categorically negative affective states, there will be a significant reduction in responses to the ambiguous stimuli after experiencing threat induction compared to baseline judgement bias measurements in the entire sample. Lastly, if judgement bias measurements capture the interaction between trait affect and affective states, temporally stable interindividual differences in judgement biases should remain consistent, despite an overall negative shift in judgement biases caused by repeated threat induction.

2. Methods

All aspects of the study were approved by the University of California, Davis Institutional Animal Care and Use Committee (IACUC) [Protocol #22389] and carried out at the California National Primate Research Center (CNPRC).

2.1 Subjects and Housing

Two age classes of animals were utilized in validating the judgement bias task as a reliable proxy of affective processing. Adults were long-term, unmanipulated study subjects in the laboratory tested on a baseline judgement bias task in 2017. Juveniles were involved in an ongoing study on the impact of fetal Zika virus infection and split into two cohorts (2021: N = 9; 2022: N = 10) that were tested on both a baseline and a post-threat induction judgement bias task in 2021 and 2022, respectively. In conjunction with post-threat induction judgement bias testing, juveniles also underwent a Human Intruder Test to measure affective reactivity in response to threat induction via an unfamiliar human.

Adults

Seven male and seven female adult rhesus macaques (M_{age}= 11.89 years; SD_{age}= 2.14 on the first day of testing) housed at the California National Primate Research Center (CNPRC) at the University of California, Davis were tested on a baseline judgement bias task in 2017. Data was collected by previous researchers using the methods described below and utilized in this current study to address the effects of age and sex on baseline judgement biases. Monkeys were housed indoors in male-female pairs and allowed access to their social partner depending on pair compatibility, with six of the seven pairings being together for 7 hours a day and one pairing being together for 24 hours per day. They were housed in standard adult macaque laboratory cages (112 cm width × 69 cm length × 93 cm height) in rooms that were maintained on a 12-hour light/dark cycle (lights on at 0600) at ~24 degrees C. Subjects were fed twice daily (Lab Diet #5047, PMI Nutrition International INC, Brentwood, MO), supplemented biweekly with fresh fruit and vegetables, had access to water *ad libitum*, and had daily enrichment (including toys, forage, etc.).

Juveniles

Subjects were ten male and nine female juvenile rhesus macaques (M_{age}= 1.79 years; SD_{age}=0.05 on the first day of baseline testing) housed at the California National Primate Research Center (CNPRC) at University of California, Davis. The study was conducted with one cohort of monkeys in 2021 (N = 9) and a second cohort in 2022 (N = 10). One individual from the 2021 cohort was removed from the study due to refusal to participate in the training of the task, leaving 18 total participants in the study. While in utero, monkeys underwent either fetal Zika virus infection or control procedures (as described in Moadab et al., 2022) and were involved in studies monitoring social, cognitive, and affective outcomes after fetal infection from birth to two years of life. Behavior and neurological development were compared between individuals infected with fetal Zika virus and control animals in each cohort. In both cohorts, monkeys were housed in an indoor social group enclosure, composed of juveniles and two adult monkeys selected for compatibility. All individuals were allowed to interact 24 hours a day outside of testing. The indoor social group enclosure (1.78 m width × 6.10 m length × 1.93 m height) was maintained on a 12-hour light/dark cycle (lights on at 0600) at ~24 degrees C.

Subjects were fed twice daily (Lab Diet #5047, PMI Nutrition International INC, Brentwood, MO), supplemented biweekly with fresh fruit and vegetables, had access to water *ad libitum*, and had daily enrichment (including toys, forage, etc.). While training animals on the judgement bias task, monkeys were kept on their regular feeding schedule. Once testing commenced, to ensure adequate interest in the food reward, daily rations of monkey chow and produce were withheld prior to testing and given to monkeys after testing concluded each day.

2.2 Judgement Bias Task

Stimuli and Apparatus

The judgement bias experiment was a visual discrimination task analogous to the 'go/no-go' task developed by Harding et al. (2004) and adapted for rhesus macaques by Bethell et al. (2012). The stimuli consisted of dark gray high-density polyethylene cylinders cut into various lengths (i.e., 5 cm, 10 cm, 15 cm, 20 cm, 25 cm) (Figure 1A). These stimuli were selected because they did not readily resemble any stimuli monkeys were previously exposed to, and therefore monkeys had no previously learned positive or negative associations. Additionally, the use of stimuli on a single physical axis of length difference allowed for stimuli to be numerically coded as a continuous variable and for responses to the ambiguous stimuli to be evaluated as graded and monotonic. Training stimuli were two cylinders, with one cylinder being the shortest (5cm) and one cylinder being the longest (25cm) (Figure 1A). These cylinders were used in training to allow the animal to learn to touch the positive signal stimulus (S+) to receive a “Good” verbal cue and food reward (raisin, dried papaya, and marshmallow) and to refrain from touching the negative signal stimulus (S-) to avoid a “No” verbal cue and five-second timeout. Although monkeys had no prior experience with these verbal cues, the link between the verbal cues and the associated reward/punishment was promptly formed during training. The assignment of length to the rewarded (S+) and unrewarded (S-) stimuli was counterbalanced such that half of the animals had the longest (25cm) cylinder as the reinforced S+ stimulus while the other half had the shortest (5cm) cylinder as the reinforced S+ stimulus. In addition to their use in the training of the visual discrimination task, the signal stimuli were also used for control trials during testing to signal an animal to adjust behavior accordingly and ensure understanding of the task. Three unreinforced ambiguous cylinders of various lengths intermediate to both the rewarded and unrewarded training stimuli (i.e., 10cm, 15cm, 20cm) were used during testing to probe the influence of affect on an animal's response to ambiguity (Figure 1A). The ambiguous stimuli were comprised of three cylinders, with each cylinder having a 5cm length difference compared to the nearest two stimuli. This resulted in a probe stimulus (PI) intermediate in length between the two signal stimuli (15cm) and two probe stimuli (P+ and P-) which were located between the signal stimuli and the intermediate probe stimulus (10cm and 20cm). The length of the P+ and P- stimuli depended on the length of the S+ and S- stimuli such that P+ was always closer in length to S+ and P- was always closer in length to S- (Figure 1A).

The apparatus used in the task consisted of a Plexiglass panel secured to the front of the cage with two holes in which the monkeys were able to reach through. The lower hole allowed the monkeys to reach through and touch the stimuli placed on a metal platform attached to the panel, while the upper hole led to a cavity in which a food reward was placed when S+ was touched (Figure 1B).

Design and Procedure

The judgement bias task consists of two phases: a training phase and a testing phase. The training phase involved all subjects being presented with the positive signal stimuli (S+) and negative signal stimuli (S-) only to learn the correct expression and suppression of touch behavior. The training phase was conducted in three stages, with monkeys advancing to the next stage when they correctly responded to 80% of trials (touch S+ and do not touch S-) over three consecutive days. The testing phase involved the presentation of the positive and negative signal stimuli (S+ and S-) with pseudorandom interspersal of the ambiguous probe stimuli (P-, PI, P+) to capture the influence of affective processing on an animal's decision-making. While adults only underwent a baseline testing phase, the testing phase for the juveniles was split into two testing conditions- baseline testing and testing after threat induction via Human Intruder.

While all animals were socially housed, either in a dyad or a group enclosure, animals were separated to be trained and tested individually in their own cage. Adult animals were trained and tested in their indoor standard adult home cages (112 cm width × 69 cm length × 93 cm height). Juveniles were moved from their group enclosure (1.78 m width × 6.10 m length × 1.93 m height) into a testing cage (59 cm width × 69 cm length × 83 cm height) adjacent to the indoor enclosure within their housing building for training and baseline judgement bias testing.

Training Phase. The first stage of training was designed to motivate the animals to touch the rewarded signal stimulus (S+) to receive a food reward. An experimenter stood to the side of the apparatus platform (Figure 1B), in profile to the monkey, with the S+ stimulus and food rewards (raisins, dried papaya, marshmallows) located in proximity. A trial began with the placement of a stimulus on the platform and ended when the animal either touched the cylinder and received a food reward or after 5 seconds elapsed. After each trial ended, the S+ stimulus was removed from the platform. The food reward was placed in the top well of the apparatus

(Figure 1B) as a reward if the monkey performed the correct touch behavioral response to S+. An animal's behavioral response was recorded for each S+ presentation and coded as a correct response if the stimulus was touched and incorrect if the animal did not touch within the 5-second interval. Additionally, a stopwatch recorded latency to respond (in hundredths of a second) for juveniles. Animals were presented with 20 trials of S+ per training day. Monkeys were required to touch the S+ stimulus on at least 16 of the presentation trials in a single training session over three consecutive days to advance to the second training stage.

The second training stage was designed to allow the animal to visually discriminate between the S+ and S- stimuli and learn appropriate responses. In this stage of training, animals were required to touch S+ to receive a food reward and not touch S- to avoid receiving a verbal "No" cue and a 5-second timeout from training. This stage consisted of 22 trials of stimuli presentation, presented in a fixed repeated sequence of S+, S+, S-, S- trials, with the first and last trials always being S+ presentations. Therefore, a day of Stage 2 training involved the presentation of 12 S+ and 10 S- trials in the fixed repeated pattern. A trial consisted of a single stimulus being placed on the platform until either the monkey touched the stimulus or 2 seconds elapsed. The stimulus was removed from the platform if the animal did not respond within 2 seconds following presentation, and the lack of behavioral response was recorded as a “NO GO”. If the stimulus removed from the platform after a “NO GO” was S-, the experimenter issued a verbal reinforcer of 'Good' when removing the stimuli. Like Stage 1 of training, touching the S+ when presented yielded a food reward. Thus, a correct response was coded “GO” if S+ was touched and “NO GO” if S- was not touched. Monkeys were required to provide correct responses (touch S+, do not touch S-) to at least 18 of the trials presented on each training day of Stage 2 for three consecutive days to advance to Stage 3.

The third stage involved presenting the monkeys with the S+ and S- in a pseudorandom order, where no more than two S+ or S- stimuli were presented consecutively and where an equal number of S+ and S- were presented. Each animal performed 20 trials, with a trial beginning with the stimulus placement on the platform. Similar to Stage 2, animals were given 2 seconds to respond, with a touch to the S+ stimulus yielding a food reward and a touch to the S- yielding a verbal "No" reprimand and a 5-second timeout. If the animal refrained from touching the S- stimulus, they were given a verbal cue of "Good" to reinforce the training. All monkeys were required to provide correct responses to at least 16 trials for three consecutive training days to be eligible to move on to the testing phase.

Baseline Judgement Bias Testing Phase. The animals completed three days of baseline judgement bias testing, with either 20 trials (for adults) or 34 trials (for juveniles) occurring per testing day. Pseudorandom training days similar to Stage 3 of training were conducted between each testing day, such that days 1, 3, and 5 were testing, and 2 and 4 were pseudorandom training days. The pseudorandom training days were used to re-establish the learned discrimination task and ensure the animals performed up to the criterion of 75% correct responses to signal stimuli (touch S+, do not touch S-) during testing. The threshold of 75% correct responses during testing as a criterion was determined empirically based upon the number of signal stimuli presentations and observations of testing performance. During testing sessions, monkeys were presented with the signal stimuli (S+ and S-) on control trials interspersed with ambiguous probe stimuli (P+, PI, P-) on experimental trials in a pseudorandom counterbalanced order. For adults, each testing block consisted of 4 trials for each of the five stimuli (S-, P-, PI, P+, S+) such that 20 trials were conducted. Each testing block for juveniles contained 22 control trials (12 S+ trials, 10 S- trials) and 12 experimental trials (4 P- trials, 4 PI trials, 4 P+ trials) randomly interspersed such that 34 trials were conducted. The first and last trials were always rewarded signal stimulus (S+) trials within each testing block. The exact order of presentation of stimuli over the three days of baseline testing for adults and juveniles can be found in Table 1. Similar to training sessions, touching S+ yielded a food reward, while touching S- resulted in a 'No' verbal response and a 5-second delay of testing. All ambiguous stimuli yielded neither a reward nor punishment for a response. Control trials using the S+ and S- signal stimuli were included to gauge animals' understanding of the task, while ambiguous stimuli were included to test for experimental hypotheses regarding judgement bias. Stimuli were situated out of view of the animals to avoid any anticipation of the stimulus to be presented, and food rewards were hidden behind the back of the experimenter. A testing trial began with the placement of a stimulus on the apparatus platform and ended either after the monkey responded or after 2 seconds had elapsed. For adult monkeys, the behavioral response (touched, did not touch) to stimuli was recorded for each trial over the three testing days. For juveniles, both the behavioral response (touched, did not touch) and latency to respond (in hundredths of seconds) to stimuli were recorded for each trial over the three testing days.

Threat Induction via Human Intruder Test. A Human Intruder Test was conducted with the juvenile monkeys using procedures similar to those used previously by our group and others (e.g., Bliss-Moreau & Baxter, 2018; Bliss-Moreau & Moadab, 2016; Gottlieb et al., 2013). During the Human Intruder Test, an unfamiliar human (a male experimenter) positioned themselves in various spatial positions relative to the caged macaque, either making or not making direct eye contact, to elicit behavioral responses. Testing first involved the transportation of subjects from their indoor home cages to a testing room. For ease of transportation of juveniles, individuals were moved from their group enclosure into adjacent testing cages (59 cm width × 69 cm length × 83 cm height) and separated by individual. Monkeys were transported individually into a testing cage (85.5 cm width × 68 cm length × 82 cm height) in the testing room and given 1 minute of acclimation to the cage and testing environment before the test began. After the one-minute acclimation phase, the unfamiliar human entered the room and began the test. Subjects were presented with the unfamiliar human in four different positions for one minute each: 1) Profile-Far: human standing ≈1 meter from cage facing 90 degrees away with their left side facing the animal; 2) Profile-Close: human standing 0.3 meters from cage facing 90 degrees away with their left side facing the animal; 3) Stare-Far: human standing ≈1 meter from cage facing monkeys directly and making eye contact; 4) Stare-Close: human standing 0.3 meters from cage facing the cage and making eye contact with subjects. Behavioral data recorded from the task was collected by a trained observer using a 1/0 sampling in 10-second bins to analyze the frequency of affect-related behaviors generated by the monkey during each testing context. These affect-related behaviors included: cage shake, coo, grunt, scream, bark, yawn, lip smack, barred teeth, self-scratch, stereotypy, freeze, threat, and tooth grind (Table 2). Behaviors displayed within a 10-second bin were coded as a 1, and those not performed within the bin were given a score of 0. The behavioral observer was trained to the lab standard of greater than 85% inter-rater reliability. The summation of affect-related behaviors produced by a monkey in a given period was then used to quantify affective reactivity, a valuable indicator of individual variation in affective responsiveness to stimuli used in previous studies (Bliss-Moreau et al., 2021; Bliss-Moreau & Baxter, 2018; Bliss-Moreau & Moadab, 2016). The Human Intruder Test was also used in the juvenile monkeys to induce a presumed momentary negative affective state through threat induction via the unfamiliar human intruder. After the 5-

minute test finished, the human experimenter left the room, and a familiar human boxed the animal and transported them to an adjacent room for judgement bias testing.

Post Human Intruder Judgment Bias Testing Phase. Juvenile monkeys underwent judgement bias testing sessions immediately after being exposed to an unfamiliar male experimenter in the Human Intruder Test. After the Human Intruder Test was concluded, juveniles were transferred into a testing cage (59 cm width×69 cm length×83 cm height) in an adjacent room for judgement bias testing. Methods for testing after exposure to the human intruder were similar to those conducted during baseline testing a week prior. During testing sessions, monkeys were presented with the signal stimuli (S+ and S-) on control trials interspersed with ambiguous probe stimuli (P+, PI, P-) on experimental trials in a pseudorandom counterbalanced order. Each testing block contained 22 control trials (12 S+ trials, 10 S- trials) and 12 experimental trials (4 P- trials, 4 PI trials, 4 P+ trials) randomly interspersed. The order of stimuli presentation for juveniles was the same as baseline testing and can be found in Table 1. After a juvenile finished testing, they were transported back to their home building and put back into the individual testing cages. Once all individuals finished judgement bias testing after the Human Intruder task, experimenters performed refresher pseudorandom training (similar to Stage 3 of training) for the first two testing days to ensure the animals retained the learned discrimination task. Therefore, a testing day consisted of the judgement bias task in a testing room after the Human Intruder Test, followed by a refresher training session in the monkey's homeroom. Thus, rather than having a span of 5 days for a testing period (three days of testing with two days of training in-between), this condensed testing into three consecutive days.

2.3 Statistical Analysis

Effects of Parameters on Judgement Bias Task Training Duration

Non-parametric comparisons were conducted in SPSS 28.0 (IBM Corp., 2021) to assess the effects of parameters (ZIKV treatment, age, sex, stimulus length) on the duration required to pass the second and third stages of training, where individuals learn to differentiate between the S+ and S- stimuli. Non-parametric comparisons were used due to the assumption of normality for training duration data being violated for all group combinations of ZIKV treatment, age, sex, and stimulus length, as assessed by the Shapiro-Wilk's test ($p < 0.05$). Mann-Whitney U tests were used to compare differences in the duration of training between juveniles infected with fetal Zika virus and control individuals, males and females, adults and juveniles, and those trained with a 25cm long S+ versus those trained with a 5cm long S+. A Kruskal-Wallis H test was used to analyze the effects of the interaction between age and sex on differences in the number of training days, with comparisons occurring between adult males, adult females, juvenile males, and juvenile females. Significance in the Kruskal-Wallis H test indicated whether at least one age*sex group stochastically dominated another group, but pairwise Mann-Whitney U tests were used to analyze the specific sample pairs for stochastic dominance.

Influence of Parameters on Judgement Biases

To analyze the influence of various parameters on judgement biases, responses to stimuli during the judgement bias testing phase were specified with a binary variable (1- "touched stimulus", 0- "refrained from touching stimulus") and analyzed using generalized linear mixed-effects models (Pinheiro & Bates, 2006) in R 4.1.0 (R Core Team, 2021) using the 'lme4' package (Bates et al., 2015). Models used a logit link function and binomial error distribution to account for the binary nature of the response variable. Assumptions regarding homoscedasticity of errors and normality of random effects and errors were examined using a graphical analysis of residuals. Subject identity was included as a random effect in all models to control for repeated measures within the same subjects. Stimuli were included in all models as a fixed effect and transformed to a continuous predictor variable ‘stimulus value’ per statistical recommendations (Gygax, 2014). ‘Stimulus value’ was coded as the difference in length of the stimulus relative to the length of S-, such that S- was 0, P- was 5, PI was 10, P+ was 15, and S+ was 20. In generalized linear mixed-effects models involving interactions of factors with several levels, models were optimized using a Bound Optimization BY Quadratic Approximation (BOBYQA) proposed by Powell (2009). Models were selected based on their ranking according to the model weights derived from the Bayesian information criterion (BIC). Additionally, the 'anova' function from the ‘stats’ package (R Core Team, 2021**)** was used to compare the fit of the final model against the null model (a model retaining the random effect of individual but with the fixed effects removed and an intercept of 1 specified), and the final model was accepted if it provided a significantly better fit than the null at $p < 0.05$. The ‘Anova’ function from the ‘car’ package (Fox & Weisberg, 2019) was used for Type III Wald chi-squared tests of the generalized linear mixed-effects models to determine the significant effects of predicator variables on the dependent response variable. The ‘emmeans’ package (Lenth, 2022) was used for planned comparisons of predictor variables from the generalized linear mixed-effects models using the Holm-Bonferroni method of correction for multiple comparisons.

Three generalized linear mixed-effects models were implemented to analyze the effects of fetal Zika Virus treatment, age and sex, and threat induction on judgement biases. An exploratory analysis of the influence of fetal Zika virus on baseline judgement bias testing was performed using a binomial generalized linear mixed-effects model that was fitted to the dependent binary response variable with juvenile identity as the random effect. Only binary response data from the baseline judgement bias testing context was used in this model. Fixed effects in this model were ZIKV treatment (control, infected), stimulus value (S-: 0, P-: 5, PI: 10, P+: 15, S+: 20), reinforced S+ stimulus length (long, short), and testing day (1, 2, and 3). An interaction between stimulus value and ZIKV treatment was also included in the model. To explore the influence of age and sex on response during baseline judgement bias testing in adults and juveniles, a binomial generalized linear mixed-effects model was fitted to the dependent binary response variable with animal identity as the random effect. Fixed effects of the model were age class (adult, juvenile), sex (female, male), stimulus value, reinforced S+ stimulus length, and testing day. Age class, sex, and stimulus value were also included as interaction terms. To understand the effects of threat induction on judgement biases in juveniles, an additional binomial generalized linear mixed-effects model using response as the dependent variable and juvenile identity as the random effect was fitted. Fixed effects of this model were testing condition (baseline, post threat induction), stimulus value, sex, reinforced S+ stimulus length, and testing day. Interaction terms also included were testing condition and stimulus value, sex and stimulus value, sex and testing type, and testing day and stimulus value.

Transformation of Responses to Stimuli into a Judgement Bias Index

A judgement bias index (JBI) was created to allow for more accessible analyses of individual differences in decision-making by generating a singular measure of judgement bias within an individual- either per testing day, per testing type, or overall. The JBI utilized the trapezoidal rule to approximate the area under the curve generated by the proportion of responses to the five stimuli. The JBI was approximated by partitioning the integration interval into four subintervals with a width of 5 and summing each of the four subintervals into a singular area under the curve. Each subinterval had a width of 5 due to each stimulus differing in length from adjacent stimuli by 5cm. The use of proportion data for calculations of the JBI allowed for higher JBI values for an individual to reflect that individual demonstrating a higher proportion of responses to stimuli during judgement bias testing. However, this index could not differentiate to which specific stimuli an individual had a higher proportion of response. Figure 2 demonstrates a visualization of the use of the trapezoidal rule for the transformation of proportion data into a singular JBI. A simplified formula for calculating the JBI can be presented as:

$$\mathrm{JBI} = \mathbf{5}\left(\frac{\boldsymbol{x_{prop\,S-} + 2(x_{prop\,P-} + x_{prop\,PI} + x_{prop\,P+}) + x_{prop\,S+}}}{\mathbf{2}}\right)$$

Temporal and Context Stability in Individual Differences in Judgement Biases

Intra-class correlation coefficient (ICC) analysis was performed in R 4.1.0 (icc function, irr library; Gamer et al., 2019) (R Core Team, 2021) to confirm both temporal and context stability in individual differences in judgement biases in both testing contexts (baseline, post-threat induction). To analyze temporal stability, an individual's JBI was first calculated for each of the three days of testing, and an ICC was used to reflect the degree of stability in individual differences in judgement biases across the days of testing. Temporal stability ICC estimates and their 95% confident intervals for the JBI were calculated based on a mean measurement (k = 3), absolute-agreement, 2-way mixed-effects model. To analyze context stability, an individual's JBI was averaged over the three days of baseline testing and over the three days of post-threat induction testing to obtain trait judgement bias indexes between both testing contexts. The context stability analysis then correlated individuals' average baseline JBI with their post-threat induction JBI using a single rater, absolute-agreement, 2-way mixed-effects model to understand rank-order stability across testing contexts. The magnitude of rank-order stability was used to determine whether individuals' relative ranking of JBI remained stable across contexts; high rank-order stability would indicate that individuals demonstrating the highest baseline trait JBI would also demonstrate the highest trait JBI post-threat induction, and that other individuals' relative rankings would remain stable across contexts. All interpretations of ICC measures were based upon recommendations by Koo and Li (2016) to allow for the evaluation of the levels of both temporal and context stability.

Influence of Threat Induction on Affective Reactivity

A linear mixed-effects (LME) model with restricted maximum likelihood was implemented in R 4.1.0 (R Core Team, 2021) using the ‘lme4’ package (Bates et al., 2015) to analyze the influence of threat induction on affective reactivity in the Human Intruder Test. The ‘Anova’ function from the ‘car’ package (Fox & Weisberg, 2019) was used for Type II Wald chi-squared tests to investigate the significant effects of close proximity and direct eye contact on increases in affective reactivity. The linear mixed-effects model was also used to validate affective reactivity as both a trait and state measure in preparation for exploring the relationship between affective reactivity and both trait and state judgement bias measures. An animal’s affective reactivity was used as the dependent variable in the model, while juvenile identity was included as a random effect. The fixed effects of this model were testing condition (profile-far, profile-close, stare-far, stare-close) and testing day (1, 2, and 3).

Behavioral Components of Affective Reactivity

While thirteen affective behaviors were recorded during the Human Intruder test (Table 2), two behaviors (scream and bark) were excluded from data analyses due to no occurrences of these behaviors displayed during the task. A Principal Component Analysis (PCA) was used to transform the remaining eleven affective behaviors collected during the Human Intruder Test into four behavioral components using a Varimax rotation with Kaiser Normalization **(**IBM SPSS Statistics 28, Armonk, NY, USA**)**. Prior to performing the PCA, the sampling adequacy of the data was examined through confirmation that Bartlett's test of sphericity was significant ($p < 0.001$), the Kaiser-Meyer-Olkin measure was sufficiently high (KMO = 0.54), and examination of the subject-to-variable ratio (1.73:1). Although the PCA was conducted, interpretations of results were regarded with caution due to the low KMO value and small subject-to-variable ratio, as per recommendations (Budaev, 2010). Factors were identified at eigenvalues greater than one and through examination of the scree plot. Behaviors were grouped into components based on a factor loading cut-off of ≥ 0.45 per recommendations by Tabachnick and Fidell (2007) and Comrey and Lee (1992) to obtain a more stringent loading threshold. The four-factor structure of affective behaviors explained 62.9% of the total variance in the data and can be seen in Table 3. Factor scores were also generated for each individual from the PCA using a least-squares regression approach to refine analyses further to focus on the relationship between the specific components of affective reactivity and judgement biases.

To further characterize the grouping of affective behaviors, the loading of behaviors onto the four components was used to generate intuitive labels to aid in interpreting analyses. The first behavioral grouping contained the self-scratch, tooth grind, and yawn behaviors and was labeled "Behavioral Tension" due to these behaviors being involved in a monkey's internal response to tension (J. R. Anderson, 2010; Hinde & Rowell, 1962; Schino et al., 1988). The second behavioral grouping included the threat and cage shake behaviors and was labeled "Threat Display" as behaviors were an outward display of threat back at the unfamiliar human. The third grouping was labeled “Vocalizations” due to both the coo and grunt vocalization behaviors loading on to a component together. The fourth behavioral grouping, which included the lip smack and barred teeth behaviors, was labeled as “Appeasement Behaviors” as these behaviors are involved in reducing the risk of aggression and in appeasing threats in social settings. (Hinde & Rowell, 1962; Maestripieri & Wallen, 1997; Petit & Thierry, 1992). However, it must be acknowledged that assessing the meaning behind behaviors, as noted earlier, is a contentious issue that may result in errors in interpretations of the underlying intent of behaviors. Thus, our labels serve as a single categorization of these behavioral groupings, but the interpretation of these specific behavioral groupings should remain open to other evaluations.

Relationship between Judgement Biases and Affective Reactivity

To analyze the relationship between affective reactivity from the Human Intruder Test and judgement biases, mixed-effects models were implemented in R 4.1.0 (R Core Team, 2021) using the ‘lme4’ package (Bates et al., 2015). The ‘Anova’ function from the ‘car’ package (Fox & Weisberg, 2019) was then used for Type III Wald chi-squared tests to determine the significant effects of fixed effects on the dependent variable in these models. A binomial generalized linear mixed-effects model, using binary response as the dependent variable and juvenile identity as the random effect, was created to analyze the possible effects of an individual’s four PCA factor scores on responses to stimuli during subsequent post-threat induction testing. Thus, only binary response data from the post-threat induction judgement bias testing context was used in this model. Fixed effects of this model were stimulus value, individuals’ four PCA factor scores, and testing day. The interaction between each of the four factor scores and stimulus value was also inputted into the model to test for the effects of the specific affective reactivity components on responses to specific stimuli. Additionally, a linear mixed-effects model was used to evaluate the interaction between an individual's affective reactivity (as well as the four behavioral components of reactivity) and trait average baseline JBI in generating post-threat induction JBI per day of testing. The dependent variable of this model was individuals' post-threat induction JBI and juvenile identity was included as the random effect. Fixed effects included the interaction between affective reactivity and baseline JBI, the interaction between each of the four factor scores and baseline JBI, and testing day.

Kendall rank correlation tests were conducted in R 4.1.0 using the 'cor.test' function from the 'stats' package (R Core Team, 2021**)** to evaluate the ordinal association between affective behaviors collected during the Human Intruder Test and judgement bias measures. A Kendall rank correlation test was performed to investigate the relationship between an individual's trait affective reactivity during the Human Intruder Test and average baseline JBI. Additional Kendall rank correlation tests were performed on individuals' average PCA factor scores and average baseline JBI. These correlations were used to further validate the judgement bias task as a proxy for affective traits. A Kendall rank correlation test was also performed to understand the strength of the relationship between an individual's state affective reactivity during the Human Intruder Test and judgement biases during subsequent post-threat induction judgement bias testing. This correlation was used to validate the judgement bias measure as an adequate proxy of shifts in affective state. The 'p.adjust' function from the 'stats' package (R Core Team, 2021) was used to perform Holm-Bonferroni adjustments of p-values to counteract multiple comparisons.

3. Results

3.1 Impact of Fetal Zika Virus on Judgement Bias Task Training and Testing in Juveniles

Due to juvenile monkeys being involved in an ongoing study on the impact of fetal Zika virus infection, there was a possibility that Zika virus infection could have led to cognitive impairment and thus greatly impacted judgement bias training and/or testing. Previous studies in adult mice and humans have shown the virus preferentially targeting mature memory-related brain regions, potentially leading to cognitive deficits (Figueiredo et al., 2019). However, our results showed that the effects of Zika virus did not significantly influence the training or testing of the judgement bias task. There was no significant difference in the number of training days required to learn the discrimination of the S+ and S- stimuli during the second training stage [$U = 37$, $z = 0.22$, $p = .83$] between control ($M = 6.5$ days, $SD = 2.55$ days) and ZIKV-infected animals ($M = 6.7$ days, $SD = 2.33$ days). During baseline judgement bias testing, there was no significant effect of ZIKV infection on responses to stimuli during the task ($\chi 2$ (1) = 0.15, $p = 0.70$; $OR_{Control/Zika} = 0.89$, $p = 0.79$). Thus, this effect was not considered in future analyses and all juveniles, regardless of treatment, were grouped together.

3.2 Influence of Parameters on Training of Judgement Bias Task

A Kruskal-Wallis H test revealed a significant interaction between age and sex on the number of training days required to pass the second training stage (H (3) = 11.37, $p = 0.009$; Figure 3). Adult females required significantly longer to pass the second training stage ($M = 14.40$ days, $SD = 2.87$ days) compared to juvenile females ($M = 7.11$ days, $SD = 2.51$ days [$U = 1.50$, $z = -2.73$, $p = 0.006$], adult males ($M = 6.20$ days, $SD = 2.32$ days) [$U = 0.50$, $z = 2.40$, $p = 0.016$], and juvenile males ($M = 6.11$ days, $SD = 2.23$ days) [$U = 1$, $z = -2.80$, $p = 0.003$].

Analyzing the effect of sex, females required significantly longer (*M* = 9.71 days, *SD* = 4.38 days) to consistently discriminate between the positive and negative signal stimuli in the second training stage compared to males (*M* = 6.14 days, *SD* = 2.26) [*U* = 49.50, *z* = 2.21, *p* = .027]. Examining the effects of age, adults showed a near significant increase (*U* = 49.50, *z* = -1.92, *p* = .055) in days required to pass the second training stage (*M* = 10.30 days, *SD* = 4.86 days) compared to juveniles (*M* = 6.61 days, *SD* = 2.43 days). Furthermore, the counterbalanced difference in lengths of S+ and S- had no significant effect on the number of days required to pass the second training stage (*U* = 76, *z* = -0.97, *p* = .332) between those trained with S+ as the longest stimulus (*M* = 7.60 days, *SD* = 4.44 days) and those with S+ as the shortest stimulus (*M* = 8.31 days, *SD* = 3.17 days).

A Kruskal-Wallis H test showed no significant interaction between age and sex in influencing the duration of the third training stage (*H* (3) = 1.49, *p* = 0.68), with each age*sex group requiring around 4 days to pass (*M* = 3.79 days, *SD* = 1.99 days). Furthermore, there was no significant effect of sex on the number of days required to pass the third training stage (**Female**: *M* = 3.14 days, *SD* = 0.52 days; **Males**: *M* = 4.43 days, *SD* = 2.61 days) [*U* = 75, *z* = -1.03, *p* = .303]. Additionally, there was no significant effect of age on the duration of the third training stage (**Adults**: *M* = 4.10 days, *SD* = 2.70 days; **Juveniles**: *M* = 3.61 days, *SD* = 1.42 days) [*U* = 87, *z* = -0.12, *p* = .91]. Lastly, the counterbalanced difference in lengths of S+ and S- had no significant effect on the number of days required to pass the third training stage (*U* = 71, *z* = -1.14, *p* = .254) between those trained with S+ as the longest stimulus (*M* = 3.31 days, *SD* = 1.21 days) and those with S+ as the shortest stimulus (*M* = 4.42 days, *SD* = 2.56 days). Ultimately, there were no effects of parameters on the duration of the third training stage, as all individuals quickly moved through this stage once the discrimination between stimuli was learned in the second stage of training.

3.3 Parameters Influencing Baseline Judgement Bias Testing

General Response During the Baseline Judgment Bias Task

The criterion for data inclusion in analyses of baseline judgement biases was correct responses (touch the S+ stimulus within 2 seconds; refrain from touching the S- stimulus for at least 2 seconds) on at least 75% of control trials where these stimuli were presented during testing. While all 18 juveniles reached this criterion for all three days of baseline judgement bias testing, three out of the 14 adults included in the experiment failed to reach the criterion on at least one day of testing. Two adults failed a single day of baseline testing, while one adult failed all three days, resulting in the removal of five testing sessions or 100 trials [40 control trials (S+, S-), 60 experimental trials (P-, PI, P+)] of data from analyses of the baseline testing context.

Analyses of the proportion of responses during baseline judgement bias testing showed that animals had a higher likelihood to touch the food-reinforced S+ cylinder ($P = 0.98$, $SE = 0.01$) when presented compared to the punished S- cylinder ($P = 0.03$, $SE = 0.01$) (χ^2 (1) = 51.61, $p < 0.001$), confirming that the monkeys successfully demonstrated the expected discrimination of S+ and S- stimuli. Additionally, the monkeys generalized across the three intermediate ambiguous probe stimuli as expected (χ^2 (2) = 35.39, $p < 0.001$), with the middle PI probe having a significantly higher likelihood of response ($P = 0.45$, $SE = 0.05$) compared to the near negative P- probe ($P = 0.17$, $SE = 0.03$) [$OR_{P-/PI} = 0.24$, $p < 0.001$] and significantly lower likelihood compared to the near positive P+ probe ($P = 0.84$, $SE = 0.03$) [$OR_{PI/P+} = 0.16$, $p < 0.001$]. The likelihood to respond to stimuli and pairwise comparisons of responses to adjacent stimuli can be seen in Table 4. Furthermore, there was a significant effect of signal stimuli length (χ^2 (1) = 11.26, $p < 0.001$) on responses to the ambiguous stimuli during baseline testing, with those trained to touch the 25cm long stimulus as S+ having an overall higher likelihood to respond ($P = 0.67$, $SE = 0.06$) compared to those trained with S+ as the 5cm long stimulus ($P = 0.31$, $SE = 0.06$). Thus, monkeys presented with the 10 cm long P-, 15 cm long PI, and 20 cm long P+ had a significantly higher proportion of responses to these stimuli ($OR_{Long/Short} = 4.52$, $p < 0.001$). However, there was no significant effect of presentation number of the ambiguous stimuli on the proportion of response across all individuals during baseline testing. The monkeys were equally likely to respond to the ambiguous probe stimulus whether it was the 1st, 2nd, 3rd, 4th, or 5th time seeing it during testing sessions ([**P-**: χ^2 (4) = 0.54, $p = 0.97$], [**PI**: χ^2 (3) = 1.52, $p = 0.68$], [**P+**: χ^2 (4) = 3.27, $p = 0.51$]).

Effects of Sex and Age

A generalized linear mixed-effects model was utilized to analyze the effects of age and sex on the likelihood to respond to the ambiguous stimuli during baseline judgement bias testing. There was no significant effect of sex on overall likelihood to respond during baseline testing (**Male**: $P = 0.48$, $SE = 0.06$; **Female**: $P = 0.50$, $SE = 0.06$) [χ^2 (1) = 0.02, $p = 0.90$]. There was, however, a near significant effect of age (χ^2 (1) = 3.57, $p = 0.06$), as adults had a higher likelihood to respond to the ambiguous stimuli compared to juveniles (**Adult**: $P = 0.54$, $SE = 0.07$; **Juvenile**: $P = 0.43$, $SE = 0.06$) [$OR_{Adult/Juvenile} = 1.52$, $p = 0.25$]. Furthermore, while there was no significant effect of the interaction between age and sex on the overall response to the ambiguous stimuli (χ^2 (1) = 0.08, $p = 0.78$), a significant interaction between age, sex, and responses to specific stimuli emerged (χ^2 (2) = 10.01, $p = 0.007$) (Figure 4). Analyzing the effects of sex in the two age groups separately, adult males had a near significantly lower proportion of response to near positive P+ stimulus during baseline testing ($z = -1.81$, $p = 0.07$) compared to adult females (**Adult Male**: $P = 0.67$, $SE = 0.12$; **Adult Female**: $P = 0.85$, $SE = 0.07$) [$OR_{Adult\ Male\ P+/Adult\ Female\ P+} = 0.33$, $p = 0.89$]. In contrast, juvenile males showed a significantly higher proportion of response to the near positive P+ stimulus ($z = 3.13$, $p = 0.002$) compared to juvenile females (**Juvenile Male**: $P = 0.93$, $SE = 0.03$; **Juvenile Female**: $P = 0.74$, $SE = 0.07$) [$OR_{Juvenile\ Male\ P+/Juvenile\ Female\ P+} = 4.73$, $p = 0.39$].

3.4 Temporal Stability in Baseline Judgement Bias Testing of Juvenile Monkeys

As the days of baseline testing progressed, there was a significant reduction in the overall likelihood to respond to the ambiguous stimuli (χ^2 (2) = 10.24, $p = 0.006$) for the sample of juvenile monkeys (**Day 1**: $P = 0.57$, $SE = 0.07$; **Day 2**: $P = 0.42$, $SE = 0.07$; **Day 3**: $P = 0.30$, $SE = 0.06$). Pairwise comparisons revealed a significantly higher likelihood of respond on Day 1 that decreased over the following two days of baseline testing ($OR_{Day\ 1/Day\ 2} = 1.88$, $p = 0.04$; $OR_{Day\ 2/Day\ 3} = 1.69$, $p = 0.07$), with a significant difference in likelihood to respond between Day 1 and Day 3 ($OR_{Day\ 1/Day\ 3} = 3.16$, $p = 0.0001$) [Figure 5]. To analyze temporal stability in an individual's judgement bias index (JBI), the intraclass correlation coefficient (ICC) was calculated for stability across the three days of baseline testing. A moderate degree of temporal stability in individual differences in JBI across the three days of baseline testing was found for juveniles. The average measures ICC estimate was 0.59 with a 95% confidence interval from 0.14 to 0.83 (F(17, 31.5) = 2.66, $p = 0.009$), indicating that the judgement bias index for the group, on average, was significantly correlated across days.

3.5 Influence of Threat Induction via Human Intruder on Judgement Biases in Juvenile monkeys

Threat induction via the Human Intruder Test led to a significant reduction in the likelihood to respond to the ambiguous stimuli (χ^2 (1) = 46.97, $p < 0.001$; **P-:** $P = 0.03$, $SE = 0.01$; **PI:** $P = 0.12$, $SE = 0.03$; **P+:** $P = 0.37$, $SE = 0.05$), as well as the rewarded S+ stimulus (χ^2 (1) = 151.38, $p < 0.001$; **S+:** $P = 0.72$, $SE = 0.08$), during the judgement bias task in the overall sample of juveniles. The significant reduction in the likelihood to respond to stimuli due to threat induction for the overall sample of juveniles can be seen in Figure 6. There was no significant interaction between threat induction and stimulus characteristics in influencing the magnitude of change in response to the ambiguous stimuli (χ^2 (2) = 4.25, $p = 0.12$), with similar significant reductions being seen at all three ambiguous stimuli. However, analyzing responses to all stimuli, a significant interaction between threat induction and stimulus value emerged (χ^2 (4) = 34.97, $p < 0.0001$), with the largest decrease in response being seen at the S+ stimulus [$OR_{Post\ Threat\ Induction\ S+/Baseline\ S+} = 0.04$, $p < 0.0001$]. It should be noted that the significant reduction in response to S+ between baseline ($M = 0.98$, $SD = 0.15$) and post-threat induction ($M = 0.66$, $SD = 0.47$) led to 9 individuals performing below the threshold of 75% correct responses during testing sessions and the removal of 20 testing sessions, or 680 trials [440 control trials, 240 experimental trials], of data from the post threat induction context.

After correcting the data through the removal of non-criterion reaching testing sessions, a significant effect of testing context on response to the ambiguous stimuli remained (χ^2 (1) = 24.64, $p < 0.001$); however, the significant interaction between threat induction and stimulus value in influencing response to the ambiguous stimuli was lost (χ^2 (2) = 1.67, $p = 0.43$). As a result of data correction, there was a loss of significant difference in response to the near negative probe stimulus P- ($z = 1.91$, $p = 0.22$). Nevertheless, the significant reduction in response to the intermediate probe stimulus PI ($z = 2.82$, $p = 0.024$) and near positive probe stimulus P+ (*z* = 4.39, *p* = 0.0001) remained. Additionally, the significant effect of threat induction (χ^2 (1) = 10.34, *p* = 0.001) on the reduction of response to the positive signal stimulus S+ (*z* = 3.05, *p* = 0.014) remained after data correction. Figure 7 shows the differences in the likelihood to respond to all stimuli between testing contexts for the corrected data.

3.6 Temporal Stability in Judgement Bias Following Threat Induction in Juveniles

Contrary to baseline judgement bias measures, there was no significant effect of testing day on the overall response to the ambiguous stimuli during post-threat induction judgement bias testing in the sample of juveniles (χ^2 (2) = 4.66, *p* = 0.098) [Figure 8]. However, analyzing the effects of testing day on responses to each ambiguous stimuli separately revealed a significantly higher proportion of response to PI during testing Day 2 (β = 2.56, *SE*= 1.07, *z* = 2.40, *p* = 0.017) compared to the other two days of testing. There were, however, no significant differences in response to P- (χ^2 (2) = 4.07, *p* = 0.131) or P+ (χ^2 (2) = 2.69, *p* = 0.261) between the three days of testing, with responses to these stimuli decreasing across testing days (**P-**: β = -0.06, *SE*= 0.023; **P+**: β = -0.14, *SE*= 0.11). Analyzing temporal stability in individual differences in post-threat induction judgement biases, a good degree of stability was found across the three days. The average measures ICC estimate was 0.80 with a 95% confidence interval from 0.57 to 0.92 (F(17, 34) = 4.95, *p*<0.001), indicating that temporal stability in individual differences in judgement biases was stronger compared to baseline measures.

3.7 Rank-Order Stability in Individual Differences in Judgement Biases Across Testing Contexts

An Intraclass Correlation (ICC) analysis was used to assess rank-order stability, or the degree to which the ordering of individuals' judgement biases relative to one another remained stable between testing contexts, despite the overall group decreasing in responses to stimuli following threat induction. A poor degree of rank-order stability was found for individual differences in JBI between the baseline and threat induction testing contexts. The single measures ICC estimate was 0.03 with a 95% confidence interval from -0.15 to 0.31 (F(17, 18) = 1.12, p = 0.41). Individuals demonstrating the highest baseline judgement bias index did not also demonstrate the highest relative judgement bias index post-threat induction. Analyses instead found that juveniles did not maintain their ranking of trait JBI relative to one another between testing contexts, with 16 individuals' ranks decreasing and two individuals' ranks increasing.

3.8 Relationship Between Response During the Human Intruder Test and Judgement Bias Measures

General Parameters of Human Intruder Measures

In preparation for investigating the relationship between affective reactivity from the Human Intruder Test and judgement biases in the juveniles, the effects of threat induction on individuals during the Human Intruder Test were explored. Affective reactivity during the Human Intruder Test was generated from the summation of behaviors expressed during each day of testing. A linear mixed-effects model revealed a significant main effect of threat induction on generating increased affective reactivity in the juveniles ($\chi 2(1) = 116.71$, $p < 0.001$) as expected, demonstrating that direct eye contact and close proximity by an unfamiliar human generated a significantly more affect-related behavior than the other testing conditions. This allowed for testing whether behavioral responses during a session of Human Intruder predicted subsequent judgement biases post-threat induction to validate the judgement bias task as a proxy for momentary changes in affective state. Additionally, the behavioral reaction to the unfamiliar human did not change across the three days of testing in the overall sample ($\chi 2$ (4) = 4.85, p =

0.30), suggesting that the same magnitude of affective response was induced across all three test days. Furthermore, a good degree of stability in an individual's affective reactivity across the three days of Human Intruder Testing was found (ICC: 0.80, $p < 0.001$), suggesting trait-like temporal stability in behavioral measures from the Human Intruder Test. This allowed for correlations between an individual's behavioral response to threat in the Human Intruder Test and average response during judgement bias testing to validate whether the judgement bias task captures affective traits.

Relationship Between Human Intruder Measures and Baseline Judgement Biases

There was no significant correlation between an individual's average affective reactivity during the Human Intruder task and their average baseline JBI ($r_{Tb} = 0.099$, $p = 0.57$; Figure 9). To explore this relationship further, factor scores were generated from the PCA analysis conducted on the behavioral data collected during the Human Intruder Test. An individual's factor scores were then correlated with their baseline judgement biases to analyze the relationship between the four behavioral components of affective reactivity- "Behavioral Tension", "Threat Display", "Vocalizations", and "Appeasement Behaviors"- and average baseline JBI separately. While a significant negative correlation between "Vocalizations" and average baseline JBI was found ($r_{Tb} = -0.433$, $p = 0.049$; Figure 10), there was no significant correlation between "Behavioral Tension" ($r_{Tb} = -0.171$, $p = 0.973$) , "Threat Display" ($r_{Tb} = 0.131$, $p = 0.973$), , or "Appeasement Behaviors" ($r_{Tb} = 0.118$, $p = 0.973$) and average baseline JBI.

Relationship Between Human Intruder Response and Post-Threat Induction Decision-Making Per Testing Day

A Kendall rank correlation test was performed to analyze whether the magnitude of an individual's affective reactivity during the Human Intruder Test significantly influenced judgement biases during subsequent post threat induction testing. For a single day of testing, there was no significant correlation between the total number of affective behaviors expressed during the Human Intruder test and the proportion of responses to stimuli during subsequent judgement bias testing ($r_{Tb} = 0.086$, $p = 0.37$; Figure 11). However, factor scores were generated for each individual per testing day from the PCA using a least-squares regression approach to refine analyses further to focus on the effects of specific components of affective reactivity on responses in subsequent judgement bias testing. A generalized linear mixed-effects model analyzed the interaction between an individual's four factor scores and responses to stimuli during post-threat induction judgement bias testing for all individuals (including non-criterion testing sessions). Significant interactions between "Threat Display" and response to the PI and S+ stimuli ($\chi 2$ (4) = 12.98, $p = 0.011$) and between "Appeasement Behaviors" and response to the PI and S+ stimuli ($\chi 2$ (4) = 11.65, $p = 0.02$) during judgement bias testing per day were found. Individuals who displayed a high number of threat and cage shake behaviors, and low number of stereotypies, showed a near significant decrease in responses to the PI ($\beta = -0.81$, $SE= 0.49$, $z = -1.66$, $p = 0.098$) and S+ ($\beta = -0.77$, $SE= 0.46$, $z = -1.68$, $p = 0.093$) stimuli in the subsequent judgement bias testing. Additionally, individuals who displayed a high number of barred teeth and lip smack behaviors during a Human Intruder Testing session also showed a decrease in response to the PI ($\beta = -0.02$, $SE= 0.38$, $z = -0.58$, $p = 0.954$) and S+ ($\beta = -0.42$, $SE= 0.35$, $z = -1.22$, $p = 0.221$) stimuli during judgement bias testing on that same day. Additional analyses excluding non-criterion testing sessions revealed a loss of the significant effect of "Appeasement Behaviors" on responses to the PI and S+ stimuli ($\chi 2$ (4) = 4.71, $p = 0.319$), however the significant effect of “Threat Display” on individual’s responses to the S+ stimulus remained (β = -0.99, *SE*= 0.46, *z* = -2.13, *p* = 0.033).

Interaction Between Affective Reactivity and Trait-Like Baseline Judgement Bias in Generating State-Like Changes in Post-Threat Induction Judgement Bias

A linear mixed-effects model evaluated the interaction between an individual’s trait average baseline judgement bias index and state affective reactivity in generating post-threat induction judgement biases per day of testing. There was no significant interaction between an individual’s average baseline judgement bias and the total affective reactivity displayed during the Human Intruder Test in influencing how that individual responded in post-threat induction judgement bias testing ($\chi2$ (1) = 1.02, *p* = 0.313). Exploring the interaction in further depth, an individual’s factor scores for the four behavioral components (Table 2) were used to understand each behavioral group's interaction separately and average baseline JBI in generating post-threat induction judgement biases per testing day. There were no significant interactions between “Behavioral Tension” ($\chi2$ (1) = 0.466, *p* = 0.495), “Threat Display” ($\chi2$ (1) = 0.847, *p* = 0.357), “Vocalizations” ($\chi2$ (1) = 1.58, *p* = 0.357), or “Appeasement Behaviors” ($\chi2$ (1) = 0.358, *p* = 0.55) and average baseline JBI in generating post threat induction judgement biases per testing day. Thus, judgement biases following threat induction were not predicted by an interaction between affective reactivity and consistent baseline judgement biases, whether analyzing affective reactivity broadly or each behavioral component separately.

3.9 Summary of Significant Findings

The interaction between age and sex significantly influenced responses to the near positive (P+) stimulus during baseline testing. While adult males showed a lower response to the

P+ stimulus compared to adult females, juvenile males showed a higher response to this stimulus during baseline testing compared to juvenile females. Furthermore, as days of baseline testing progressed, there was a significant reduction in the overall likelihood of responding to the ambiguous stimuli, leading to an overall 27% decrease in the likelihood to respond between Day 1 and Day 3 in the overall sample of juveniles. Analyzing temporal stability in individual differences in judgement biases, a moderate degree of temporal stability in individuals' JBI across the three days of baseline testing was found for juveniles.

The evaluation of juveniles' judgement biases after threat induction via the Human Intruder Test showed a significant reduction in their proportion of responses to all stimuli except the negative signal stimulus. However, the significant decrease in response to the positive signal stimulus (S+) post-threat induction led to the removal of 680 trials [440 control trials, 240 experimental trials] of data from the post-threat induction testing context. After correction of the data, a significant effect of threat induction on reduction of response to the intermediate probe (PI) stimulus, near positive probe (P+) stimulus, and positive signal (S+) stimulus remained.

Despite the overall sample of juveniles shifting responses to ambiguous stimuli from one day to another, a good degree of temporal stability was found for individual differences in judgement biases across the three days. Nevertheless, we did not find rank-order stability in individual differences across testing contexts- the relative ranking of an individual's baseline judgement bias did not predict their ranking of post-threat induction judgement bias relative to other individuals. Therefore, although temporal stability in individual differences in judgement biases was found for each testing context separately, there was no stability in individual differences in judgement biases between contexts.

The relationship between behavioral responses during the Human Intruder Test and judgement bias measures were analyzed to bolster the judgement bias task as a proxy measure for affective traits and states. As trait measures, there was no significant correlation between an individual's average affective reactivity during the Human Intruder task and their average baseline judgement bias. However, correlations between PCA factor scores and an individual's average baseline judgement bias index revealed that individuals with consistently negative judgement biases also showed higher amounts of coo and grunt vocalizations. As state measures, individuals who displayed a higher number of barred teeth, lip smack, threat, and cage shake behaviors showed a decrease in responses to the intermediate probe (PI) stimulus and positive signal (S+) stimulus during subsequent judgement bias testing. Additionally, there was no significant interaction between an individual's average baseline judgement bias and total affective reactivity displayed during the Human Intruder Test in influencing how that individual responded in post-threat induction judgement bias testing. Furthermore, there was no significant correlation between PCA factor scores and average baseline JBI in generating post-threat induction judgement biases. Higher total affective reactivity displayed by an individual during Human Intruder testing did not lead to a greater change in judgement bias from average baseline measures during post-threat induction judgement bias testing.

4. Discussion

4.1 Aims

This study aimed to investigate the validity of the judgement bias task as a reliable proxy for affective traits and states in captive rhesus macaques (*Macaca mulatta*) by focusing on four specific aims. The first focus was on determining whether individuals displayed temporally stable baseline judgement biases suggestive of the influence of trait affect. Our results demonstrated a moderate degree of temporal stability in individual differences in baseline judgement bias, suggesting that the judgement bias task may index features of trait affect. The second focus of the study was to examine whether the judgement bias task could capture variation in momentary negative affective states induced via a standardized threat induction. Results showed a significant negative shift in judgement biases following threat induction compared to baseline judgement bias measures in the overall sample of juveniles, suggesting the task was sensitive to a categorically negative shift in affective state. The third focus was on analyzing the interaction between affective traits and states in influencing the generation of judgement biases after repeated threat inductions. While our results showed a lack of rank-order stability in individual differences in judgement biases across contexts, a good degree of temporal stability was found for an individual's judgement bias after repeated threat induction. The fourth focus was exploring the link between affective reactivity and judgment biases to further validate the judgement bias task as a tool to index affective processing. Results showed no significant correlation between an individual's overall affective reactivity and average baseline judgement bias. However, it was found that individuals with consistently negative baseline judgement biases expressed a higher frequency of vocalizations during the Human Intruder test, suggesting a link between negative trait baseline judgement biases and increased susceptibility to threat induction. Additionally, there was a significant interaction between increased human-directed reactivity during the Human Intruder Test and decreased response to the intermediate probe (PI) and positive signal (S+) stimuli during subsequent post-threat induction judgement bias testing, further implying that greater shifts in affective state due to human reactivity could be captured by alterations to how certain stimuli were processed and appraised.

4.2 General Response During the Baseline Judgement Bias Task

Monkeys were trained to touch a positive signal stimulus (S+) associated with a food reward and refrain from touching a negative stimulus (S-) to avoid receiving a verbal reprimand and a 5-second timeout. Thirty monkeys (12 adults and 18 juveniles) successfully demonstrated a visual discrimination between the two stimuli and consistently selected the correct response according to the stimulus (touch S+, do not touch S-) prior to testing. Animals were subsequently presented with stimuli that were ambiguous in length between the two reference training stimuli, and responses to these stimuli were used to measure judgement bias. On average, monkeys showed a monotonic graded response to the ambiguous stimuli during baseline testing, demonstrating that the animals based their responses to the ambiguous stimuli upon the nearest trained reference stimulus (S+ or S-). That is, as similarity in length to the positive rewarded S+ stimulus increased, the proportion of response to the stimulus increased. Thus, the near positive P+ probe stimulus was interpreted as similar to the positive rewarded S+ stimulus and touched significantly more than the intermediate PI probe stimulus and near negative P- probe stimulus. This difference in the proportion of response to the ambiguous stimuli provides supportive evidence for the ability of the judgement bias task to capture baseline affect-mediated judgement biases in rhesus macaques, similar to that seen in various other species (Bethell et al., 2012; Lagisz et al., 2020; Mendl et al., 2009).

4.3 Factors Influencing Baseline Judgement Bias

In validating the judgement bias task as a proxy for affective processing, it is essential to understand factors contributing to the generation of judgement biases, such as sex and age. While some human studies demonstrate sex differences in affective processing between males and females (e.g., Altemus, 2006; Watson et al., 1988), the majority of judgement bias studies in animals lack an investigation into the effects of sex on response during the task (Lagisz et al., 2020; Whittaker & Barker, 2020). Although a portion of studies have utilized a mixed-sex design, the contrary direction of effects in studies have led to an unclear understanding of sex differences in judgement bias (Barker et al., 2016, 2017; Briefer & McElligott, 2013; Brown et al., 2016; Takeshita & Sato, 2016; Whittaker & Barker, 2020). Our study further complicates our understanding of sex differences in judgement biases, with inverse sex differences in response during baseline testing being found in the two age classes. Adult males demonstrated a more 'negative' judgment bias during baseline testing, while juvenile males showed a more 'positive' judgement bias compared to females in their respective age groups. This effect was not driven by age alone, as there was a lack of significant difference in magnitude of judgement bias effects between adults and juveniles, similar to that found in multiple previous studies (Lagisz et al., 2020). Instead, the interaction between sex and age led to a significant difference in judgement bias.

Very few judgment bias studies have explored the interaction between sex and age in the generation of judgement biases in animals. Still, human studies have found that sex differences play a prominent role in modulating age-related decline in cognition, with males declining faster in specific tasks that rely on attention, inhibition, and memory (Gur & Gur, 2002). Furthermore, it has been postulated that sex differences in rodents may emerge from differences in learning, with males being quicker to extinguish previously learned associations than females (Dalla & Shors, 2009). Therefore, a more 'negative' judgement bias demonstrated by the adult males may be attributed to the 10-year age difference between adults and juveniles causing an age-related cognitive decline in learning and memory, resulting in a reduction in response during the judgement bias task. Sex also plays a role in differences in impulsivity, being higher in female than male rats (Weafer & de Wit, 2014). Additionally, age differences in impulsivity have emerged, with impulsivity being associated with younger rats (Andrzejewski et al., 2011; Burton & Fletcher, 2012) and declining as age increases (Doremus et al., 2004; Laviola et al., 2004). Impulsivity also displays an interaction between age and sex similar to that found in our study, with adolescent male rats being more sensitive to the impulsivity-inducing effects of amphetamine compared to adult males (Burton & Fletcher, 2012). Therefore, the interaction between age and sex in generating differences in impulsivity may explain the shift from a high proportion of response in juvenile males to a lower proportion of response in adult males. Adult males may be less impulsive than juvenile males and display a higher inhibitory control over responding to stimuli, thus demonstrating a more ‘negative’ judgment bias. Therefore, we recommend that more studies be dedicated to understanding the impacts of age and sex on judgement biases in a range of species and that the impact of these factors are considered in experimental design and analyses.

4.4 Temporally Stable Individual Differences in Baseline Judgement Bias

Our results indicate that juvenile monkeys display a moderate degree of temporal stability in baseline judgement biases over the three days of testing, suggesting that the task may capture individual differences in trait affect similar to that discussed in previous scientific reviews (Faustino et al., 2015; Roelofs et al., 2016). While a judgment bias measure collected on a single day of baseline testing provides a momentary 'snapshot' of an animal's affective state, momentary affective states within an individual must arise from a temporally stable baseline trait affect. Thus, the presence of moderately stable individual differences in judgement biases can be conceptualized as the joining together of three 'snapshots' of short-term affective states to reveal a fuller 'picture' of the long-term persistent effects of baseline trait affect on judgement biases within individuals. Our findings of moderate temporal stability, albeit on a shorter time span, suggest that individuals persistently differ in their baseline judgment of ambiguous situations. This temporally persistent difference in baseline judgement biases might then be explained by a relative temporal stability in baseline affective processing (trait affect). However, while baseline trait affect may have possibly been captured by the moderately stable individual differences in baseline judgement biases, an important confirmation of these findings would be the existence of correlation to other consistent measures of affective processing, which is discussed later.

To our knowledge, our results are the first time that consistency in the proportion of responses to stimuli over time has been demonstrated, with other longitudinal studies instead relying on consistency in latency to respond (Asher et al., 2016; Bateson & Nettle, 2015; Clegg et al., 2017; Lecorps, Kappel, et al., 2018; Lecorps, Weary, et al., 2018; Verjat et al., 2021). While a recent meta-analysis of judgement bias studies (Lagisz et al., 2020) revealed no significant impact of measurement type of behavioral responses on judgement bias effects, there are conceptual reasons for the use of different measurement types, with latency and proportion data representing different aspects of cognitive processes and utilized in different ways depending on the task (Lagisz et al., 2020). Most longitudinal judgement bias studies relied on spatial tasks, which require animals to move towards stimuli, rather than the direct presentation of stimuli used in our task. Additionally, proportion and latency data have different statistical distributions that require different transformations and utilize different formulas for calculating judgment bias indexes (Gygax, 2014). While latency data was collected in this study, it was not utilized in data analysis due to large variability in general activity and overall speed of response, causing difficulty in interpretation of data as a measure of judgement biases. Thus, to further explore temporal stability in judgement biases, we encourage future studies to develop judgement bias tasks that utilize proportion data to measure judgement biases.

While our results showed a moderate degree of temporal stability in individual differences in baseline judgement biases, the overall sample of juveniles showed a significant decrease in responses to the ambiguous stimuli across the three days of testing. Individuals displayed a relatively high likelihood of response to the ambiguous stimuli on Day 1 of baseline testing that significantly decreased across the subsequent two additional days of testing (Figure 5). Juveniles were 3.16 times more likely to respond to the ambiguous stimuli on Day 1 of baseline testing compared to Day 3. This behavioral change cannot be linked to a change in affective processing due to no experimental modulation of affect used during baseline testing. Instead, the decrease in response to the ambiguous stimuli may have been due to a loss of ambiguity, as individuals learned that the ambiguous stimuli yielded no reward for response (Roelofs et al., 2016). Baseline judgement bias testing in sheep has already found that repeated testing in the absence of affective modulations can lead to a reluctance to respond to ambiguous stimuli, as animals begin to learn the outcome, or lack thereof, of ambiguous trials (Doyle et al., 2010). Consequently, this lack of temporal stability in response in the overall sample, and possible loss of ambiguity, over the three days of baseline testing calls into question whether temporal stability in individual differences in baseline judgement biases is genuinely being captured. Therefore, additional days of testing would be required to understand whether temporally stable individual differences in judgement bias indeed emerged to validate the use of the task as a proxy measure for affective traits.

4.4.1 Limitations in Measurement of Temporal Stability in Baseline Judgement Bias

While additional days of testing would be required to further elucidate proxy measures of affective traits from judgement biases, researchers must be cognizant of the loss of ambiguity and implement strategies to mitigate this effect. Our study demonstrating a possible loss in ambiguity joins previous studies in hamsters, sheep, pigs, and dogs have also found a reduction in response to ambiguous stimuli over repeated testing sessions (E. J. Bethell & Koyama, 2015; Doyle et al., 2010; Scollo et al., 2014; Starling et al., 2014). Therefore, simply extending the number of testing days would not be advantageous. Instead, repeated testing should be conducted over a span of weeks, similar to methods used in other judgement bias studies (Bateson & Nettle, 2015; Clegg et al., 2017; Lecorps, Weary, et al., 2018; Murphy et al., 2013; Rygula et al., 2013; Starling et al., 2014). This testing design would allow for a greater understanding of temporal stability in judgement bias without sacrificing ambiguity of stimuli utilized to understand judgement bias in the task.

4.5 Judgement Bias Task as a Proxy for Affective States

Overall, threat induction via the Human Intruder Test significantly reduced responses to all stimuli during the judgement bias task. This significant decrease in response to stimuli during judgement bias testing after threat induction provides further evidence that the judgement bias task can capture momentary shifts in affective state through changes in affect-mediated judgement biases. Additionally, using a within-subjects repeated measures design allowed for the reduction in response post-threat induction compared to baseline measurements to reflect a categorically negative affective state within individuals in response to threat induction. Our results share similarities with studies from humans and animals that show that experimental modulations presumed to cause negative affective states are associated with a reduction in the proportion of response to ambiguous stimuli (Mendl et al., 2009). However, this link between our findings and earlier research is further complicated by the finding that threat induction also caused a substantial decrease in response to the unambiguous reinforced positive signal stimulus (S+) during judgement bias testing.

The significant reduction in response to the reinforced positive stimulus (S+) and overall refusal to participate in the judgement bias task after threat induction led to numerous individuals failing to reach the criterion of correct responses on a given testing day. As such, data was corrected to include only testing sessions in which animals reached the 75% criterion of correct responses. After data correction, the significant effect of threat induction on the overall reduction of response in the sample remained. Specifically, there was a significant decrease in the likelihood of responding to the intermediate probe (PI), near positive probe (P+), and positive signal (S+) stimuli post threat induction compared to baseline judgement bias testing after correction. This shift in judgement biases, mainly to the near positive P+ and positive S+ stimuli, may reflect mechanisms sensitive to reward rather than punishment within the monkeys. This is because the anticipated reinforcement yielded by the response to the ambiguous stimuli is theorized to be most strongly influenced by the reinforcer associated with the nearest training stimulus. Thus, presentations of the near positive probe (P+) would tend to elicit an anticipation of the positive reward, and therefore a decreased response may reflect a lower expectation of the positive outcome. Furthermore, a reduction in response to the positive signal (S+) stimulus itself suggests a modulation of reward valuation. However, measures of feeding motivation and satiation, which would allow us to accurately measure changes in food reward valuation, were not implemented in this study. Thus, this significant difference in response to S+ trials after threat induction increases the challenge of interpretation of results from the task and suggests the requirement of further exploration.

4.5.1 Limitations of Task as a Valid Proxy for Affective States

The judgement bias task hinges upon an animal’s ability to consistently respond to the S+ stimulus and refrain from responding to the S- stimulus. These “rules” of the task are imperative for determining what responses, or lack thereof, to the ambiguous stimuli theoretically signify and thus are crucial to interpreting judgment biases. Despite the theoretical and empirical reasons for why judgment biases should not occur at the training stimuli, our study joins multiple other studies that have found large effects with regard to how animals respond to the positive and negative reference stimuli (e.g., Deakin et al., 2018; Horváth et al., 2016; Zidar et al., 2018). Even the first judgement bias study, which provided the foundation from which numerous studies have adapted, found fewer responses to the positive stimulus itself after negative affect induction (Harding et al., 2004). However, there seems to be a lack of consensus among these studies on what factors may cause changes in response to the trained positive S+ stimulus.

It has been theorized that a reduction in responses to the trained positive stimulus may be related to shifts in arousal, attention, activity, motivation, satiation, or stress-related impairment of task performance (Mendl et al., 2009). Our results may reflect an increase in satiation prior to post-threat induction testing leading to a decrease in food valuation and thus a lower propensity to respond to the positive S+ stimulus. Our study found a possible association between testing order and motivation to participate in post-threat induction judgement bias testing, perhaps due to satiation. Juveniles tested earlier in the sequence tended to have minimal participation in the task, notably refusing to touch the S+ stimulus when presented. This may have been due to post-threat induction judgement bias testing being conducted earlier in the day and thus closer to the morning feeding time by the husbandry staff, possibly altering the monkey's food motivation and interest in responding to the S+ stimulus and task as a whole. However, the alternative explanations of results being caused by changes in arousal, attention, activity, or risk-taking behavior cannot be ruled out and may also work in tandem with the effects of satiation. Thus, to address these limitations, future studies should implement independent tests of feeding motivation, anhedonia, engagement in the task, and activity in addition to measures of judgement biases to aid in the interpretation of results.

Furthermore, a significant reduction in response to the S+ stimulus may reflect a sensitivity limit of the task when an acute threatening stressor is used. Judgement bias studies have often used chronic affect manipulations to evaluate shifts in judgement biases consistent with temporally extended affective states, with manipulations usually lasting days or weeks before judgement biases are measured (Lagisz et al., 2020; Mendl et al., 2009). However, the use of the short-term, acute threat induction prior to the task may have led to such a pronounced negative shift in affective states that performance in the task was impaired. Other studies that have relied on short-term, acute experimental modulation have also found contrasting effects on judgement biases (e.g., Doyle, Fisher, et al., 2010; Sanger et al., 2011), suggesting this task may be better suited for longer-term affect manipulations. To expand upon the connection between affect manipulation and the generation of measurable judgement biases, future studies should implement an experimental design that compares the effects of both acute and chronic affect manipulation on the generation of judgement biases in the task.

4.6 Interactions between Affective States and Affective Traits in Generation of Post Threat Induction Judgement Biases

In humans (Mathews & MacLeod, 1994b) and animals (Faustino et al., 2015; Roelofs et al., 2016), it has been suggested that both state and trait variables interact to provide an essential contribution to affect-mediated decision-making patterns. Additionally, the repeated generation of affective states interacts with trait affect to cause dynamic changes in affective processing over time (Bliss-Moreau & Rudebeck, 2021). In our study, we strived to understand whether the repeated generation of negative affective states via recurrent threat induction caused consistent shifts in baseline trait judgement biases reflective of a dynamic shift in trait affect. Therefore, to understand whether the judgement bias task can capture this interaction between affective states and traits, both rank-order stability across contexts and temporal stability in individual differences in post-threat induction judgment biases were analyzed. Our results showed a lack of rank-order stability in individual differences in judgement biases across testing contexts, with the relative ranking of individual differences in baseline judgement biases not predicting how individuals will differ in response post threat induction relative to others. However, despite this lack of context stability in individual differences, there was a good degree of temporal stability in individual differences in post-threat induction judgement biases. Similar to recommendations stated earlier, however, additional testing days would allow for a more robust understanding of the temporality of the dynamic shift in an individual's judgement bias due to repeated generation of affective states. Nevertheless, again researchers must be careful about the effects of increased testing on the loss of ambiguity, although our results do not suggest the presence of this effect. Overall, results suggest that temporally stable individual differences in post-threat induction judgment biases do not simply depend on trait baseline judgement biases alone, but also on how influential threat induction is on shifts in an individual's affective states. Therefore, to understand how threat induction influenced individuals' affective processing and further bolster judgement bias results, the relationship between affective reactivity and judgement biases was explored.

4.7 Relationship between Affective Reactivity and Judgement Bias

Our study supports previous literature that has extensively shown that threat induction via the Human Intruder Test causes an increase in affective reactivity (Bliss-Moreau et al.., 2021; Bliss-Moreau & Baxter, 2018; Gottlieb & Capitanio, 2013; Kalin & Shelton, 1989, 1998). Of importance to this study, measures of affective reactivity from this task can reflect both a trait and state measure. Affective reactivity may represent a trait, with some animals being consistently more reactive overall than others across multiple testing days. Affective reactivity can also be a state, with individuals displaying a higher magnitude of adverse reactions on a single day of testing compared to their reactions on the other two days of testing. This reflection of both a trait and a state allowed for the exploration of the correlation between affective reactivity and both trait and state judgement bias measures. Additionally, the use of a Principal Component Analysis identified four components of affective reactivity- “Threat Display”, “Behavioral Tension”, “Appeasement Behaviors”, and “Vocalizations”- with which we were able to conduct additional analyses of these affective components’ relationship to judgement biases. An important means of validating the judgement bias task in animals is the correlation of additional behavioral tests assumed to assess affective states and/or traits (Roelofs et al., 2016). To our knowledge, this is the first study to directly assess the extent to which affective reactivity assessed through the Human Intruder task is associated with judgement biases in rhesus macaques.

4.7.1 Relationship Between Human Intruder Measures and Baseline Judgement Biases

As a trait measure, there was a lack of significant correlation between an individual’s average affective reactivity during the Human Intruder Test and average baseline judgement bias. These results contradict the prediction that individuals displaying a consistently negative baseline judgement bias would also demonstrate a consistently higher average affective reactivity following threat induction. Multiple studies in animals have explored the relationship between individual variation in judgement biases and responses to uncontrollable stressors (Drozd et al., 2017; Enkel et al., 2010; Lecorps et al., 2020, 2021; Lecorps, Kappel, et al., 2018;

Rygula et al., 2013) and have found that negative judgement biases have been linked to a higher intensity of responses to stressors. Studies in humans have also shown that trait pessimism is linked to a stronger reaction to emotional challenges (Costa & McCrae, 1980; Scheier & Carver, 1992). However, our results demonstrated no such association, which may call into question the proposed link between consistent judgement biases and trait affect, or it may simply indicate that trait affect influences behavior differently in these two tasks. Nevertheless, results found that only a single component of affective reactivity, vocalizations, correlated with an individual's baseline trait judgement bias.

Our results showed that individuals with a consistently negative baseline judgement bias displayed consistently higher amounts of vocalizations, specifically coos and grunts, during the Human Intruder Test (Figure 10). This relationship between a consistent negative judgement bias and increased vocalizations has also been found in a previous study in dairy calves (Lecorps, Kappel, et al., 2018). Vocalizations are affective in nature, being validated as a measure of high arousal (Briefer et al., 2015) and shown to be displayed by juvenile rhesus macaques during distress (Maestripieri et al., 2000; Maestripieri & Call, 1996; Tomaszycki et al., 2001). Both coos and grunts are vocalizations occurring in social interactions in nonhuman primates, with coos associated with separation from conspecifics (Hauser, 1991) and grunts associated with approaching others (Seyfarth & Cheney, 1997) and during post-conflict reconciliation (Silk et al., 1996). Therefore, results may suggest that individuals with consistently negative judgement biases may be more affected by the social isolation of the Human Intruder Test and thus have consistently higher levels of vocalization. This link between consistently negative baseline judgement biases and overall higher frequency of vocalizations in individuals may be capturing the influence of trait affect on both the generation of overall decision-making and a greater susceptibility to the stress of separation from conspecifics.

4.7.2 Relationship Between Human Intruder Measures and Post Threat Induction Decision-Making Per Day

While an overall increased affective reactivity during the Human Intruder Test did not predict a lower response to stimuli on subsequent judgement bias testing, components of reactivity, specifically higher “Threat Display” and “Appeasement” behaviors directed towards the unfamiliar human, were found to predict a significant decrease in responses to the intermediate probe (PI) and positive signal (S+) stimuli during subsequent judgment bias testing. It was established earlier that our judgement bias task captured momentary shifts in affective processing in response to threat induction, and these results further elucidate how an individual’s behavioral response to a threat may generate specific judgment biases during the task. The intermediate probe stimulus (PI) should theoretically be the stimulus with the greatest amount of ambiguity and thus uncertainty for responses. Therefore, individuals displaying a more robust human-directed reactivity may have shifted their interpretation of this stimulus away from an anticipation of the positive rewarded event and thus decreased response as a result. Additionally, an effect at the rewarded positive signal (S+) stimulus might further support this theory by suggesting that individuals with greater human-directed reactivity demonstrated a decreased expectation of the positive rewarded event and shifted expectations toward the negative outcome. The previously discussed significant reduction in the proportion of response to the near positive probe (P+) stimulus after threat induction may further strengthen the connection that higher human-directed reactivity to the threat caused a reduction in the expectation of the positive reward. However, there was no significant effect of higher human-directed reactivity on a reduction in response to the P+ stimulus. Thus, an alternative explanation for the results may also simply be that individuals who displayed higher human-directed reactivity in the Human Intruder Test were also influenced by the human experimenter presenting the stimuli when deciding how to respond to the PI and S+ stimuli in subsequent judgement bias testing.

4.7.3 Interaction Between Trait Baseline Judgment Biases and State Affective Reactivity in Generating State-Like Shifts in Post Threat Induction Judgement Biases

Our study sought to understand how an individual's state affective reactivity interacts with their trait average baseline judgement bias to generate post-threat induction judgement biases per day of testing. We predicted that individuals experiencing higher affective reactivity would demonstrate a more negative judgement bias post-threat induction if they also displayed a consistently negative baseline judgement bias. This *a priori* assumption stemmed from human studies, which demonstrated that elevated state anxiety only generated attention biases if individuals also had high trait vulnerability to anxiety (C. MacLeod & Mathews, 1988). However, our results showed that higher state affective reactivity displayed by an individual did not interact with an individual's baseline judgment bias trait to predict how individuals respond post threat induction. Furthermore, none of the four behavioral components- "Behavioral Tension", "Threat Display", "Vocalizations", "Appeasement Behaviors"- interacted with an individual's trait baseline judgement bias to predict how they would respond post threat induction. Instead, results showed increased variability in how an individual's total affective reactivity interacted with their baseline judgement bias trait to generate shifts in judgement biases. This suggests that there are individual differences in how influential threat induction is on baseline decision-making when generating judgement biases. While an individual's judgement bias is not static and instead shifts in response to threat induction, affective behaviors from the

Human Intruder Test alone cannot explain how decision-making will shift over time. Additionally, coupled with the lack of rank-order reliability in individual differences in judgement biases across contexts, results may suggest that an individual's baseline judgement bias does not predict how influential threat induction will be on an individual's affective processing, and thus the generation of judgement biases.

4.8 Additional Limitations

Although some limitations of this study have been highlighted in their respective sections, additional limitations arose due to the experimental design. A "go/no-go" task was adopted to evaluate judgement biases in this thesis, however a meta-analysis found that "active choice go/go" tasks tended to yield larger judgement bias effects than "go/no-go" tasks (Lagisz et al., 2020). While a "go/no-go" task involves the suppression of response, an "active choice go/go" task requires animals to choose between two alternative responses (e.g., touch the left or right stimuli presented). Smaller effects in a "go/no-go" task may be due to an increased susceptibility of these tasks to the influences of Pavlovian action predispositions (e.g., go for reward; no-go to avoid punishment; Guitart-Masip et al., 2014; Jones et al., 2017), which could unintentionally bias responding (Mendl and Paul, 2020) and obscure the effects of affect manipulation. An additional significant issue in the "go/no-go" task is that subjects may sometimes refrain from responding for reasons unrelated to affect manipulations (e.g., failing to detect or attend to a cue, variation in impulsivity and behavioral inhibition; Bethell, 2015; Jones et al., 2018; Roelofs et al., 2016), making "go/no go" tasks less dependable. Although currently "go/no-go" tasks are predominantly used to measure judgement biases (Lagisz et al., 2020), the use of a "go/go active choice" task would require animals to engage in the task to produce meaningful decisions that allow for enhanced interpretability.

Furthermore, transitioning from baseline judgement bias testing in a social context- where animals could interact with, see, and hear other conspecifics before and during testing- to individual testing post-threat induction may have significantly altered judgement bias results. The testing of animals individually in the judgement bias task was previously found to require extensive habituation and pre-training before judgement biases could be adequately assessed (Krasheninnikova & Schneider, 2014; Murphy et al., 2013, 2015). While social isolation may be thought of as a part of the negative affective induction, the testing of animals individually may have also contributed to the challenge of assessing judgement bias post-threat induction. This is further supported by our results that showed that individuals with a consistently negative baseline judgement bias displayed consistently higher amounts of social vocalizations, possibly reflecting the increased susceptibility to social isolation stress in these individuals.

Another confounding factor that may have influenced responses during the post-threat induction judgement bias testing is the involvement of a human experimenter in both the Human Intruder and judgment bias tests. The Human Intruder Test involved threat induction via an unfamiliar human in close proximity to and making eye contact with the monkeys. Similarly, the judgement bias task involved a human experimenter in close proximity presenting stimuli to the monkeys. The same human experimenter presented all the stimuli during judgement bias training and testing sessions to allow monkeys to develop familiarity, and there seemed to be no effects of the human experimenter on responses during baseline testing. However, exposure to threat induction via the unfamiliar human during the Human Intruder test may have caused changes in the interpretation of the familiar human experimenter, thus causing a general suppressive effect on responses to stimuli. Voluntary touchscreen testing devoid of human intervention, similar to that conducted by Bethell et al. (2012), may have alleviated any suppressive effects on response due to the human presence. Self-initiation of trials would also allow monkeys to have control over the task, which is expected to increase motivation to participate (Hintze et al., 2018). However, it is unclear whether animals would be willing to self-initiate testing after experiencing a threatening stressor or be motivated to continue to participate once testing has commenced.

Lastly, analyses of the relationship between judgement biases and affective reactivity must be interpreted with caution due to statistical limitations. The high dimensionality of both the judgement bias and affective reactivity datasets necessitated the use of transformation methods that allowed for a low-dimensional representation of both datasets to explore their relationship. However, the use of a single judgement bias index to represent responses to the five different stimuli may have resulted in an oversimplification that did not allow for differentiation of which specific stimuli an individual showed a higher proportion of responses to, which is conceptually relevant to interpreting judgement biases. Additionally, a PCA was used for the affective reactivity data to group affective behaviors into the four different components, despite a low KMO and small subject-to-variable ratio suggesting that affective behaviors may not have been best suited for a factor analysis. However, despite the oversimplification of the datasets for analyses, our current results suggest that the use of different statistical methods (e.g., Kendall rank correlation tests between each stimuli and affective behavior separately using a Holm-Bonferroni correction for multiple comparisons) may still yield a similar relationship between judgement biases and affective reactivity. An additional statistical consideration is that the small sample of juveniles used in analyses may have reduced statistical power and caused difficulty in correctly identifying significant relationships between judgement biases and affective reactivity. Finally, it must be noted that all statistical analyses conducted were autodidactic. While a statistician was consulted throughout the process, the intricate nature of statistics and the novelty of my statistical understanding may have resulted in unforeseen errors.

4.9 Future Considerations

While much has been explored in this thesis to validate the judgement bias task as a reliable measure of state and trait affective processing, additional considerations must be considered when designing future studies to further explore many of the concepts touched upon. First, the interaction between age and sex in generating differences in baseline judgement biases highlights the need for future judgment bias studies to consider these factors in experimental design to allow for a deeper understanding of their impact. Additional days of judgment bias testing must also be implemented to allow for a more concrete measure of trait affect. However, researchers must be cognizant of the possible effects of additional testing days on the loss of ambiguity and implement strategies to mitigate this effect. To allow for a deeper understanding of the link between affective valence and the generation of judgement biases within an individual, multiple positive and negative experimental modulations of affective state should be implemented along with baseline judgement bias measures. Furthermore, studies should also further explore the differences between the effects of acute versus chronic experimental modulation on generating judgement biases. Additional behavioral and physiological proxies of affective processing should be implemented to bolster judgement bias findings. Judgement biases offer a singular viewpoint of affective processing and should be paired with multiple other behavioral and physiological metrics to allow for a clearer picture of affective processing within an individual. Lastly, a lack of motivation to engage in the task was observed in numerous individuals, causing the judgement bias task to theoretically break down. The lack of engagement caused an inability to interpret ‘no go’ responses as either deliberate inhibition of response due to affective processing or simply disengagement due to other factors. Therefore, more go-go active choice judgement bias tasks should be developed that require animals to make meaningful decisions during the task rather than disengage. While findings from this study have provided insight into the link between judgment biases and affective processing in rhesus macaques, they have also unearthed additional concepts that must be explored in greater detail in future studies.

5. Conclusion

Over the last decade, evidence has emerged suggesting that measures of judgement biases may prove to be a reliable proxy for affective processing in a range of animals. However, there is a current lack of studies focusing on elucidating underlying affective mechanisms that generate judgement biases. While numerous studies have explored the capturing of affective states by the task, further investigation of trait differences in judgement biases and their reflection on affective traits was needed to validate the judgement bias task. Furthermore, there has been a lack of judgement bias studies conducted in non-human primates, specifically rhesus macaques, which have the potential to explore the evolutionary development of the relationship between cognition and affect in humans. Therefore, this thesis aimed to examine the underlying affective mechanisms that generate judgement biases in captive rhesus macaques (*Macaca mulatta*) to validate the judgement bias task as a reliable proxy for affective processing.

This thesis was designed to focus on four specific objectives that explored the connection between cognition and affective processing in captive rhesus macaques. The first focus was on investigating the emergence of temporally stable differences in baseline judgement biases indicative of the influence of trait affect. The second focus was to assess whether the judgement bias task could capture the generation of momentary negative affective states via an acutely negative threat induction. The third focus was on analyzing whether the interaction between affective traits and states is involved in the generation of judgement biases after repeated threat induction. The fourth focus explored the link between affective reactivity and judgment biases to further validate judgement biases as a proxy for affective processing.

Results from this thesis demonstrated the capturing of temporally stable individual differences in baseline judgement biases, suggesting that the emergence of trait judgement biases in the task may index trait affect. Furthermore, the relationship between affective reactivity and judgement biases demonstrated that negative trait baseline judgement biases corresponded with a consistently higher frequency of social vocalizations. Results also showed a capturing of categorically negative affective states within individuals in response to threat induction through a significant reduction in response to the intermediate probe (PI), near positive probe (P+), and positive rewarded signal stimulus (S+). All individuals, especially those who expressed a high frequency of human-directed affective behaviors during the Human Intruder Test, showed a negative shift in judgement biases post-threat induction. Temporal stability of individual differences in post-threat induction judgement biases was also found, suggesting a possible interaction between affective traits and affective states. However, the lack of rank-order stability in individual differences across contexts, coupled with no interaction between trait baseline judgment biases and state affective reactivity, suggests that an individual’s trait baseline judgment bias alone cannot predict how judgement biases will shift in response to threat induction.

This thesis adds to the growing number of studies utilizing the judgement bias task as a proxy for affective processing by capturing affect-mediated judgement biases in captive rhesus macaques. Through identifying temporally stable individual differences in baseline judgement biases, we expand our understanding of how measures from the judgement bias task can provide a proxy for individual differences in affective traits. Furthermore, the findings of individuals with consistently negative baseline judgement biases having higher contact calls during the Human Intruder Test provides a deeper insight into the linkage between sociality and judgement biases, with consistent negative judgement biases reflecting a possible susceptibility to social separation stress. The use of a within-subjects repeated measure design allowed for a significant negative shift in judgement biases after threat induction to reflect the task capturing a categorically negative shift in affective state within animals. However, the pronounced negative shift in affect due to the acutely negative threat induction led to a drastic impairment in the performance of the task in multiple individuals, demonstrating a fault in the judgment bias task. Lastly, findings of the temporal stability of individual differences in judgement biases post-threat induction reveal the capturing of the interaction between affective traits and states in generating judgement biases. However, the lack of rank-order stability in individual differences in judgment biases across contexts and variability in how influential threat induction was on affective processing indicate that an individual's baseline judgement bias trait alone cannot predict how influential threat induction will be on affective processing.

While this thesis has shed new light on the connection between judgement biases and affective processing, this new light has also revealed multiple paths that need further exploration. First, the influence of age and sex on the training and testing of the judgement bias task underscores the need to develop judgement bias studies focused on elucidating the underlying mechanisms that generate these differences. Furthermore, if animal welfare science is to be moved forward, future studies should work to better understand whether trait differences in judgment biases influence an animal’s ability to cope with challenges to their welfare. The capturing of affect-mediated judgment biases in rhesus macaques similar to those found in humans also reveals an immense opportunity to explore the evolutionary mechanisms that generate the connection between affect and cognition in both human and non-human animals. As of now, several questions about how this connection evolved remain to be answered. Thus, considerably more judgement bias studies must be conducted in a range of non-human primate species to allow for empirical testing of the evolution of emotion through a phylogenetic comparative approach. However, if scientists are to address these concepts, future studies should be carried out using the “go-go active choice” task rather than the “go/no-go” task used in this thesis. Lack of engagement in our “go/no-go” task caused numerous individuals to be removed from the task, causing the removal of significant information that may better reveal the intricate connection between affect and cognition. Instead, a “go-go active choice” task requires animals to engage with the task to provide meaningful decisions that allow for better interpretability and less data removal. Additionally, judgement biases offer a limited understanding of affective processing if not paired with other behavioral and physiological proxies. Thus, to substantiate the specificity of judgement bias measures for indexing affect, a greater focus must be placed on understanding the relationship between judgement biases and other measures of affective processing. By addressing these considerations, future judgement bias studies may develop more accurate assessments of affective processing that offer a promising insight into the internal experiences of animals.

6. References

Altemus, M. (2006). Sex differences in depression and anxiety disorders: Potential biological determinants. *Hormones and Behavior*, *50*(4), 534–538. https://doi.org/10.1016/j.yhbeh.2006.06.031

Anderson, D. J., & Adolphs, R. (2014). A Framework for Studying Emotions across Species. *Cell*, *157*(1), 187–200. https://doi.org/10.1016/j.cell.2014.03.003

Anderson, J. R. (2010). Non-Human Primates: A Comparative Developmental Perspective on Yawning. In O. Walusinski (Ed.), *Frontiers of Neurology and Neuroscience* (Vol. 28, pp. 63–76). KARGER. https://doi.org/10.1159/000307082

Andrzejewski, M. E., Schochet, T. L., Feit, E. C., Harris, R., McKee, B. L., & Kelley, A. E. (2011). A comparison of adult and adolescent rat behavior in operant learning, extinction, and behavioral inhibition paradigms. *Behavioral Neuroscience*, *125*(1), 93–105. https://doi.org/10.1037/a0022038

Asher, L., Friel, M., Griffin, K., & Collins, L. M. (2016). Mood and personality interact to determine cognitive biases in pigs. *Biology Letters*, *12*(11), 20160402. https://doi.org/10.1098/rsbl.2016.0402

Aue, T., & Okon-Singer, H. (2015). Expectancy biases in fear and anxiety and their link to biases in attention. *Clinical Psychology Review*, *42*, 83–95.

Baldock, N. M., & Sibly, R. M. (1990). Effects of handling and transportation on the heart rate and behaviour of sheep. *Applied Animal Behaviour Science*, *28*(1), 15–39. https://doi.org/10.1016/0168-1591(90)90044-E

Baldock, N. M., Sibly, R. M., & Penning, P. D. (1988). Behaviour and seasonal variation in heart rate in domestic sheep, Ovis aries. *Animal Behaviour*, *36*(1), 35–43. https://doi.org/10.1016/S0003-3472(88)80247-1

Barker, T., Bobrovskaya, L., Howarth, G., & Whittaker, A. (2017). Female rats display fewer optimistic responses in a judgment bias test in the absence of a physiological stress response. *Physiology & Behavior*, *173*, 124–131.

Barker, T., Howarth, G., & Whittaker, A. (2016). The effects of metabolic cage housing and sex on cognitive bias expression in rats. *Applied Animal Behaviour Science*, *177*, 70–76.

Barrett, L. F. (2004). Feelings or Words? Understanding the Content in Self-Report Ratings of Experienced Emotion. *Journal of Personality and Social Psychology*, *87*(2), 266–281. https://doi.org/10.1037/0022-3514.87.2.266

Barrett, L. F. (2006a). Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2006b). Valence is a basic building block of emotional life. *Journal of Research in Personality*, *40*(1), 35–55. https://doi.org/10.1016/j.jrp.2005.08.006

Barrett, L. F. (2006c). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2017a). *How emotions are made: The secret life of the brain.* (pp. xv, 425). Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(11), 1833. https://doi.org/10.1093/scan/nsx060

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a Psychological Primitive. In *Advances in Experimental Social Psychology* (Vol. 41, pp. 167–218). Elsevier. https://doi.org/10.1016/S0065-2601(08)00404-8

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The Experience of Emotion. *Annual Review of Psychology*, *58*, 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bateson, M., Desire, S., Gartside, S. E., & Wright, G. A. (2011). Agitated Honeybees Exhibit Pessimistic Cognitive Biases. *Current Biology*, *21*(12), 1070–1073. https://doi.org/10.1016/j.cub.2011.05.017

Bateson, M., & Matheson, S. (2007). *Performance on a categorisation task suggests that removal of environmental enrichment induces 'pessimism' in captive European starlings (Sturnus vulgaris)*. 4.

Bateson, M., & Nettle, D. (2015). Development of a cognitive bias methodology for measuring low mood in chimpanzees. *PeerJ*, *3*, e998. https://doi.org/10.7717/peerj.998

Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, *55*(1), 30–40. https://doi.org/10.1016/j.bandc.2003.04.001

Bethell, E., Holmes, A., Maclarnon, A., & Semple, S. (2012). Cognitive bias in a non-human primate: Husbandry procedures influence cognitive indicators of psychological well-being in captive rhesus macaques. *Animal Welfare*, *21*(2), 185–195. https://doi.org/10.7120/09627286.21.2.185

Bethell, E. J. (2015). A "How-To" Guide for Designing Judgment Bias Studies to Assess Captive Animal Welfare. *Journal of Applied Animal Welfare Science*, *18*(sup1), S18–S42. https://doi.org/10.1080/10888705.2015.1075833

Bethell, E. J., & Koyama, N. F. (2015). Happy hamsters? Enrichment induces positive judgement bias for mildly (but not truly) ambiguous cues to reward and punishment in Mesocricetus auratus. *Royal Society Open Science*, *2*(7), 140399. https://doi.org/10.1098/rsos.140399

Blanchard, R. J., Hebert, M., Sakai, R. R., McKittrick, C., Henrie, A., Yudko, E., McEwen, B. S., & Blanchard, D. C. (1998). Chronic social stress: Changes in behavioral and physiological indices of emotion. *Aggressive Behavior*, *24*(4), 307–321. https://doi.org/10.1002/(SICI)1098-2337(1998)24:4<307::AID-AB6>3.0.CO;2-F

Bliss-Moreau, E. (2017). Constructing nonhuman animal emotion. *Current Opinion in Psychology*, *17*, 184–188. https://doi.org/10.1016/j.copsyc.2017.07.011

Bliss-Moreau, E., & Baxter, M. G. (2018). Estradiol treatment in a nonhuman primate model of menopause preserves affective reactivity. *Behavioral Neuroscience*, *132*(4), 224–229. https://doi.org/10.1037/bne0000253

Bliss-Moreau, E., & Moadab, G. (2016). Variation in behavioral reactivity is associated with cooperative restraint training efficiency. *Journal of the American Association for Laboratory Animal Science*, *55*(1), 41–49.

Bliss-Moreau, E., & Moadab, G. (2017). The faces monkeys make. In *The science of facial expression.* (pp. 153–171). Oxford University Press.

Bliss-Moreau, E., & Rudebeck, P. H. (2021). Animal models of human mood. *Neuroscience & Biobehavioral Reviews*, *120*, 574–582. https://doi.org/10.1016/j.neubiorev.2020.06.024

Bliss-Moreau, E., Santistevan, A. C., Bennett, J., Moadab, G., & Amaral, D. G. (2021). Anterior Cingulate Cortex Ablation Disrupts Affective Vigor and Vigilance. *Journal of Neuroscience*, *41*(38), 8075–8087. https://doi.org/10.1523/JNEUROSCI.0673-21.2021

Bliss-Moreau, E., Toscano, J. E., Bauman, M. D., Mason, W. A., & Amaral, D. G. (2010). Neonatal amygdala or hippocampus lesions influence responsiveness to objects. *Developmental Psychobiology*, *52*(5), 487–503.

Bliss-Moreau, E., Toscano, J. E., Bauman, M. D., Mason, W. A., & Amaral, D. G. (2011). Neonatal amygdala lesions alter responsiveness to objects in juvenile macaques. *Neuroscience*, *178*, 123–132. https://doi.org/10.1016/j.neuroscience.2010.12.038

Blumberg, M. S., & Sokoloff, G. (2001). Do infant rats cry? *Psychological Review*, *108*(1), 83.

Boissy, A. (1995). Fear and Fearfulness in Animals. *The Quarterly Review of Biology*, *70*(2), 165–191.

Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., & Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior*, *92*(3), 375–397. https://doi.org/10.1016/j.physbeh.2007.02.003

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148. https://doi.org/10.1037/0003-066X.36.2.129

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. *Cognitive Neuroscience of Emotion*, *25*, 49–59.

Bradshaw, R. H., Parrott, R. F., Forsling, M. L., Goode, J. A., Lloyd, D. M., Rodway, R. G., & Broom, D. M. (1996). Stress and travel sickness in pigs: Effects of road transport on plasma concentrations of cortisol, beta-endorphin and lysine vasopressin. *Animal Science*, *63*(3), 507–516. https://doi.org/10.1017/S135772980001540X

Briefer, E. F., & McElligott, A. G. (2013). Rescued goats at a sanctuary display positive mood after former neglect. *Applied Animal Behaviour Science*, *146*(1–4), 45–55.

Briefer, E. F., Sypherd, C. C.-R., Linhart, P., Leliveld, L. M. C., Padilla de la Torre, M., Read, E. R., Guérin, C., Deiss, V., Monestier, C., Rasmussen, J. H., Špinka, M., Düpjan, S., Boissy, A., Janczak, A. M., Hillmann, E., & Tallet, C. (2022). Classification of pig calls produced from birth to slaughter according to their emotional valence and context of production. *Scientific Reports*, *12*(1), 3409. https://doi.org/10.1038/s41598-022-07174-8

Briefer, E. F., Tettamanti, F., & McElligott, A. G. (2015). Emotions in goats: Mapping physiological, behavioural and vocal profiles. *Animal Behaviour*, *99*, 131–143.

Broom, D. M., & Johnson, K. G. (1993). *Stress and Animal Welfare*. Springer Netherlands. https://doi.org/10.1007/978-94-024-0980-2

Brown, G. R., Cullum, P., Martin, S., & Healy, S. D. (2016). Sex differences in performance on a cognitive bias task in Norway rats. *Behavioural Processes*, *133*, 52–55.

Brydges, N. M., Leach, M., Nicol, K., Wright, R., & Bateson, M. (2011). Environmental enrichment induces optimistic cognitive bias in rats. *Animal Behaviour*, *81*(1), 169–175. https://doi.org/10.1016/j.anbehav.2010.09.030

Budaev, S. V. (2010). Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, *116*(5), 472–480.

Burke, M., & Mathews, A. (1992). Autobiographical memory and clinical anxiety. *Cognition and Emotion*, *6*(1), 23–35. https://doi.org/10.1080/02699939208411056

Burman, O. H. P., & Mendl, M. T. (2018). A novel task to assess mood congruent memory bias in non-human animals. *Journal of Neuroscience Methods*, *308*, 269–275. https://doi.org/10.1016/j.jneumeth.2018.07.003

Burton, C. L., & Fletcher, P. J. (2012). Age and sex differences in impulsive action in rats: The role of dopamine and glutamate. *Behavioural Brain Research*, *230*(1), 21–33. https://doi.org/10.1016/j.bbr.2012.01.046

Caeiro, C. C., Burrows, A. M., & Waller, B. M. (2017). Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions? *Applied Animal Behaviour Science*, *189*, 66–78.

Caeiro, C., Waller, B., Zimmerman, E., Burrows, A., & Davila Ross, M. (2013). OrangFACS: A muscle-based movement coding system for facial communication in Orangutans. *International Journal of Primatology, 34*, 115–129.

Carver, C. S., Scheier, M. F., & Segerstrom, S. C. (2010). Optimism. *Clinical Psychology Review*, *30*(7), 879–889. https://doi.org/10.1016/j.cpr.2010.01.006

Charbonneau, J. A., Maister, L., Tsakiris, M., & Bliss-Moreau, E. (2022). Rhesus monkeys have an interoceptive sense of their beating hearts. *Proceedings of the National Academy of Sciences*, *119*(16), e2119868119. https://doi.org/10.1073/pnas.2119868119

Clegg, I. L. K., Rödel, H. G., & Delfour, F. (2017). Bottlenose dolphins engaging in more social affiliative behaviour judge ambiguous cues more optimistically. *Behavioural Brain Research*, *322*, 115–122. https://doi.org/10.1016/j.bbr.2017.01.026

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis, 2nd ed.* (pp. xii, 430). Lawrence Erlbaum Associates, Inc.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, *38*(4), 668.

Crump, A., Arnott, G., & Bethell, E. J. (2018). Affect-Driven Attention Biases as Animal Welfare Indicators: Review and Methods. *Animals : An Open Access Journal from MDPI*, *8*(8), 136. https://doi.org/10.3390/ani8080136

Dalla, C., & Shors, T. J. (2009). Sex differences in learning processes of classical and operant conditioning. *Physiology & Behavior*, *97*(2), 229–238.

Daros, R. R., Costa, J. H. C., Keyserlingk, M. A. G. von, Hötzel, M. J., & Weary, D. M. (2014). Separation from the Dam Causes Negative Judgement Bias in Dairy Calves. *PLOS ONE*, *9*(5), e98429. https://doi.org/10.1371/journal.pone.0098429

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

Dawkins, M. S. (1990). From an animal's point of view: Motivation, fitness, and animal welfare. *Behavioral and Brain Sciences*, *13*(1), 1–9. https://doi.org/10.1017/S0140525X00077104

Dawkins, M. S. (1993). *Through our eyes only?: The search for animal consciousness*.

De Waal, F. B., & Yoshihara, D. (1983). Reconciliation and redirected affection in rhesus monkeys. *Behaviour*, *85*(3–4), 224–241.

Deakin, A., Mendl, M., Browne, W. J., Paul, E. S., & Hodge, J. J. L. (2018). State-dependent judgement bias in Drosophila: Evidence for evolutionarily primitive affective processes. *Biology Letters*, *14*(2), 20170779. https://doi.org/10.1098/rsbl.2017.0779

Doremus, T. L., Varlinskaya, E. I., & Spear, L. P. (2004). Age-Related Differences in Elevated Plus Maze Behavior between Adolescent and Adult Rats. *Annals of the New York Academy of Sciences*, *1021*(1), 427–430. https://doi.org/10.1196/annals.1308.057

Doyle, R. E., Vidal, S., Hinch, G. N., Fisher, A. D., Boissy, A., & Lee, C. (2010). The effect of repeated testing on judgement biases in sheep. *Behavioural Processes*, *83*(3), 349–352. https://doi.org/10.1016/j.beproc.2010.01.019

Drozd, R., Rojek-Sito, K., & Rygula, R. (2017). The trait 'pessimism' does not interact with cognitive flexibility but makes rats more vulnerable to stress-induced motivational deficits: Results from the attentional set-shifting task. *Behavioural Brain Research*, *335*, 199–207. https://doi.org/10.1016/j.bbr.2017.08.028

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition & Emotion*, *21*(6), 1184–1211. https://doi.org/10.1080/02699930701437931

Durán, J. I., & Fernández-Dols, J.-M. (2021). Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. *Emotion*, *21*(7), 1550–1569. https://doi.org/10.1037/emo0001015

Ekman, P., & Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., & others. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712.

Enkel, T., Gholizadeh, D., von Bohlen und Halbach, O., Sanchis-Segura, C., Hurlemann, R., Spanagel, R., Gass, P., & Vollmayr, B. (2010). Ambiguous-Cue Interpretation is Biased Under Stress- and Depression-Like States in Rats. *Neuropsychopharmacology*, *35*(4), 1008–1015. https://doi.org/10.1038/npp.2009.204

Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychology*, *100*(2), 144–150. https://doi.org/10.1037//0021-843x.100.2.144

Faustino, A. I., Oliveira, G. A., & Oliveira, R. F. (2015). Linking appraisal to behavioral flexibility in animals: Implications for stress research. *Frontiers in Behavioral Neuroscience*, *0*. https://doi.org/10.3389/fnbeh.2015.00104

Figueiredo, C. P., Barros-Aragão, F. G. Q., Neris, R. L. S., Frost, P. S., Soares, C., Souza, I. N. O., Zeidler, J. D., Zamberlan, D. C., de Sousa, V. L., Souza, A. S., Guimarães, A. L. A., Bellio, M., Marcondes de Souza, J., Alves-Leon, S. V., Neves, G. A., Paula-Neto, H. A., Castro, N. G., De

Felice, F. G., Assunção-Miranda, I., … Ferreira, S. T. (2019). Zika virus replicates in adult human brain tissue and impairs synapses and memory in mice. *Nature Communications*, *10*(1), 3890. https://doi.org/10.1038/s41467-019-11866-7

Forgas, J. P. (2001). *Feeling and thinking: The role of affect in social cognition*. Cambridge University Press.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fraser, D., & Duncan, I. J. (1998). *'Pleasures', 'pains' and animal welfare: Toward a natural history of affect*.

Fraser, D., Weary, D. M., Pajor, E. A., & Milligan, B. N. (1997a). *A scientific conception of animal welfare that reflects ethical concerns*.

Fraser, D., Weary, D., Pajor, E., & Milligan, B. (1997b). A Scientific Conception of Animal Welfare that Reflects Ethical Concerns. *Ethics and Animal Welfare Collection*. https://www.wellbeingintlstudiesrepository.org/ethawel/1

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Gamer, M., Lemon, J., & Singh <puspendra.pusp22@gmail.com>, I. F. P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. https://CRAN.R-project.org/package=irr

Gottlieb, D. H., & Capitanio, J. P. (2013). Latent Variables Affecting Behavioral Response to the Human Intruder Test in Infant Rhesus Macaques (Macaca mulatta). *American Journal of Primatology*, *75*(4), 314–323. https://doi.org/10.1002/ajp.22107

Griffin, D. R. (1998). From cognition to consciousness. *Animal Cognition*, *1*(1), 3–16.

Gross, J. J. (1998). The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*, *2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Guillory, S. A., & Bujarski, K. A. (2014). Exploring emotions using invasive methods: Review of 60 years of human intracranial electrophysiology. *Social Cognitive and Affective Neuroscience*, *9*(12), 1880–1889. https://doi.org/10.1093/scan/nsu002

Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in decision making. *Trends in Cognitive Sciences*, *18*(4), 194–202. https://doi.org/10.1016/j.tics.2014.01.003

Gur, R. E., & Gur, R. C. (2002). Gender differences in aging: Cognition, emotions, and neuroimaging studies. *Dialogues in Clinical Neuroscience*, *4*(2), 197–210.

Gygax, L. (2014). The A to Z of statistics for testing cognitive judgement bias. *Animal Behaviour*, *95*, 59–69. https://doi.org/10.1016/j.anbehav.2014.06.013

Harding, E. J., Paul, E. S., & Mendl, M. (2004). Cognitive bias and affective state. *Nature*, *427*(6972), 312–312. https://doi.org/10.1038/427312a

Hauser, M. D. (1991). Sources of Acoustic Variation in Rhesus Macaque (Macaca mulatta))Vocalizations. *Ethology, 89*(1), 29–46. https://doi.org/10.1111/j.1439-0310.1991.tb00291.x

Hinde, R. A. (1985). Was 'The expression of the emotions'a misleading phrase? *Animal Behaviour*, *33*(3), 985–992.

Hinde, R. A., & Rowell, T. E. (1962). Communication by postures and facial expressions in the rhesus monkey (Macaca mulatta). *Proceedings of the Zoological Society of London*, *138*(1), 1–21.

Horváth, M., Pichová, K., & Košťál, Ľ. (2016). The effects of housing conditions on judgement bias in Japanese quail. *Applied Animal Behaviour Science*, *185*, 121–130. https://doi.org/10.1016/j.applanim.2016.09.007

Howarth, E. R. I., Kemp, C., Thatcher, H. R., Szott, I. D., Farningham, D., Witham, C. L., Holmes, A., Semple, S., & Bethell, E. J. (2021). Developing and validating attention bias tools for assessing trait and state affect in animals: A worked example with Macaca mulatta. *Applied Animal Behaviour Science*, *234*, 105198. https://doi.org/10.1016/j.applanim.2020.105198

IBM Corp. (2021). *IBM SPSS Statistics for Windows* (28.0). https://www.ibm.com

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* Free press.

Jones, S., Neville, V., Higgs, L., Paul, E. S., Dayan, P., Robinson, E. S. J., & Mendl, M. (2018). Assessing animal affect: An automated and self-initiated judgement bias task based on natural investigative behaviour. *Scientific Reports*, *8*(1), 12400. https://doi.org/10.1038/s41598-018-30571-x

Jones, S., Paul, E. S., Dayan, P., Robinson, E. S. J., & Mendl, M. (2017). Pavlovian influences on learning differ between rats and mice in a counter-balanced Go/NoGo judgement bias task. *Behavioural Brain Research*, *331*, 214–224. https://doi.org/10.1016/j.bbr.2017.05.044

Kalin, N. H., & Shelton, S. E. (1989). Defensive behaviors in infant rhesus monkeys: Environmental cues and neurochemical regulation. *Science*, *243*(4899), 1718–1721.

Kalin, N. H., & Shelton, S. E. (1998). Ontogeny and stability of separation and threat-induced defensive behaviors in rhesus monkeys during the first year of life. *American Journal of Primatology*, *44*(2), 125–135. https://doi.org/10.1002/(SICI)1098-2345(1998)44:2<125::AID-AJP3>3.0.CO;2-Y

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (2003). *Facial expression of emotion.*

Kindt, M., & Van Den Hout, M. (2001). Selective attention and anxiety: A perspective on developmental issues and the causal status. *Journal of Psychopathology and Behavioral Assessment*, *23*(3), 193–202. https://doi.org/10.1023/A:1010921405496

Knutson, B., Burgdorf, J., & Panksepp, J. (2002). Ultrasonic vocalizations as indices of affective states in rats. *Psychological Bulletin*, *128*(6), 961.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, *42*(2), 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, *18*(1), 64–71.

Krasheninnikova, A., & Schneider, J. M. (2014). Testing problem-solving capacities: Differences between individual testing and social group setting. *Animal Cognition*, *17*(5), 1227–1232.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Lagisz, M., Zidar, J., Nakagawa, S., Neville, V., Sorato, E., Paul, E. S., Bateson, M., Mendl, M., & Løvlie, H. (2020). Optimism, pessimism and judgement bias in animals: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *118*, 3–17. https://doi.org/10.1016/j.neubiorev.2020.07.012

Landgraf, R., & Wigger, A. (2002). High vs Low Anxiety-Related Behavior Rats: An Animal Model of Extremes in Trait Anxiety. *Behavior Genetics*, *32*(5), 301–314. https://doi.org/10.1023/A:1020258104318

Lane, R. D., Nadel, L., Allen, J. J., & Kaszniak, A. W. (2000). *The study of emotion from the perspective of cognitive neuroscience*.

Larsen, R. J. (2000). Toward a Science of Mood Regulation. *Psychological Inquiry*, *11*(3), 129–141. https://doi.org/10.1207/S15327965PLI1103_01

Laviola, G., Adriani, W., Rea, M., Aloe, L., & Alleva, E. (2004). Social withdrawal, neophobia, and stereotyped behavior in developing rats exposed to neonatal asphyxia. *Psychopharmacology*, *175*(2), 196–205. https://doi.org/10.1007/s00213-004-1800-3

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46*(4), 352.

Lazarus, R. S. (1999). The cognition-emotion debate: A bit of history. *Handbook of Cognition and Emotion*, *5*(6), 3–19.

Lecorps, B., Kappel, S., Weary, D. M., & von Keyserlingk, M. A. G. (2018). Dairy calves' personality traits predict social proximity and response to an emotional challenge. *Scientific Reports*, *8*(1), 16350. https://doi.org/10.1038/s41598-018-34281-2

Lecorps, B., Nogues, E., von Keyserlingk, M. A., & Weary, D. M. (2020). Pessimistic dairy calves are more vulnerable to pain-induced anhedonia. *Plos One*, *15*(11), e0242100.

Lecorps, B., Weary, D. M., & von Keyserlingk, M. A. G. (2018). Pessimism and fearfulness in dairy calves. *Scientific Reports*, *8*, 1421. https://doi.org/10.1038/s41598-017-17214-3

Lecorps, B., Weary, D. M., & von Keyserlingk, M. A. G. (2021). Negative expectations and vulnerability to stressors in animals. *Neuroscience & Biobehavioral Reviews*, *130*, 240–251. https://doi.org/10.1016/j.neubiorev.2021.08.025

Ledoux, J. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon and Schuster.

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Lindquist, K. A., Siegel, E. H., Quigley, K. S., & Barrett, L. F. (2013). The Hundred-Year Emotion War: Are Emotions Natural Kinds or Psychological Constructions? Comment on. *Psychological Bulletin*, *139*(1), 255–263. https://doi.org/10.1037/a0029038

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(3), 121–143. https://doi.org/10.1017/S0140525X11000446

MacLeod, A. K., & Byrne, A. (1996). Anxiety, depression, and the anticipation of future positive and negative experiences. *Journal of Abnormal Psychology*, *105*(2), 286–289. https://doi.org/10.1037/0021-843X.105.2.286

MacLeod, C., & Mathews, A. (1988). Anxiety and the Allocation of Attention to Threat. *The Quarterly Journal of Experimental Psychology Section A*, *40*(4), 653–670. https://doi.org/10.1080/14640748808402292

Macphail, E. M. (1998). *The evolution of consciousness.* Oxford University Press.

Maestripieri, D., & Call, J. (1996). Mother-infant communication in primates. *Advances in the Study of Behavior*, *25*, 613–642.

Maestripieri, D., Jovanovic, T., & Gouzoules, H. (2000). Crying and infant abuse in rhesus monkeys. *Child Development*, *71*(2), 301–309.

Maestripieri, D., & Wallen, K. (1997). Affiliative and submissive communication in rhesus macaques. *Primates*, *38*(2), 127–138. https://doi.org/10.1007/BF02382003

Manteuffel, G., Puppe, B., & Schön, P. C. (2004). Vocalization of farm animals as a measure of welfare. *Applied Animal Behaviour Science*, *88*(1–2), 163–182.

Marchant, J. N., Mendl, M. T., Rudd, A. R., & Broom, D. M. (1995). The effect of agonistic interactions on the heart rate of group-housed sows. *Applied Animal Behaviour Science*, *46*(1), 49–56. https://doi.org/10.1016/0168-1591(95)00636-2

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), 73–81. https://doi.org/10.1037/1528-3542.6.1.73

Matheson, S. M., Asher, L., & Bateson, M. (2008). Larger, enriched cages are associated with ‘optimistic’ response biases in captive European starlings (Sturnus vulgaris). *Applied Animal Behaviour Science*, *109*(2), 374–383. https://doi.org/10.1016/j.applanim.2007.03.007

Mathews, A., & MacLeod, C. (1994a). Cognitive approaches to emotion and emotional disorders. *Annual Review of Psychology*, *45*, 25–50. https://doi.org/10.1146/annurev.ps.45.020194.000325

Mathews, A., & MacLeod, C. (1994b). Cognitive Approaches to Emotion and Emotional Disorders. *Annual Review of Psychology*, *45*(1), 25–50. https://doi.org/10.1146/annurev.ps.45.020194.000325

Mathews, A., & MacLeod, C. (2002). Induced processing biases have causal effects on anxiety. *Cognition & Emotion*, *16*(3), 331–354.

McCarty, R., Horwatt, K., & Konarska, M. (1988). Chronic stress and sympathetic-adrenal medullary responsiveness. *Social Science & Medicine (1982)*, *26*(3), 333–341. https://doi.org/10.1016/0277-9536(88)90398-x

Mendl, M. (2001). Assessing the welfare state. *Nature*, *410*(6824), 31–32. https://doi.org/10.1038/35065194

Mendl, M., Burman, O., Parker, R., & Paul, E. (2009). Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science - APPL ANIM BEHAV SCI*, *118*, 161–181. https://doi.org/10.1016/j.applanim.2009.02.023

Middlebrooks, P. G., & Sommer, M. A. (2012). Neuronal correlates of metacognition in primate frontal cortex. *Neuron*, *75*(3), 517–530.

Mineka, S., Watson, D., & Clark, L. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annu. Rev. Psychol*, *49*, 377–412. https://doi.org/10.1146/annurev.psych.49.1.377

Mogg, K., & Bradley, B. P. (1998). A cognitive-motivational analysis of anxiety. *Behaviour Research and Therapy*, *36*(9), 809–848. https://doi.org/10.1016/s0005-7967(98)00063-1

Morris, W. N. (1992). *A functional analysis of the role of mood in affective systems.*

Murphy, E., Kraak, L., van den Broek, J., Nordquist, R. E., & van der Staay, F. J. (2015). Decision-making under risk and ambiguity in low-birth-weight pigs. *Animal Cognition*, *18*(2), 561–572.

Murphy, E., Nordquist, R. E., & van der Staay, F. J. (2013). Responses of conventional pigs and Göttingen miniature pigs in an active choice judgement bias task. *Applied Animal Behaviour Science*, *148*(1–2), 64–76.

Neave, H. W., Daros, R. R., Costa, J. H. C., Keyserlingk, M. A. G. von, & Weary, D. M. (2013). Pain and Pessimism: Dairy Calves Exhibit Negative Judgement Bias following Hot-Iron Disbudding. *PLOS ONE*, *8*(12), e80556. https://doi.org/10.1371/journal.pone.0080556

Nelson, N. L., & Russell, J. A. (2013). Universality Revisited. *Emotion Review*, *5*(1), 8–15. https://doi.org/10.1177/1754073912457227

Neville, V., Nakagawa, S., Zidar, J., Paul, E. S., Lagisz, M., Bateson, M., Løvlie, H., & Mendl, M. (2020). Pharmacological manipulations of judgement bias: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *108*, 269–286. https://doi.org/10.1016/j.neubiorev.2019.11.008

Nygren, T. E., Isen, A. M., Taylor, P. J., & Dulin, J. (1996). The Influence of Positive Affect on the Decision Rule in Risk Situations: Focus on Outcome (and Especially Avoidance of Loss) Rather Than Probability. *Organizational Behavior and Human Decision Processes*, *66*(1), 59–72. https://doi.org/10.1006/obhd.1996.0038

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, *1*(1), 29–50.

Ortony, A., Clore, G. L., & Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.

Panksepp, J. (2003). At the interface of the affective, behavioral, and cognitive neurosciences: Decoding the emotional feelings of the brain. *Brain and Cognition*, *52*(1), 4–14.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80.

Panksepp, J. (2011). Cross-Species Affective Neuroscience Decoding of the Primal Affective Experiences of Humans and Related Animals. *PLOS ONE*, *6*(9), e21236. https://doi.org/10.1371/journal.pone.0021236

Panksepp, J., & Burgdorf, J. (2003). "Laughing" rats and the evolutionary antecedents of human joy? *Physiology & Behavior*, *79*(3), 533–547.

Parkinson, B., Totterdell, P., Briner, R. B., & Reynolds, S. A. (1996). *Changing Moods: The Psychology of Mood and Mood Regulation*. Longman. https://ueaeprints.uea.ac.uk/id/eprint/14584/

Parr, L. A. (2001). Cognitive and physiological markers of emotional awareness in chimpanzees (Pan troglodytes). *Animal Cognition*, *4*(3), 223–229. https://doi.org/10.1007/s100710100085

Parr, L. A., Waller, B. M., Burrows, A. M., Gothard, K. M., & Vick, S. J. (2010). MaqFACS: A Muscle-Based Facial Movement Coding System for the Rhesus Macaque. *American Journal of Physical Anthropology*, *143*(4), 625–630. https://doi.org/10.1002/ajpa.21401

Paul, E. S., Harding, E. J., & Mendl, M. (2005a). Measuring emotional processes in animals: The utility of a cognitive approach. *Neuroscience & Biobehavioral Reviews*, *29*(3), 469–491. https://doi.org/10.1016/j.neubiorev.2005.01.002

Paul, E. S., Harding, E. J., & Mendl, M. (2005b). Measuring emotional processes in animals: The utility of a cognitive approach. *Neuroscience & Biobehavioral Reviews*, *29*(3), 469–491. https://doi.org/10.1016/j.neubiorev.2005.01.002

Paul, E. S., Sher, S., Tamietto, M., Winkielman, P., & Mendl, M. T. (2020). Towards a comparative science of emotion: Affect and consciousness in humans and animals. *Neuroscience & Biobehavioral Reviews*, *108*, 749–770. https://doi.org/10.1016/j.neubiorev.2019.11.014

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., & others. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, *7*(3), e1001342.

Perrett, D. I., & Mistlin, A. J. (1990). Perception of facial characteristics by monkeys. *Comparative Perception, Vol. 2: Complex Signals.*, 187–215.

Petit, O., & Thierry, B. (1992). Affiliative function of the silent bared-teeth display in moor macaques (Macaca maurus): Further evidence for the particular status of sulawesi macaques. *International Journal of Primatology*, *13*(1), 97. https://doi.org/10.1007/BF02547729

Phillips, K. A., Bales, K. L., Capitanio, J. P., Conley, A., Czoty, P. W., ‘t Hart, B. A., Hopkins, W. D., Hu, S.-L., Miller, L. A., Nader, M. A., Nathanielsz, P. W., Rogers, J., Shively, C. A., & Voytko, M. L. (2014). Why Primate Models Matter. *American Journal of Primatology*, *76*(9), 801–827. https://doi.org/10.1002/ajp.22281

Pinheiro, J., & Bates, D. (2006). *Mixed-Effects Models in S and S-PLUS*. Springer Science & Business Media.

Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, *26*.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramos, A., & Mormède, P. (1997). Stress and emotionality: A multidimensional and genetic approach. *Neuroscience & Biobehavioral Reviews*, *22*(1), 33–57. https://doi.org/10.1016/S0149-7634(97)00001-8

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America*, *113*(6), 3390–3402.

Roelofs, S., Boleij, H., Nordquist, R. E., & van der Staay, F. J. (2016). Making Decisions under Ambiguity: Judgment Bias Tasks for Assessing Emotional State in Animals. *Frontiers in Behavioral Neuroscience*, *10*, 119. https://doi.org/dar

Rosati, A. G., & Santos, L. R. (2016). Spontaneous metacognition in rhesus monkeys. *Psychological Science*, *27*(9), 1181–1191.

Ruis, M. A. W., Te Brake, J. H. A., Engel, B., Ekkel, E. D., Buist, W. G., Blokhuis, H. J., & Koolhaas, J. M. (1997). The Circadian Rhythm of Salivary Cortisol in Growing Pigs: Effects of Age, Gender, and Stress. *Physiology & Behavior*, *62*(3), 623–630. https://doi.org/10.1016/S0031-9384(97)00177-7

Rushen, J. (1986). Some problems with the physiological concept of "stress." *Australian Veterinary Journal*, *63*(11), 359–361. https://doi.org/10.1111/j.1751-0813.1986.tb02896.x

Rushen, J. (1991). Problems associated with the interpretation of physiological data in the assessment of animal welfare. *Applied Animal Behaviour Science*, *28*(4), 381–386. https://doi.org/10.1016/0168-1591(91)90170-3

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (2005). Emotion in human consciousness is built on core affect. *Journal of Consciousness Studies*, *12*(8–10), 26–42.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037//0022-3514.76.5.805

Rygula, R., Papciak, J., & Popik, P. (2013). Trait Pessimism Predicts Vulnerability to Stress-Induced Anhedonia in Rats. *Neuropsychopharmacology*, *38*(11), 2188–2196. https://doi.org/10.1038/npp.2013.116

Sanger, M. E., Doyle, R. E., Hinch, G. N., & Lee, C. (2011). Sheep exhibit a positive judgement bias and stress-induced hyperthermia following shearing. *Applied Animal Behaviour Science*, *131*(3), 94–103. https://doi.org/10.1016/j.applanim.2011.02.001

Sato, N., & Nakamura, K. (2001). Detection of directed gaze in rhesus monkeys (Macaca mulatta). *Journal of Comparative Psychology*, *115*(2), 115–121. https://doi.org/10.1037/0735-7036.115.2.115

Scheier, M. F., & Carver, C. S. (1992). Effects of optimism on psychological and physical well-being: Theoretical overview and empirical update. *Cognitive Therapy and Research*, *16*(2), 201–228. https://doi.org/10.1007/BF01173489

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. *Approaches to Emotion*, *2293*(317), 31.

Scherer, K. R. (1999). *Appraisal theory*.

Schino, G., Scucchi, S., Maestripieri, D., & Turillazzi, P. G. (1988). Allogrooming as a tension-reduction mechanism: A behavioral approach. *American Journal of Primatology*, *16*(1), 43–50. https://doi.org/10.1002/ajp.1350160106

Schrader, L., & Ladewig, J. (1999). Temporal Differences in the Responses of the Pituitary Adrenocortical Axis, the Sympathoadrenomedullar Axis, Heart Rate, and Behaviour to a Daily

Repeated Stressor in Domestic Pigs. *Physiology & Behavior*, *66*(5), 775–783. https://doi.org/10.1016/S0031-9384(99)00015-3

Scollo, A., Gottardo, F., Contiero, B., & Edwards, S. A. (2014). Does stocking density modify affective state in pigs as assessed by cognitive bias, behavioural and physiological parameters? *Applied Animal Behaviour Science*, *153*, 26–35. https://doi.org/10.1016/j.applanim.2014.01.006

Seyfarth, R. M., & Cheney, D. L. (1997). Some general features of vocal development in nonhuman primates. In *Social influences on vocal development* (pp. 249–273). Cambridge University Press. https://doi.org/10.1017/CBO9780511758843.013

Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science*, *20*(6), 395–399. https://doi.org/10.1177/0963721411424739

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. https://doi.org/10.1016/j.cub.2011.10.030

Shepperd, J. A., Waters, E. A., Weinstein, N. D., & Klein, W. M. (2015). A primer on unrealistic optimism. *Current Directions in Psychological Science*, *24*(3), 232–237.

Shettleworth, S. J. (2009). *Cognition, Evolution, and Behavior*. Oxford University Press.

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion Fingerprints or Emotion Populations? A Meta-Analytic Investigation of Autonomic Features of Emotion Categories. *Psychological Bulletin*, *144*(4), 343–393. https://doi.org/10.1037/bul0000128

Silk, J. B., Cheney, D. L., & Seyfarth, R. M. (1996). The form and function of post-conflict interactions between female baboons. *Animal Behaviour*, *52*(2), 259–268. https://doi.org/10.1006/anbe.1996.0171

Starling, M. J., Branson, N., Cody, D., Starling, T. R., & McGreevy, P. D. (2014). Canine Sense and Sensibility: Tipping Points and Response Latency Variability as an Optimism Index in a Canine Judgement Bias Assessment. *PLOS ONE*, *9*(9), e107794. https://doi.org/10.1371/journal.pone.0107794

Stephenson, E., & Haskell, M. J. (2020). The Use of a “Go/Go” Cognitive Bias Task and Response to a Novel Object to Assess the Effect of Housing Enrichment in Sheep (Ovis aries). *Journal of Applied Animal Welfare Science*, *0*(0), 1–13. https://doi.org/10.1080/10888705.2020.1824786

Strelau, J. (2001). The concept and status of trait in research on temperament. *European Journal of Personality*, *15*(4), 311–325. https://doi.org/10.1002/per.412

Süer, C., Dolu, N., Özesmi, Ç., Şahin, Ö., & Ülgen, A. (1998). The relation between skin conductance level and plus-maze behavior in male mice. *Physiology & Behavior*, *64*(4), 573–576. https://doi.org/10.1016/S0031-9384(98)00075-4

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics, 5th ed.* (pp. xxvii, 980). Allyn & Bacon/Pearson Education.

Takeshita, F., & Sato, N. (2016). Adaptive Sex-Specific Cognitive Bias in Predation Behaviours of Japanese Pygmy Squid. *Ethology*, *122*(3), 236–244.

Templer, V. L., & Hampton, R. R. (2012). Rhesus monkeys (Macaca mulatta) show robust evidence for memory awareness across multiple generalization tests. *Animal Cognition*, *15*(3), 409–419.

Toates, F. (1995). *Stress—Conceptual and Biological Aspects*. Wiley. http://oro.open.ac.uk/66406/

Tomaszycki, M. L., Davis, J. E., Gouzoules, H., & Wallen, K. (2001). Sex differences in infant rhesus macaque separation–rejection vocalizations and effects of prenatal androgens. *Hormones and Behavior*, *39*(4), 267–276.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405. https://doi.org/10.1177/1754073911410747

Tversky, A., & Kahneman, D. (1979). An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Verjat, A., Devienne, P., Rödel, H. G., & Féron, C. (2021). More exploratory house mice judge an ambiguous situation more negatively. *Animal Cognition*, *24*(1), 53–64. https://doi.org/10.1007/s10071-020-01414-y

Vick, S.-J., Waller, B. M., Parr, L. A., Smith Pasqualini, M. C., & Bard, K. A. (2007). A Cross-species Comparison of Facial Morphology and Movement in Humans and Chimpanzees Using the Facial Action Coding System (FACS). *Journal of Nonverbal Behavior*, *31*(1), 1–20. https://doi.org/10.1007/s10919-006-0017-z

von Borell, E., & Ladewig, J. (1992). Relationship between behaviour and adrenocortical response pattern in domestic pigs. *Applied Animal Behaviour Science*, *34*(3), 195–206. https://doi.org/10.1016/S0168-1591(05)80115-7

Waller, B. M., Lembeck, M., Kuchenbuch, P., Burrows, A. M., & Liebal, K. (2012). GibbonFACS: a muscle-based facial movement coding system for hylobatids. *International Journal of Primatology*, *33*(4), 809–821.

Waller, B. M., Peirce, K., Caeiro, C. C., Scheider, L., Burrows, A. M., McCune, S., & Kaminski, J. (2013). Paedomorphic Facial Expressions Give Dogs a Selective Advantage. *PLOS ONE*, *8*(12), e82686. https://doi.org/10.1371/journal.pone.0082686

Wathan, J., Burrows, A. M., Waller, B. M., & McComb, K. (2015). EquiFACS: The Equine Facial Action Coding System. *PLOS ONE*, *10*(8), e0131738. https://doi.org/10.1371/journal.pone.0131738

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Watts, J. M., & Stookey, J. M. (2000). Vocal behaviour in cattle: The animal’s commentary on its biological processes and welfare. *Applied Animal Behaviour Science*, *67*(1–2), 15–33.

Weafer, J., & de Wit, H. (2014). Sex differences in impulsive action and impulsive choice. *Addictive Behaviors*, *39*(11), 1573–1579. https://doi.org/10.1016/j.addbeh.2013.10.033

Weary, D. M., & Fraser, D. (1997). Vocal response of piglets to weaning: Effect of piglet age. *Applied Animal Behaviour Science*, *54*(2–3), 153–160.

Whittaker, A. L., & Barker, T. H. (2020). A consideration of the role of biology and test design as confounding factors in judgement bias tests. *Applied Animal Behaviour Science*, *232*, 105126. https://doi.org/10.1016/j.applanim.2020.105126

Wright, W. F., & Bower, G. H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes*, *52*(2), 276–291. https://doi.org/10.1016/0749-5978(92)90039-A

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151.

Zajonc, R. B. (2000). *Feeling and thinking: Closing the debate over the independence of affect.*

Zidar, J., Campderrich, I., Jansson, E., Wichman, A., Winberg, S., Keeling, L., & Løvlie, H. (2018). Environmental complexity buffers against stress-induced negative judgement bias in female chickens. *Scientific Reports*, *8*(1), 5404. https://doi.org/10.1038/s41598-018-23545-6

Section 7: Appendix

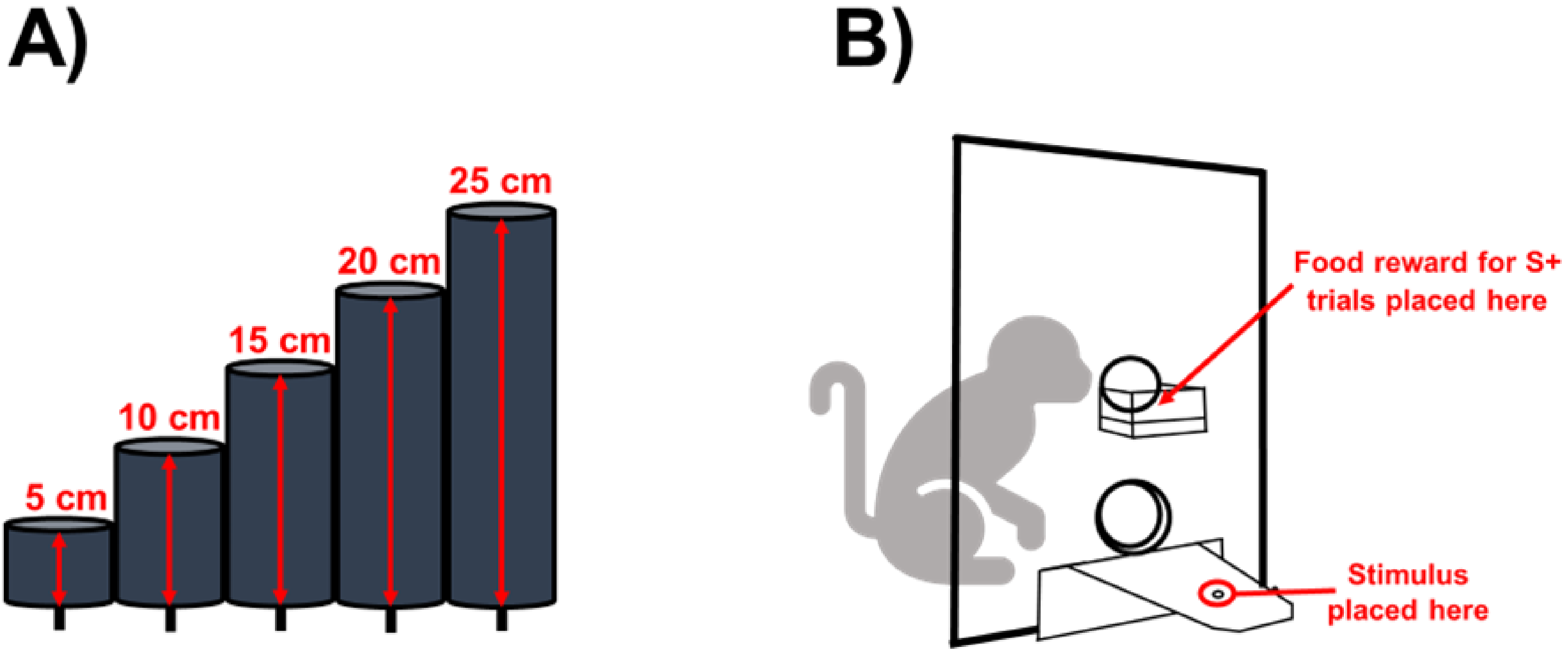


Figure 1. A) The five high-density polyethylene cylinders cut into various lengths used during the judgment bias task. The two outer stimuli (5cm and 25cm) were used for training of the task and on control trials during testing. The inner three ambiguous stimuli (10cm, 15cm, 20cm) were used on experimental trials during judgement bias testing. **B)** A plexiglass apparatus that was secured to the front of a cage during the training and testing of the judgement bias task. The apparatus had two holes: the lower hole allowed the monkeys to touch the stimulus placed on to the metal platform attached to the apparatus and the upper hole led to a cavity that allowed for retrieval of the food reward during S+ trials.

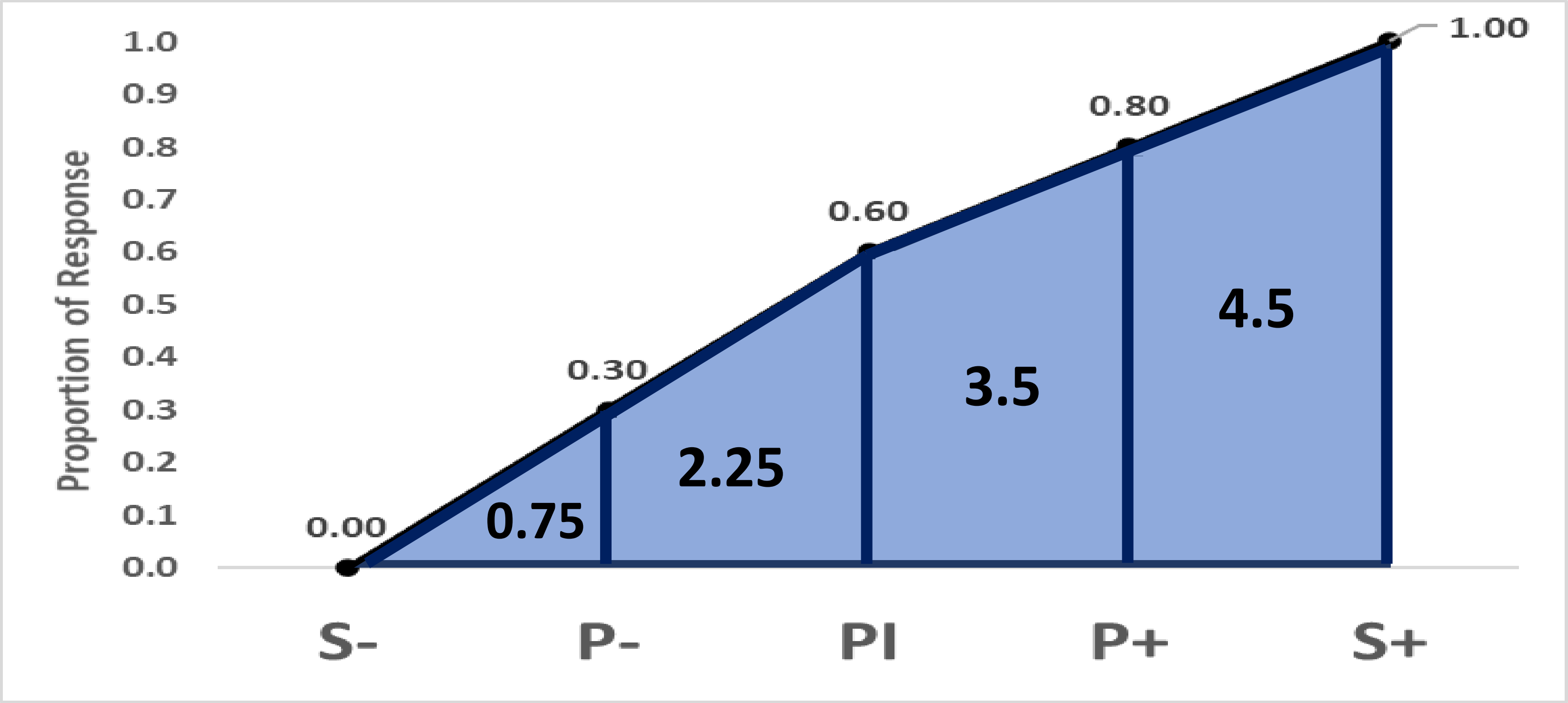
Table 1. Table showing the order of presentation of control signal stimuli (S+ and S-) and ambiguous probe stimuli (P+, PI, P-) during the three days of testing. Adults had 20 trials of stimuli presentations over the three days of baseline testing. Juveniles had 34 trials of stimuli presentations for both testing contexts (baseline and post threat induction). While the first and last trials were always rewarded S+ presentations, the presentation of stimuli was pseudorandomized and counterbalanced over the three days of testing.

Testing Day 1		Testing Day 2		Testing Day 3	
Adult	**Juvenile**	**Adult**	**Juvenile**	**Adult**	**Juvenile**
1. P-	1. S+	1. P+	1. S+	1. P+	1. S+
2. PI	2. S-	2. PI	2. S-	2. PI	2. S-
3. S+	3. P-	3. P-	3. P+	3. P-	3. PI
4. S-	4. S-	4. PI	4. S-	4. S-	4. S-
5. S+	5. S+	5. S-	5. S+	5. PI	5. S+
6. S-	6. PI	6. S+	6. P-	6. S+	6. P+
7. PI	7. S-	7. P-	7. S-	7. P+	7. S-
8. P+	8. PI	8. P-	8. P+	8. PI	8. P-
9. P+	9. S+	9. S-	9. S+	9. S-	9. S+
10. S+	10. P+	10. P+	10. PI	10. S+	10. P-
11. PI	11. S+	11. S+	11. S+	11. P-	11. S+
12. PI	12. S-	12. PI	12. S-	12. P+	12. S-
13. P+	13. S+	13. S-	13. S+	13. P-	13. S+
14. P-	14. P+	14. S+	14. P-	14. P+	14. PI
15. P+	15. S-	15. PI	15. S-	15. PI	15. S-
16. P-	16. PI	16. S+	16. P-	16. P-	16. P+
17. S-	17. S-	17. P+	17. S-	17. S+	17. S-
18. S-	18. PI	18. P+	18. P+	18. S+	18. P-
19. S+	19. S+	19. P-	19. S+	19. S-	19. S+
20. P-	20. S+	20. S-	20. S+	20. S-	20. S+
	21. P+		21. PI		21. P-
	22. S-		22. S-		22. S-
	23. S+		23. S+		23. S+
	24. P-		24. PI		24. P+
	25. S+		25. S+		25. S+
	26. P+		26. P-		26. PI
	27. S-		27. S-		27. S-
	28. S+		28. S+		28. S+
	29. S-		29. S-		29. S-
	30. P-		30. P+		30. PI
	31. S-		31. S-		31. S-
	32. P-		32. PI		32. P+
	33. S+		33. S+		33. S+
	34. S+		34. S+		34. S+

Table 2. Ethogram of affective behaviors displayed by the juvenile captive rhesus macaques (*Macaca mulatta)* during the Human Intruder Test. Frequencies of these affective behaviors were summed to represent a measure of affective reactivity.

Behavior	Description
Lip smack	Rapid lip movement usually with pursed lips and accompanied by a smacking sound
Barred Teeth	Exaggerated grin with teeth showing
Freeze	Rigid, fixed body position
Threat	At least two of the following: open mouth stare, head-bob, ear slaps, bark vocalizations, or lunges
Coo	Clear, soft vocalization that is moderate in pitch and intensity
Grunt	Deep, muffled low intensity vocalization; sometimes in conjunction with lip smacks
Bark	Low-pitched, abrupt, guttural vocalization
Scream	Intense, extremely high-pitched vocalization
Cage Shake	Grasping of cage parts and shaking
Tooth Grind	Repetitive, audible rubbing of upper and lower teeth
Yawn	Open mouth, exposing teeth
Self-Scratch	Individual scratches own body
Stereotypy	Expression of abnormal repetitive motor patterns

$$\text{JBI} = 5\left(\frac{x_{prop\,S-} + 2(x_{prop\,P-} + x_{prop\,PI} + x_{prop\,P+}) + x_{prop\,S+}}{2}\right)$$



$$\text{JBI} = 5\left(\frac{0 + 2(0.3 + 0.6 + 0.8) + 1}{2}\right) = 5\left(\frac{4.4}{2}\right) = 11$$

Figure 2. Visualization of the use of the trapezoidal rule for transformation of proportion data from judgement bias testing into a single judgement bias index (JBI). Formula for the JBI is included in figure, along with example proportion data demonstrating the calculations

Table 3. Orthogonally rotated principal component loadings of behavioral data collected from the Human Intruder test for each day of testing in the sample of juvenile monkeys. Only component loadings ≥ 0.45 are shown in the table. Four components of affective behaviors were generated that explained 62.9 percent of total variance in the original data.

Rotated Component Matrix[a]				
	Components			
	1	**2**	**3**	**4**
	“Behavioral Tension”	**“Threat Display”**	**“Vocalizations”**	**“Appeasement Behaviors”**
Self-Scratch	0.878			
Yawn	0.857			
Tooth Grind	0.629			
Threat		0.771		
Stereotypy		-0.751		
Cage Shake		0.604		-0.477
Coo			0.824	
Freeze			-0.769	
Grunt			0.467	
Lip smack				0.852
Barred Teeth				0.497

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

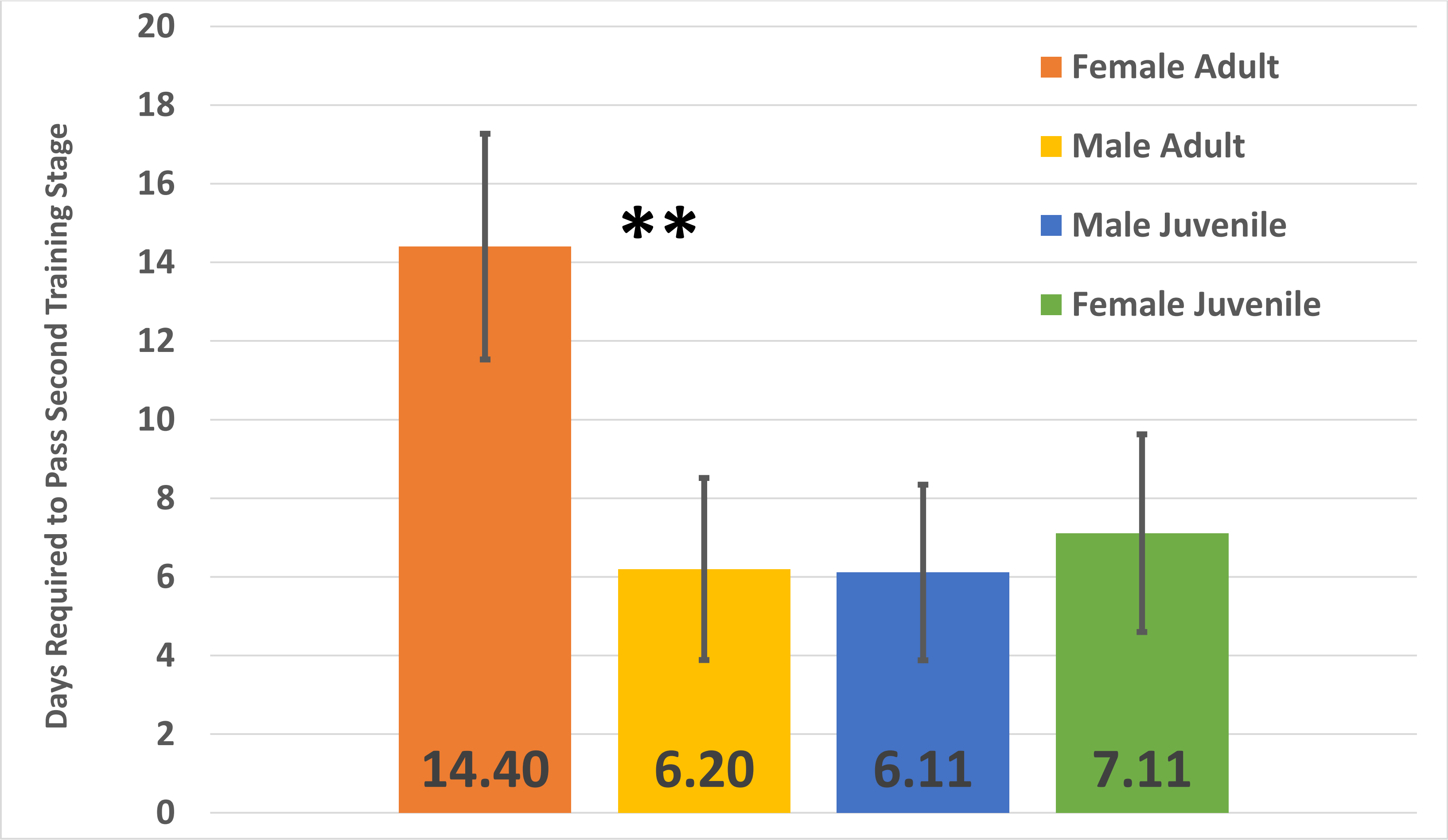


Figure 3. Interaction between age and sex on the mean number of training days (± SD) required to pass the second stage of training, where individuals learned to discriminate between the S+ and S- stimuli. A Kruskal-Wallis test showed that adult females (n = 5) required significantly longer to pass the second stage of training compared to the adult males (n = 5), juvenile males (n = 9), and juvenile females (n = 9) [*H* (3) = 11.37, *p* = 0.009]. ** indicates a p-value less than 0.01.

Table 4. Estimated marginal mean probability to respond to stimuli during baseline judgement bias testing generated by the generalized linear mixed-effects model. Results are averaged over the levels of age, sex, reinforced stimuli length, and testing day. Pairwise comparisons of adjacent stimuli are performed on the log odds ratio scale. P-value adjustments are performed using the Holm-Bonferroni method for 10 comparisons.

Stimulus	Probability to Respond	Standard Error	Pairwise Comparison	Odds Ratio	z-score	p-value
S-	0.033	0.011	S- / P-	0.153	-5.77	$p<0.0001$
P-	0.182	0.031	P- / PI	0.261	-7.06	$p<0.0001$
PI	0.461	0.047	PI / P+	0.187	-8.48	$p<0.0001$
P+	0.821	0.031	P+ / S+	0.103	-7.70	$p<0.0001$
S+	0.978	0.006	S+ / S-	1304.04	17.75	$p<0.0001$

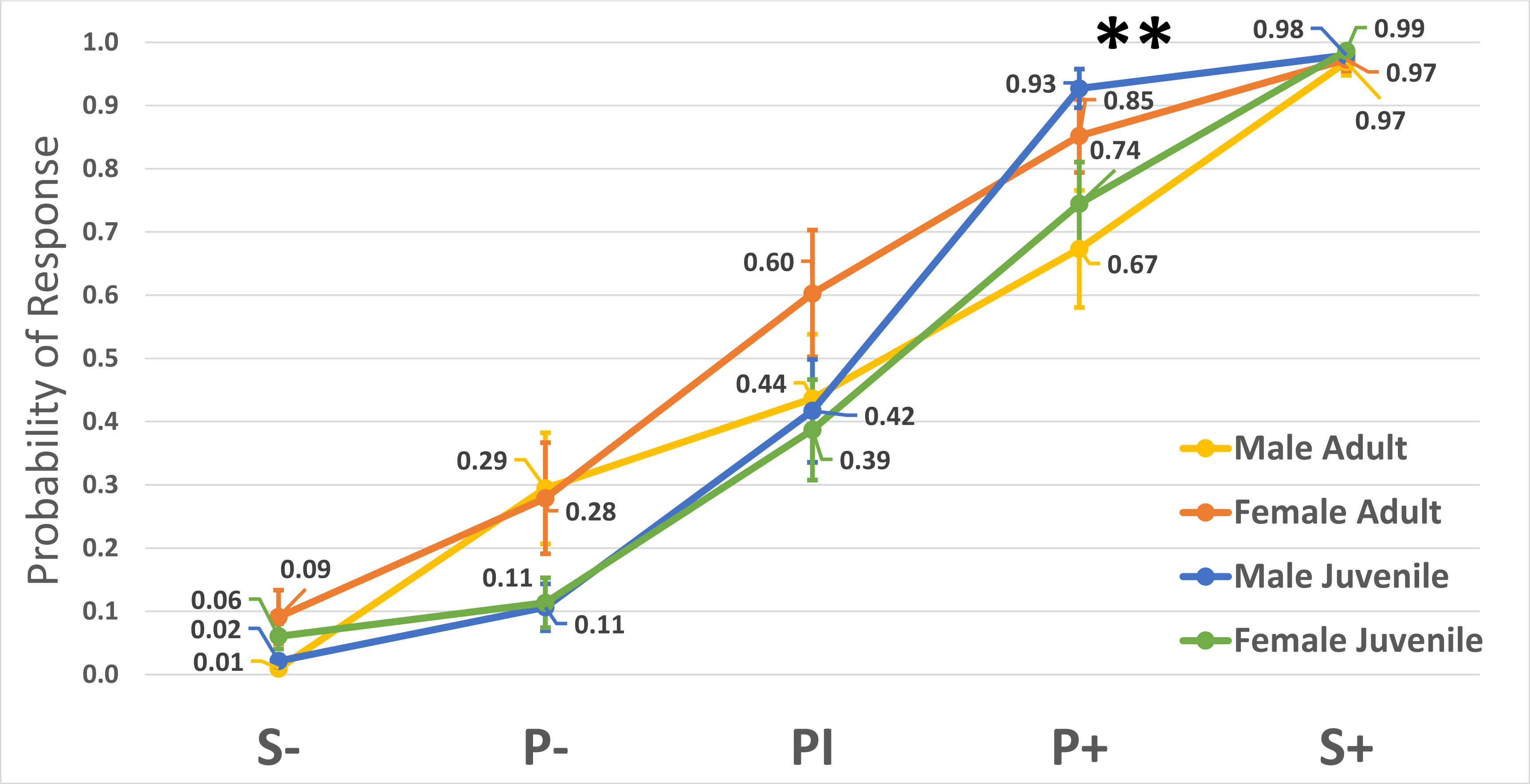


Figure 4. Interaction between age and sex in influencing the probability to respond to stimuli during baseline judgement bias testing. Probabilities are generated from the estimated marginal means of the generalized linear mixed-effects model ± SEM. Probabilities are averaged over the levels of reinforced stimuli length and testing day. A Type III Wald chi-squared test revealed a significant interaction between age and sex in influencing responses to the P+ stimulus (χ^2 (2) = 10.01, p = 0.007). ** indicates p-value less than 0.01.

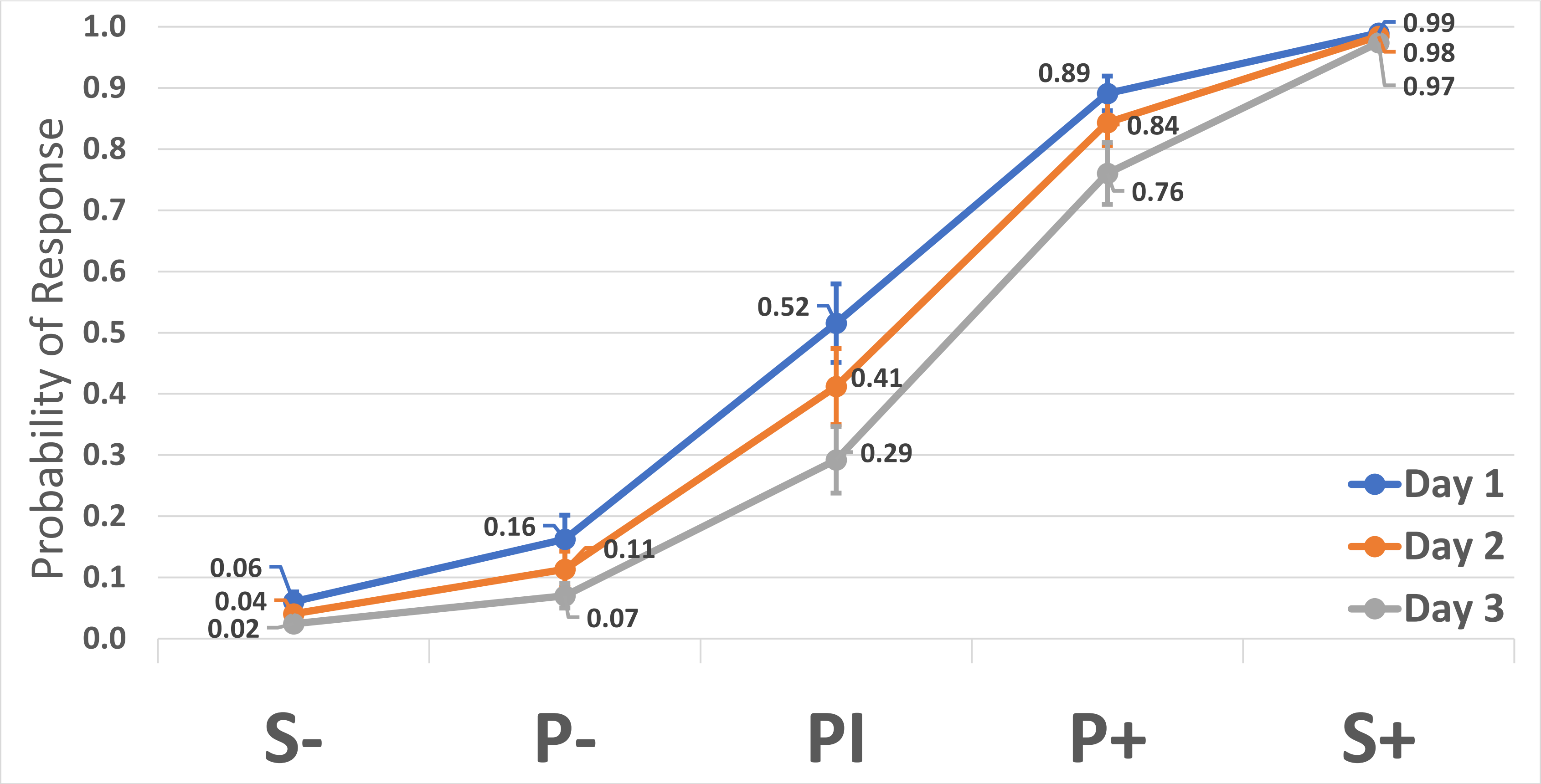


Figure 5. Influence of testing day on the probability to respond to stimuli during baseline judgement bias testing in the sample of juvenile monkeys. Probabilities are generated from the estimated marginal means of the generalized linear mixed-effects model ± SEM. Probabilities are averaged over the levels of reinforced stimuli length and sex. A Type II Wald chi-squared test revealed a significant effect of testing day on overall response during baseline testing (χ^2 (2) = 19.84, p<0.001).

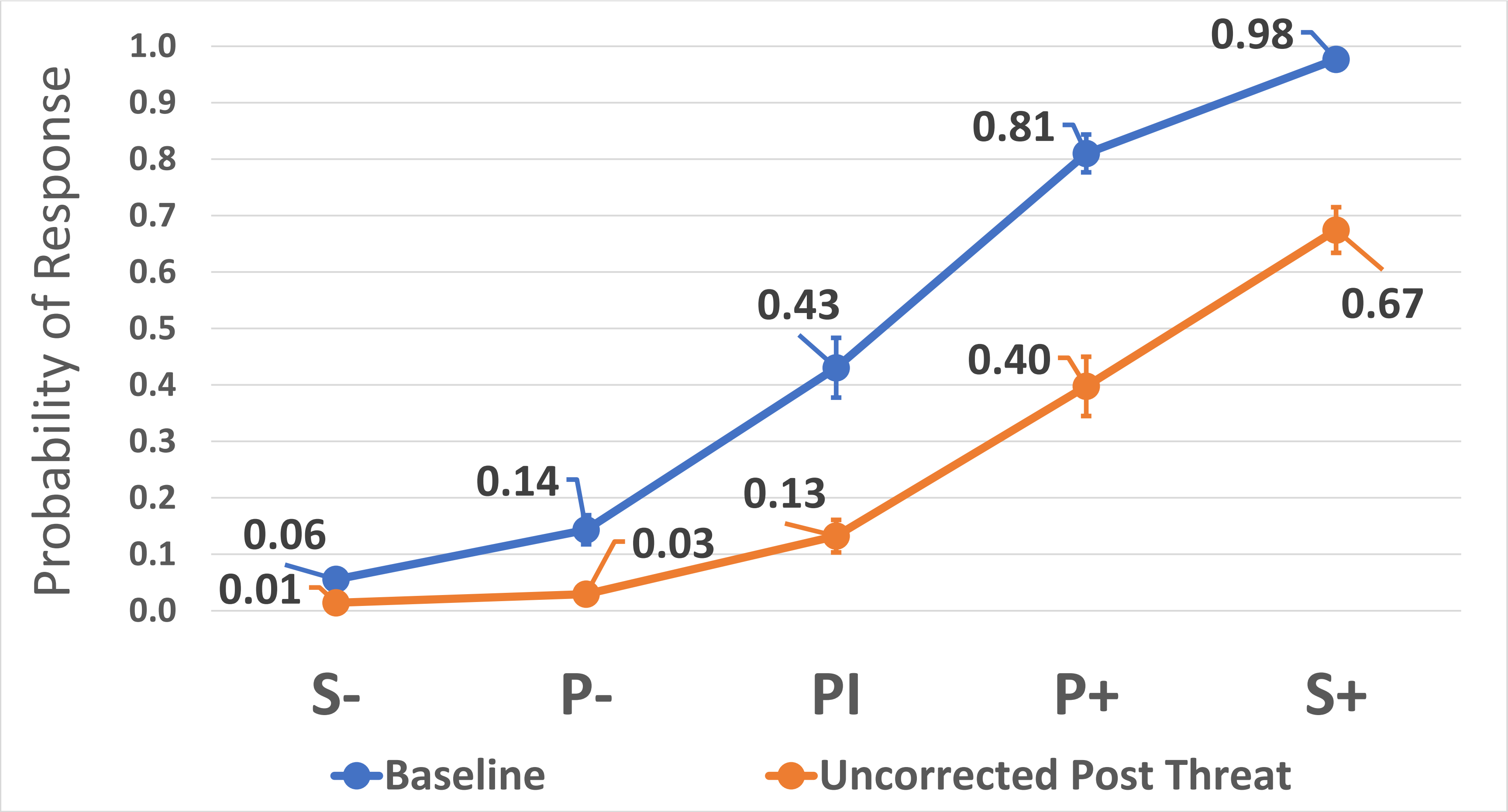


Figure 6. Effect of threat induction on the probability to respond to stimuli compared to baseline judgement bias testing in the sample of juvenile monkeys. Data from all trials included regardless of whether the 75% correct response criterion was met. Probabilities are generated from the estimated marginal means of the generalized linear mixed-effects model ± SEM. Probabilities are averaged over the levels of testing day, reinforced stimuli length, and sex. A Type III Wald chi-squared test revealed a significant effect of threat induction on a reduction in overall response to all stimuli (χ^2 (1) = 8.68, $p = 0.003$).

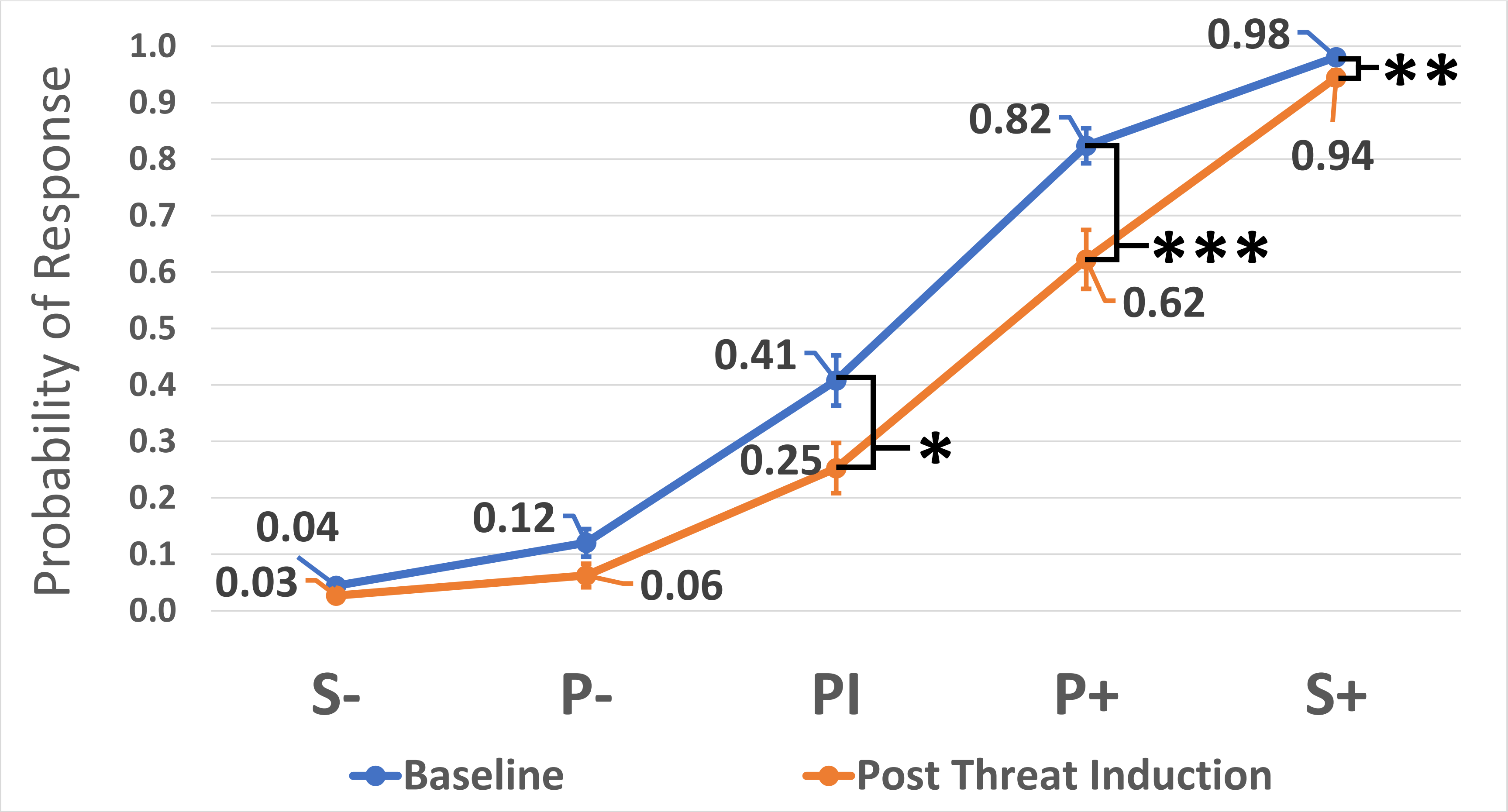


Figure 7. Influence of threat induction on the probability to respond to stimuli compared to baseline judgement bias testing for juveniles who reached the 75% correct response criterion. Probabilities are generated from the estimated marginal means of the generalized linear mixed-effects model ± SEM. Probabilities are averaged over the levels of testing day and reinforced stimuli length. A Type II Wald chi-squared test revealed a significant effect of threat induction on a reduction in overall response to all stimuli (χ^2 (1) = 35.58, p<0.001). Pairwise comparisons demonstrated that threat induction led to a significant reduction in likelihood to the PI (z = 2.837, p = 0.0228), P+ (z = 3.970, p = 0.0007), and S+ (z = 3.295, p = 0.0069) stimuli during judgement bias testing. * denotes a p-value less than 0.05; ** denotes a p-value less than 0.01; *** denotes a p-value less than 0.001.

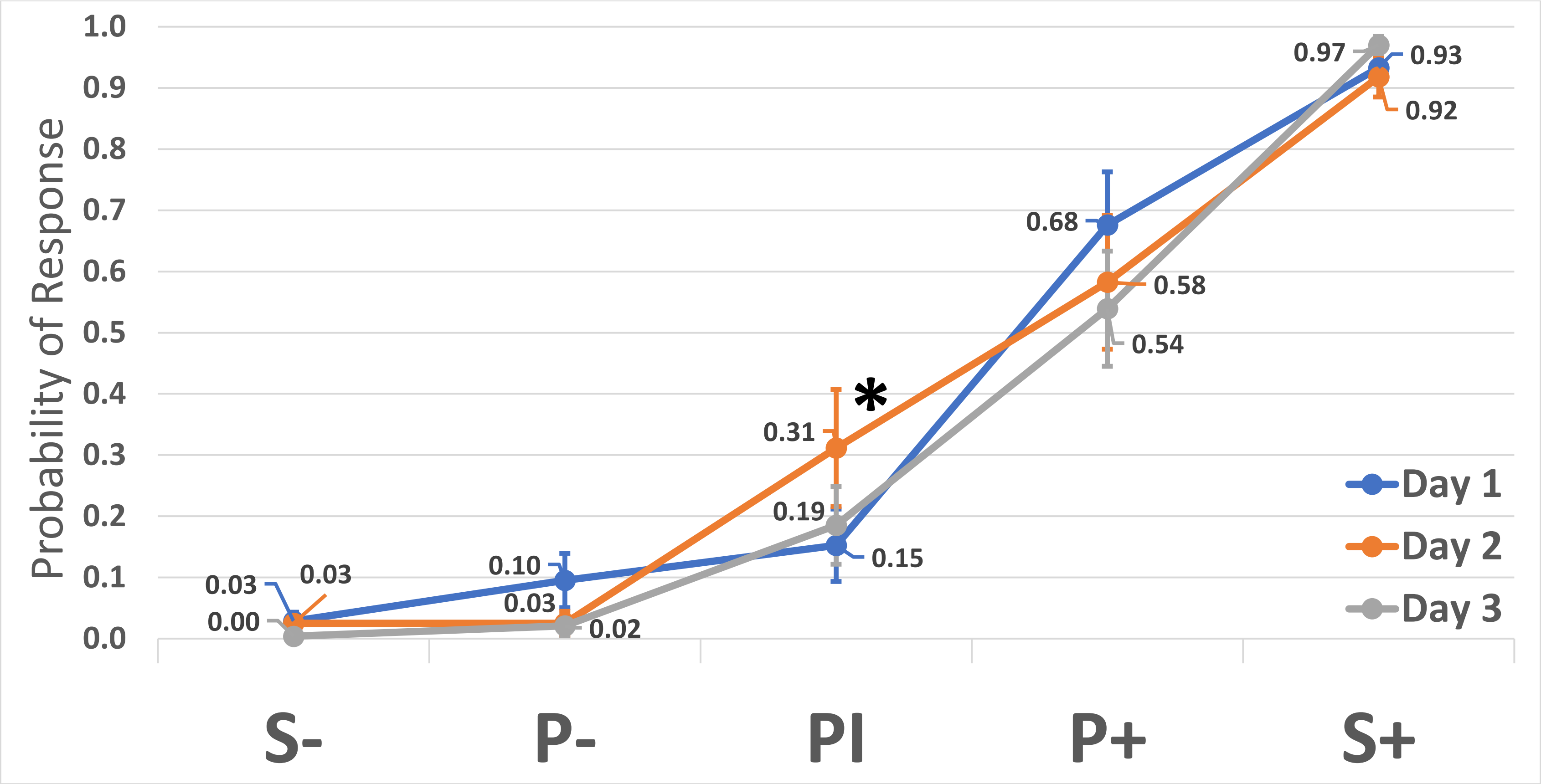


Figure 8. Effect of testing day on the probability to respond to stimuli during post threat induction judgement bias testing in juvenile monkeys who reached the 75% correct response criterion. Probabilities are generated from the estimated marginal means of the generalized linear mixed-effects model ± SEM. Probabilities are averaged over the levels of reinforced stimuli length. A Type III Wald chi-squared test showed no overall effect of testing day on responses during post threat induction testing (χ^2 (2) = 3.35, *p* = 0.19). However, there was a significant interaction between testing day and responses to specific stimuli (χ^2 (8) = 15.87, *p* = 0.044), mainly a higher response to the PI stimulus on Day 2 (*z* = 2.332, *p* = 0.02). * indicates a p-value less than 0.05.

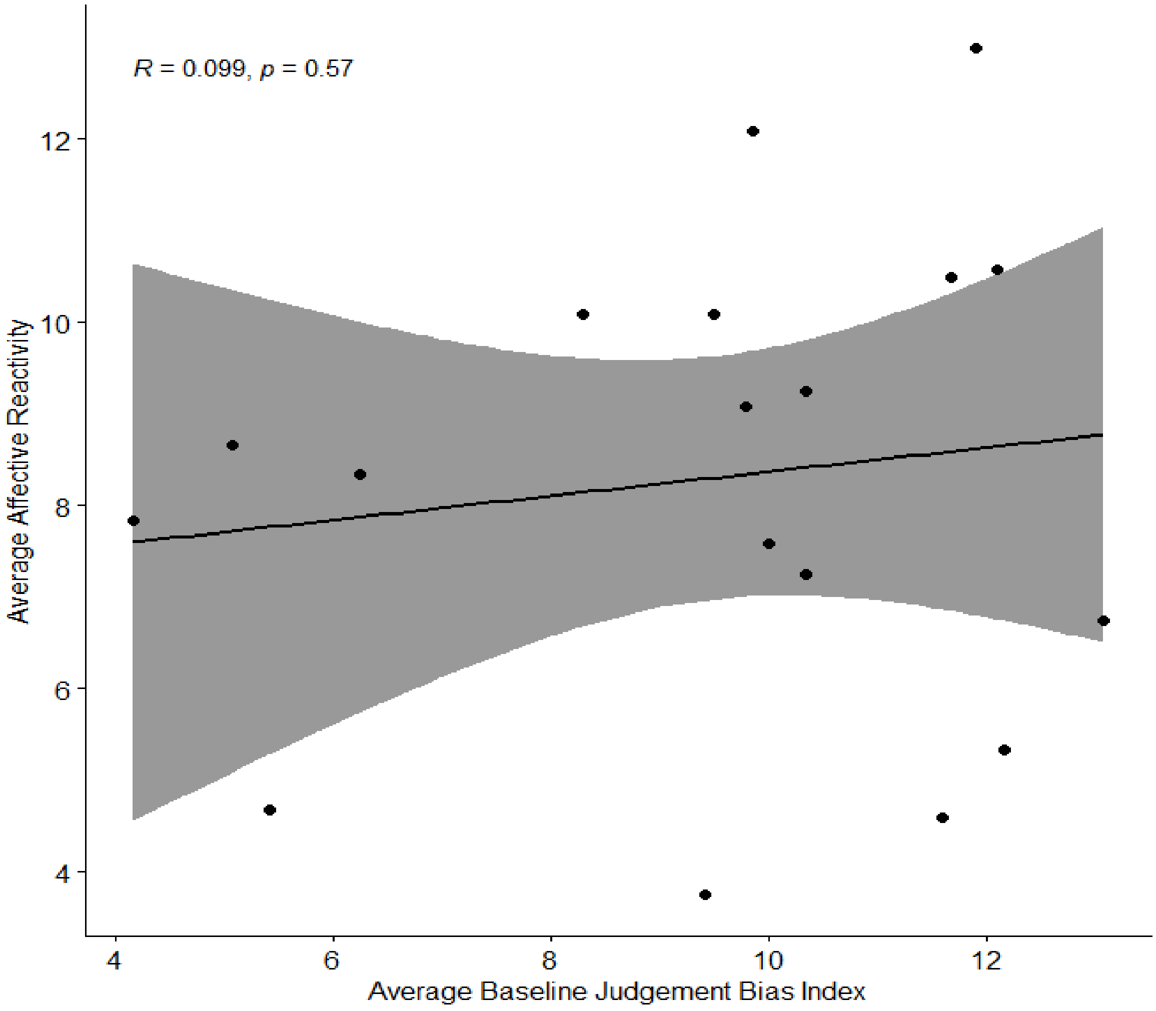


Figure 9. Relationship between an individual's average baseline judgement bias index (JBI) and average overall affective reactivity during the Human Intruder test for the sample of eighteen juvenile monkeys. A Kendall rank correlation test showed no significant relationship between these variables (r_{Tb} = 0.099, *p* = 0.57), with baseline trait judgement biases not predicting an individual's overall reactivity to threat induction and close proximity of an unfamiliar human. Each data point represents a single individual; the confidence interval of the regression line is represented by gray shading.

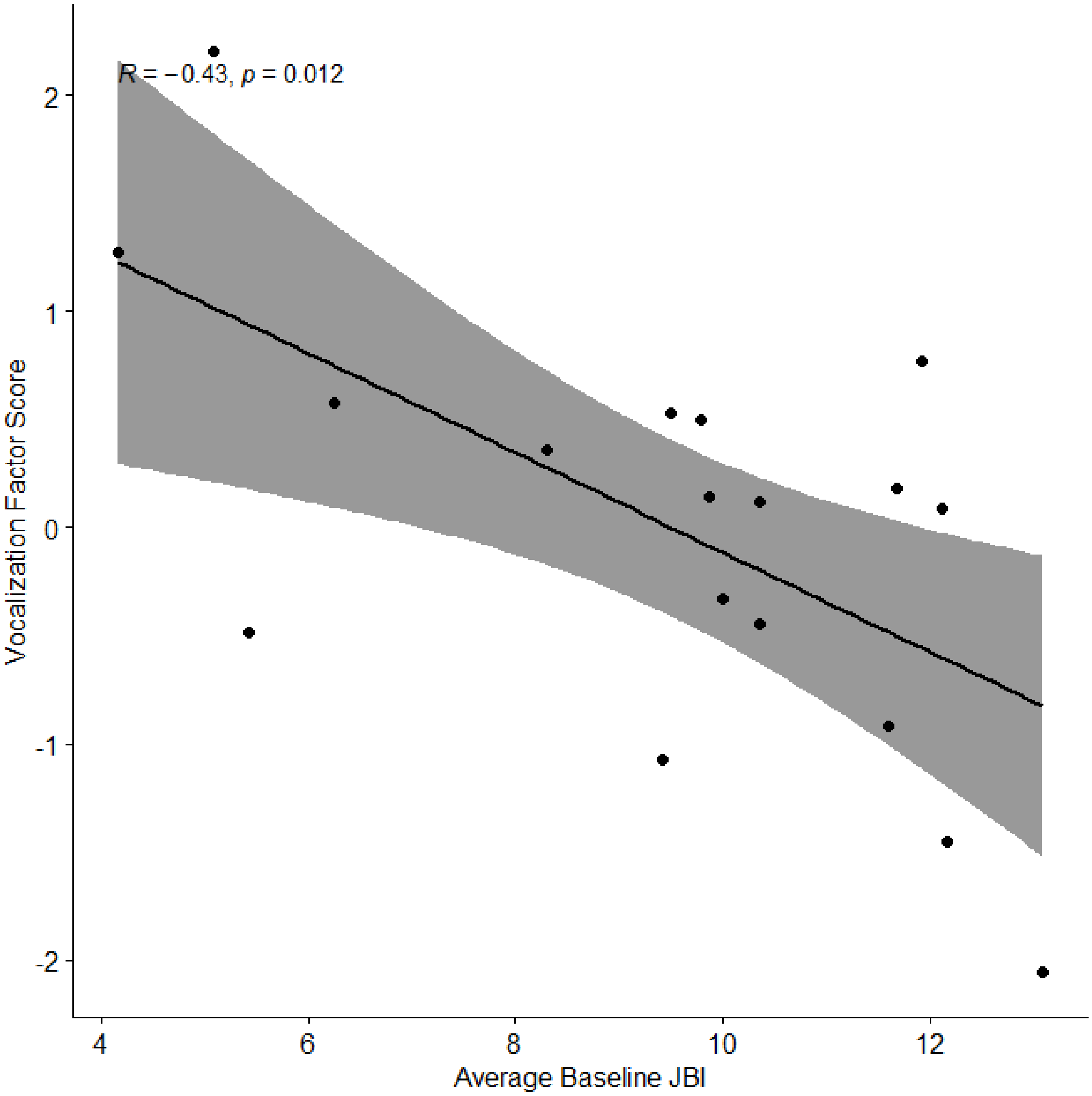


Figure 10. Relationship between an individual’s average baseline judgement bias index (JBI) and average “Vocalization” factor score during the Human Intruder test for the sample of eighteen juvenile monkeys. A Kendall rank correlation test showed that individuals with consistently lower responses during baseline judgement bias testing demonstrated a higher number of coo and grunt vocalizations during Human Intruder testing ($r_{Tb} = -0.43$, $p = 0.012$). Each data point represents a single individual; the confidence interval of the regression line is represented by gray shading.

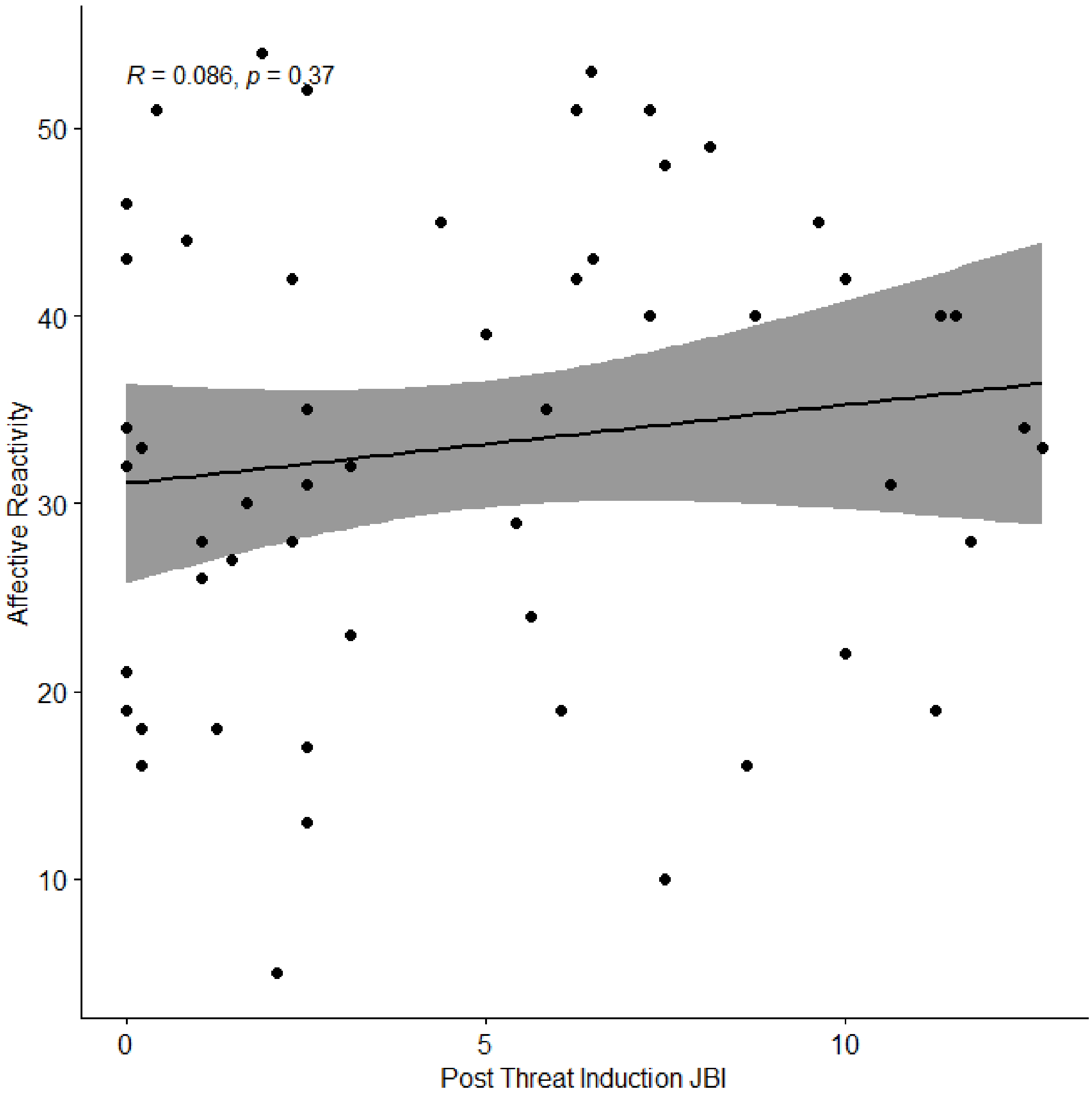


Figure 11. Relationship between an individual's affective reactivity during the Human Intruder test and an individual's judgement bias index (JBI) during subsequent post threat induction judgement bias testing for a single day of testing. A Kendall rank correlation test showed no significant relationship between an individual's reactivity during the Human Intruder test and changes in JBI on subsequent judgement bias testing (r_{Tb} = 0.086, *p* = 0.37). Individuals with higher reactivity to threat induction did not display more negative judgement biases. Each data point represents a single individual, with each individual having 3 data points corresponding to the 3 days of testing; the confidence interval of the regression line is represented by gray shading.