The Influence of Acute Stress on the Perception of Robot Emotional Body Language: Implications for Robot Design in Healthcare and Other High-Risk Domains

A Dissertation
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Zane Kali Thimmesch-Gill

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Kathleen A. Harder, Wilma Koutstaal

July 2017

Copyright © 2017 Zane Kali Thimmesch-Gill

Acknowledgments

I would like to thank my adviser Prof. Kathleen A. Harder and co-adviser Prof. Wilma Koutstaal for supporting me in this research. They made this dissertation possible through their generous dedication of time, thoughtfulness, and physical and financial resources. I would also like to thank Prof. Maria Gini and Marie Manner for support with the robot Nao and Prof. Victoria Interrante, Jung Nam, Peng Liu, and Zhihang Deng for support with virtual reality. Research was supported by an Interdisciplinary Doctoral Fellowship (ZTG) and a McKnight Presidential Fellowship (WK).

I would also like to thank my amazing partner Ani Goodenberger who is an excellent collaborator and teaches me more every day about designing for humans. Thank you also to my mom, who was the first person in her entire family to attend college – the fact that you stuck with it while experiencing intense sexism has been an inspiration and source of strength as I have made my way through school – almost always as the only trans person in the school, program, class, or group. Thank you also to my dad and sister, who are always there to cheer me on.

Abstract

In coming years, emotionally expressive social robots will permeate many facets of our lives. Yet, although researchers have explored robot design parameters that may facilitate human-robot interaction, remarkably little attention has been paid to the human perceptual and other psychological factors that may impact human ability to engage with robots. In high-risk settings, such as healthcare—where the use of robots is expected to increase markedly—it is paramount to understand the influence of a patient's stress level, temperament, and attitudes towards robots as negative interactions could harm a patient's experience and hinder recovery. Using a novel between-subject paradigm, we investigated how the experimental induction of acute physiological and cognitive stress versus low stress influences perception of normed robot emotional body language as conveyed by a physically-present versus virtual reality generated robot.

Following high or low stress induction, participants were asked to rate the valence (negative/unhappy to positive/happy) and level of arousal (calm/relaxed to animated/excited) conveyed by poses in five emotional categories: negative valence-high arousal, negative valence-low arousal, neutral, positive valence-low arousal, positive valence-high arousal. Poses from the categories were randomly intermixed and each pose was presented two or three times. Ratings were then correlated with temperament (as assessed by the Adult Temperament Questionnaire), attitudes towards and experience with robots (a new questionnaire that included measures from the Godspeed Scales and Negative Attitudes about Robots Survey), and chronic stress.

The acute stress induction especially influenced the evaluation of high arousal poses – both negative and positive – with both valence and arousal rated lower under high than low stress. Repeated presentation impacted perception of low arousal (negative and positive) and neutral poses, with increases in perceived valence and arousal for later

presentations. There were also effects of robot type specifically for positively-valenced emotions, such that these poses were rated as more positive for the physically-present than virtually-instantiated robot. Temperament was found to relate to emotional robot body language. Trait positive affect was associated with higher valence ratings for positive and neutral poses. Trait negative affect was correlated with higher arousal ratings for negative valence-low arousal poses. Subcategories within the robot attitudes questionnaire were correlated with emotional robot poses and temperament.

To our knowledge this dissertation is the first exploration of the effects of acute and chronic stress on human perception of robot emotional body language, with implications for robot design, both physical and virtual. Given the largely parallel findings that we observed for the poses presented by the physically-present versus virtually-instantiated robot, it is proposed that the use of virtual reality may provide a viable "sandbox" tool for more efficiently and thoroughly experimenting with possible robot designs, and variants in their emotional expressiveness. Broader psychological, physiological, and other factors that designers should consider as they create robots for high-risk applications are also discussed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Literature Review

Introduction

As capacity for robot design advances, researchers and designers are exploring real world applications for social robotic technology. Proposed uses include: as domestic assistants (Aldebaran Softbank Group, 2016), emergency responders during disasters (Murphy, 2014), guides for tourists, shoppers, or those unfamiliar with a certain location (Kanda et al., 2009). Although there is a plethora of proposed applications, one area of broad interest is healthcare. Rehabilitation robots can support recovery of function in lower and upper limbs (Klein, Gaedt, & Cook, 2013; Recio, Segura, Segura, & Waern, 2013), telerobots function as communication and monitoring devices (Casper & Murphy, 2003; Seelye et al., 2012), social robots can provide therapeutic companionship (Csala, Nemeth, & Zainko, 2012; Klein et al., 2013), and children with autism can learn to decode social cues through interactions with expressive robots (Pennisi et al., 2016).

In order to maximize the potential of assistive technology in all contexts, platforms must communicate effectively with humans. A major area of interest to robot designers is the potential for emotional expression to facilitate human-robot interactions (HRI). Emotion allows for quick transfer of highly complex information in human-to-human interactions (Batty & Taylor, 2003), facilitating rapid decision making and action.

Research in the area of robot emotional expression has yielded promising results for human ability to perceive robot emotion. Humans are able to distinguish emotional robot body language (Beck et al., 2013; Erden, 2013; Haring, Bee, & Andr, 2011; McColl & Nejat, 2014) and facial expression (Chammat, Foucher, Nadel, & Dubal, 2010; Dubal, Foucher, Jouvent, & Nadel, 2011) at rates significantly above chance. Use of body language has an advantage over facial expression in settings where vision is partially obscured, viewing conditions are degraded, or at great distances (Martinez, Falvello, Aviezer, & Todorov, 2015).

While it is promising that laboratory experiments have found that robot emotional expressions are capable of facilitating human-robot communication, the contexts in which robots will be functioning are far more complex. Additional factors that could influence communication must be considered for robots to be accepted and wholly effective in situ. How do individual psychological characteristics, such as a person's temperament, their current stress level, and their attitudes regarding robots, impact how an individual perceives robot interactions — and robot emotional expressiveness?

Human ability to perceive human emotional expression is altered by state-level stress (Mather, Lighthall, Nga, & Gorlick, 2010; Jefferies, 2008; Kret, Stekelenburg, Roelofs, & de Gelder, 2013; Clewett, Schoeke, & Mather, 2013). Assistive robots will inevitably interact with acutely stressed humans if they are present in our homes, work places, medical settings, and the scenes of disaster. If emotional expression is a key method of human-robot interaction, and perception of robot emotion is also altered when humans are under acute stress, the efficacy of assistive robots could be seriously undermined. People who are already stressed may experience additional frustration as they attempt to ascertain a robot's intentions, possibly leading them to refuse to engage. In some cases, this may lead to low-risk disuse of the robot, such as a child refusing robotic assistance with their homework. In other contexts, such as a stay in the hospital, disuse could lead to harm if the robot's services are instrumental for stabilization and on-going care. Despite the importance of human-robot communication in high-tension contexts, the literature is silent on the impact of stress on humans' ability to perceive robot emotions.

Another gap in the literature exists in relation to the impact of trait-level emotionality (temperament) on perception of robot emotions. Temperament has been found to impact perception of human facial expression (Yi, Murry, & Gentzler, 2016).

Those with a higher level of sociability find facial expressions easier to decode (Young & Brunet, 2011) and those who experience intense fear and happiness are able to more accurately perceive those emotions in others (Buchanan, Bibas, & Adolphs, 2010). If similar patterns of temperament informing perception of robotic emotional body language holds true, humans may individually experience the same robot very differently.

As robots permeate society there will likely be other factors that also influence human-robot communication. For example, those with prior experience with robots may find robot communication easier to decode. Opinions about the utility and role of robots in society could influence perception of their emotional expression. And chronic stress may interact differently than acute stress with emotional perception.

While existing studies have laid a solid foundation for verifying that humans can generally perceive robot emotional expression, a more sophisticated understanding of individual variance in human-robot interaction will be necessary as robots are deployed in high stakes domains. If state and trait-level stress, temperament, and attitudes towards robots are found to impact emotional expression perception, robot designers will need to explore ways to integrate this information into future designs.

A possible reason these factors may not yet have been explored by researchers is financial and other resource limitations. Robots are mechanically and computationally complex, resulting in great expense and effort to build or procure and conduct experiments. Robot proxies, in the form of 2D representations of 3D robots, have been used in studies. However, an extensive review of the literature has found that 2D robot representations yield different behavioral responses than do physically present robots (Li, 2015). Virtual reality (VR) technology, which presents images in three dimensions, may provide a more viable proxy for physical robots.

Creating virtual reality versions of robots is far more economical than building functional physical prototypes. If performance on behavioral measures is found to be similar for physical and VR robots, VR models of robots could be generated for testing design parameters. This access to economical stimuli would increase investigation of human perception of robots, greatly increasing our ability to explore the factors that impact human-robot interaction. Ultimately such knowledge will lead to well calibrated robot design, ensuring proper use of assistive robotic technology, especially in high-risk contexts such as healthcare.

Literature Review

Robots in Healthcare

The impact of stress, temperament, and attitudes towards robots have the potential to influence perception of robot communication in many domains. The inspiration for this dissertation came from healthcare, where robot presence is already increasing. Thus, results will be primarily interpreted in relation to their impact on humans in health and wellness settings, although we recognize that our findings will be generalizable to many other contexts.

As robots become more autonomous, the potential for applications in healthcare settings increases. Uses for robots in telemedicine, rehabilitation, and therapy have been investigated, with many other applications being proposed.

There is growing support for using telemedicine robots to facilitate communication between patients and doctors/surgeons who may be remotely located. Many rural healthcare systems successfully use tele-robots to consult with specialists outside the community (Marttos et al., 2013). Both remote and local physicians rate the systems highly for satisfaction with: mobility (97% and 90%, respectively), communication (91% and 97%), and visual abilities (91% and 97%) for conducting remote consultations (Marttos et al., 2013). Tele-robots are often as tall and wide as a human, with a monitor placed in the head position. The doctor's face is projected on the monitor so that patients have a sense of the doctor's presence during the consultation. Future versions of telemedicine devices may include robotic limbs that will allow physicians to conduct physical exams. Tactile sensors will allow physicians to obtain similar sensory information as compared to in-person examinations (Huang, Yu, Ming, Xiang, & Ge, 2008).

Rehabilitation robots can work with patients in healthcare facilities and at home to support recovery. Robots show potential benefits for support of children undergoing bone marrow transplant. A humanoid robot provided companionship and encouraged them to perform self-care tasks such as eating, taking medicine, waking in the morning and bathing, which are difficult for critically ill patients (Csala, Nemeth, & Zainko, 2012).

Robots have also been used to teach and reinforce physical therapy exercises for rehabilitation and physical skills learning (Malik, Yussof, & Hanapiah, 2014). Children with cerebral palsy have been found to engage significantly with a robot programmed to assist with developing skills in movement and communication (Fridin & Belokopytov, 2014). Therapeutic use of robots with children with autism has led to improved spontaneous language and reduced repetitive and stereotyped behaviors (Pennisi et al., 2016). Robots have also been used experimentally with adult stroke patients to build communication skills (Wade, Dye, Mead, & Matarić, 2011).

In order for these applications to be integrated more widely, it will be important for robots to engender the trust of patients and caregivers. When humans interact with robots in experimental conditions, a human researcher introduces the participant to the novel robot. Autonomous robots in healthcare settings will have to initiate — and sustain — interactions without the assistance of a human facilitator. Human-robot communication will be key to successful interactions and treatment.

Anthropomorphism and Robot Trust

Designers of service robots have explored anthropomorphism as a means to increase human trust of robots. According to Duffy (2003, p. 177) anthropomorphism is the "…rationalisation of animal or system behaviour through superimposing aspects of the human observer." Humans use this strategy to engage with other human agents. By assuming that another human would act as the individual would in similar circumstances, people are able to fairly accurately predict the behavior of other human agents (Dennett, 1989). By ascribing human traits to objects and animals, humans are able to explain and predict the behavior of those animals or objects. Humans have been found to readily engage in this social model with autonomous robots (Breazeal, 2003).

Increasing a robot's familiarity through anthropomorphic design leads to greater social acceptance (Fink, 2012). Humans are more likely to change their mind in order to agree with an anthropomorphic robot than one that is non-anthropomorphic (Zanatto, Patacchiola, Goslin, & Cangelosi, 2016). Further increases in decision alignment were found when an anthropomorphic robot engaged in social gaze (gazing at the object of appraisal and then the human) than non-social gaze (gazing exclusively at the object) (Zanatto et al., 2016). Humans can infer a great deal about the internal state of a robot by interpreting its gaze (Breazeal, 2003). In experiments with Kismet, a social robot, there was a significant difference in humans' sense of social connectedness with the robot depending on whether Kismet looked directly into their eyes, or merely at their face (Breazeal, 2003).

Many researchers emphasize the need to distinguish anthropomorphic design — integrating human-like qualities — from the goal of creating robots that are virtually indistinguishable from humans (Duffy, 2003; Fink, 2012; Breazeal, 2003). When humans perceive robots to be humans they create inaccurate mental models of robot behavior and capabilities. Overly high expectations of robot functionality can lead to frustration for human operators; if users' expectations are not well calibrated to the capacity of the robot, misuse or disuse can occur (Fink, 2012).

As long as a robot clearly communicates to human users that it does not have human-level abilities (e.g. through behavior and visual cues), emotional expressiveness increases human comfort with the robot (Koschate, Potter, Bremner, & Levine, 2016). Socioemotional interaction increases trust between humans and robots (Lohani, Stokes, Mccoy, Bailey, & Rivers, 2016). Students reported more trust and feelings of companionship with a robot that exhibited vulnerability, and disclosed more personal information to an expressive robot (Martelaro, Nneji, Ju, & Hinds, 2016). Robots that are not anthropomorphic are not likely to be socially accepted by humans (Duffy, 2003).

While utilizing gaze has shown promising results in developing human-robot trust, this aspect of anthropomorphism has drawbacks in healthcare contexts. One major challenge is that the working environment cannot be controlled. Medical robots will encounter patients in different positions (e.g. sitting, lying down, slightly reclined, turned to one side or another) and with differing abilities to communicate (factors include: language spoken, ability to see, hear, speak, comprehend, read). A robot that is designed to engender a sense of connectedness primarily through gaze will not develop trust with a patient whose head is turned away from the robot, or who may have limited mobility, low vision, or a line of sight that is partially obscured by medical equipment. One option to mitigate these issues is to utilize emotional body language, which can be seen and interpreted at greater distances and requires less precision in directing the gaze (Bethel & Murphy, 2008).

Perception of Robot Emotional Body Language

Emotions are often categorized on two measures: valence and arousal (Russell, 1980). Valence, or affect, is the degree of positivity or negativity. Arousal is the degree of intensity, or energy level. Negative valence-high arousal (NH) emotions include: angry, tense, afraid; negative valence-low arousal (NL) emotions include: sad, bored, depressed; positive valence-low arousal (PL) emotions include: relaxed, content, serene; positive valence-high arousal (PH) emotions include: happy, delighted, pleased (Russell, 1980).

The majority of neuroimaging studies on human perception of robot emotion have used facial stimuli. There are specific brain regions that process human facial information (Allison, Puce, & McCarthy, 2000; Puce, Allison, Asgari, Gore, & McCarthy, 1996), and even schematic faces facilitate emotion detection, which is due to the presence of salient emotion features such as two symmetrically located eyes and an expressive mouth (Eger, Jedynak, Iwaki, & Skrandies, 2003).

These findings led Dubal et al. (2011) to hypothesize that robotic faces with salient human features, but that were otherwise mechanical in appearance, would evoke similar event-related potentials (ERP) in response to happy and neutral stimuli as compared to human faces. The study used electroencephalography (EEG) to measure ERPs in the P1 and N170 components. The P1 wave is modulated by emotion while the N170 is modulated by facial configuration. They found that the occipital P1 component was enhanced for happy as compared to neutral stimuli for both human and robotic faces. The response of the temporo-occipital N170 component differentiated between robotic and human stimuli; latency was shorter and amplitude higher for human. This indicates that robot emotion can be detected, but that there is some mechanism that differentiates between human and robot faces.

Amygdala activation in response to human and avatar emotional expression has led to similar conclusions (Moser et al., 2007). Analyses of functional magnetic resonance imaging (fMRI) data showed significant amygdala activation from both avatar and human stimuli, but stronger neural responses for human stimuli in the fusiform gyri, cerebellum, left superior temporal gyrus, and rectal gyri (including several face sensitive structures). These findings indicate that patterns of human brain activity effectively distinguish between avatar and human faces, but Moser et al. (2007) questions whether this will remain true as realism increases in avatars.

Emotional facial expression and body language are processed similarly for human stimuli (de Gelder, 2006). Faces and bodies are processed at similar speeds (de Gelder, 2006), using the same neural structures (Hadjikhani & de Gelder, 2003). Both produce the N170 waveform (Meeren, van Heijnsbergen, & de Gelder, 2005), which means that it is modulated by both facial and body configuration.

Some researchers have proposed that body language may play a central role in emotion perception; in many cases it may be more salient than facial expression. When images of distorted faces and bodies were presented, both types of stimuli affected ERPs on the N1 waveform, but the effect was significantly faster for bodies as compared to faces (Gliga & Dehaene-Lambertz, 2005). This suggests that body configuration may be perceived earlier, therefore playing a key role in informing emotional perception.

When stimuli with mismatched facial and body language emotional expression were presented, identification of the facial expression was biased towards the emotional body language (EBL) (Meeren et al., 2005). In another mismatch paradigm, researchers found that body cues were more salient than were facial cues in determining overall emotional positivity or negativity (Aviezer, Trope, & Todorov, 2012).

Behavioral studies have found that humans recognize emotional robot body language at rates significantly above chance. Using the Pleasure-Arousal-Dominance model, Häring, Bee & André (2011) investigated accuracy rates for anger, sadness, fear, and joy by exploring body movement, sound, and eye color. Of the three modalities, only body movement yielded consistently above chance accuracy for emotional identification (Häring et al., 2011). The study poses were based on a variety of sources, including emotional expression libraries and theoretical work. They conducted a large pilot test to determine which poses reached the highest accuracy levels and identified options that had the following accuracy rates: anger (94.0%), fear (85.1%), sadness (95.5%), and joy (74.6%) (Häring et al., 2011).

In a study using static emotional body poses for the robot Nao, adults and children were able to identify pride (88% and 100%, respectively), happiness (73% and 83%), excitement (73% and 63%), fear (92% and 92%), sadness (85% and 92%), and anger (88% and 58%) at accuracy rates significantly above chance (Beck et al., 2013). The poses for this study were based on performances by professional actors.

In contrast, Erden (2013) attempted to generate emotional body language poses for Nao based on Coulson's (2004) research into human emotional poses. A total of 176 computer figures in various postures were generated to determine which configurations led to the highest accuracy rates for emotional body language identification (Coulson, 2004). Erden (2013) translated these poses for Nao, which necessitated some adjustments due to physical differences between humans and robots. The poses that Coulson had identified as leading to the highest rates of accurate identification were not the same that led to the best rates in the robot Nao (Erden, 2013). Anger, happiness, and sadness reached 90% concordance for numerous postures. Fear and surprise only had a few postures that reached 60-70% concordance, and none of the disgust poses reached 50%.

The variety in these approaches to pose creation and the accuracy/concordance rates that can be achieved indicate that pilot testing robot poses is important. Both Erden (2013) and Beck et al. (2013) used a forced-choice method for emotion identification. Häring et al. (2011) was less prescriptive, instead asking for each pose to be rated on pleasure (i.e., valence), arousal (i.e., intensity), and dominance (i.e., degree of control). This system allows for more fine-grained measurement and facilitates across-pose comparison. It also avoids the semantics of participants' individual definitions for particular words (e.g. how they define *happiness* versus *joy*), which are likely to be culturally influenced.

While there do not appear to be any neuroimaging studies related to perception of robot emotional body language, the behavioral studies discussed above suggest that it is processed similarly to human emotional body language. One possible reason for the paucity of neuroimaging studies on perception of robot emotional body language may be the many modalities of body language, which cause a challenge for determining how body language information is processed. The modalities of body language have been defined in various ways by researchers, but fall into three broad categories: pose, movement, proxemics (Beck et al., 2013).

Information regarding pose — the position of the body in a given moment of time — is sufficient for emotion recognition (Beck et al., 2013; Kleinsmith & Bianchi-Berthouze, 2011). Body movement can also communicate emotional state, but recognition accuracy for movement is impaired when pose information is distorted (Kleinsmith & Bianchi-Berthouze, 2011). Movement and pose are processed on separate neural pathways, with motion information only utilized to resolve inconsistencies (Lange & Lappe, 2007). Although distance (proxemics) between two agents does not give specific information about emotional states, it is a dimension that must be considered in order to design realistic emotional behavior (Beck et al., 2013). Given its primacy in body language perception, we selected pose as the modality for expression of emotion for this series of studies.

Stress, Temperament, and Emotion Perception

In general, perceiving emotions under stressful conditions, whether depicted in pictures, objects, or faces, is influenced by emotional valence and emotional arousal. As noted earlier, valence (or affect), is the degree of positivity or negativity of the emotion, whereas arousal is the degree of intensity, or energy level. Negative valence-high arousal (NH) emotions include: anger, hostility, rage. Negative valence-low arousal (NL) emotions include: sadness, boredom, guilt. Positive valence-low arousal (PL) emotions include: relaxed, content, curious. Positive valence-high arousal (PH) emotions include: happiness, joy, surprise.

Perception of emotion under stress often involves the interaction of valence and arousal. For example, a stress-inducing event may lead to hypervigilant processing of negative high-arousing stimuli (Weymar, Schwabe, Löw, & Hamm, 2012), and is associated with enhanced functional coupling between neural structures that amplify the stress response (van Marle, Hermans, Qin, & Fernández, 2010). Similarly, acute stress has been found to increase attention to threatening negative stimuli (Kret et al., 2013) and the first fixation on a threatening stimulus lasts longer for highly stressed individuals than those at baseline (Quigley, Nelson, Carriere, Smilek, & Purdon, 2012). Likewise, for stimuli with higher arousal ratings, human startle responses increase markedly for stimuli that are unpleasant versus pleasant (Lang, 1995). Eye movement data indicate that stimuli with higher levels of emotional arousal in complex scenes elicit denser attentional allocation (Ni et al., 2011). And even under neutral non-stressed conditions, evaluation response times are faster for negative high-arousal and positive low-arousal stimuli than for negative low-arousal and positive high-arousal (e.g., Robinson, Storbeck, Meier, & Kirkeby, 2004; see also Recio, Conrad, Hansen, & Jacobs, 2014).

But attention is not always pulled toward negatively-valenced emotional stimuli: Sometimes a heightened negative mood shifts an individual's attention away from negative stimuli toward positive (Ellenbogen, Schwartzman, Stewart, & Walker, 2002; Newman & Sears, 2015; Sanchez, Vazquez, Gomez, & Joormann, 2014), and decreases sensitivity to emotional facial stimuli (DeDora, Carlson, & Mujica-Parodi, 2011).

Temperament has also been found to play a role in attention to emotional stimuli. Neuroimaging has revealed individual differences in brain activation in specific regions during cognitive-affective tasks (Canli, 2004). Those scoring high in extraversion and neuroticism, respectively, on the Positive and Negative Affect Scale (PANAS) (Watson, Clark, & Tellegen, 1988), experienced different activation patterns while passively viewing stimuli from the International Affective Picture Series (IAPS) (Lang & Greenwald, 1993) (Canli, 2004). Those scoring higher on extraversion showed greater activation to positive stimuli, while neuroticism was positively correlated with increased neural activation to negative stimuli (Canli, 2004).

These findings are consistent with earlier work that found that introverts are more easily drawn to negative stimuli than are extraverts (Derryberry & Rothbart, 1988). Shifting attention towards positive stimuli can enhance or facilitate the maintenance of elicited arousal and emotion, while shifting attention away from a negative stimulus can attenuate these experiences (Derryberry & Rothbart, 1988).

Mood affects the perception of non-congruent emotions, with happy moods hampering the perception of sad expressions, and sad moods hampering the perception of happy ones (Schmid & Schmid Mast, 2012). Thus, it can be anticipated that extraverts would allocate more attention towards positive emotions and those scoring high on neuroticism would focus on negative expressions, while people scoring high on attentional control could moderate their engagement.

In Evans and Rothbart's model of adult temperament, extraversion is associated with high scores on the subconstructs of high-intensity pleasure, positive affect, and sociability (Evans & Rothbart, 2007). Neuroticism is correlated with negative affect (Costa & McCrae, 1980; Watson, Clark & Tellegen, 1988; Larsen & Ketelaar, 1991), for which there are four subconstructs in the Evans and Rothbart model: fear, sadness, discomfort, and frustration (Evans & Rothbart, 2007). Effortful control includes three subconstructs: attentional control, inhibitory control, and activation control (Evans & Rothbart, 2007). The last factor in their model is orienting sensitivity with three subconstructs: neutral perceptual sensitivity, affective perceptual sensitivity, and associative sensitivity (Evans & Rothbart, 2007).

Research in this area has traditionally investigated valence. Only recently have researchers sought to understand the relationship between intensity, or arousal, temperament, and emotion recognition. Consistent with earlier work, those scoring high on extraversion perceived positive emotions more accurately (measure of valence). In regards to arousal, those scoring higher on negativity perceive more intensity in angry expressions and lower intensity in sad expressions (Yi, Murry, & Gentzler, 2016). Given that emotion perception often involves the interaction of valence and arousal, further investigations into the relationship between these dimensions and temperament will deepen our understanding of emotion perception.

The impact of stress and temperament on perception of emotional body language has not been investigated, but similar to facial expression, different patterns of perception have been found for different emotion categories. Threatening body postures are perceived more rapidly than are happy body postures (Gilbert, Martin, & Coulson, 2011). For static emotional body poses, happy has a lower recognition rate than angry, fearful, and sad (de Gelder & Van den Stock, 2011). However, it has been found that body postures in high arousal categories can be confused for one another. Surprised body postures are confused with angry bodies, and angry bodies are confused with happy bodies (Kret et al., 2013). Since human and robot emotional body language are processed similarly, confusing happy and angry robot poses could lead to misunderstandings between humans and robots.

Given the shifts that occur in emotion perception when humans are acutely stressed, integrating robots into the healthcare domain poses challenges for communication; patients and caregivers often experience stress in medical settings. Some patients are acutely stressed when they arrive at medical facilities. Sources of stress that may be directly related to their illness or injury include pain or discomfort, fear/anxiety about a diagnosis, fear/anxiety about a medical procedure, financial concerns about paying for care, inability to fulfill family, social, occupational, or other roles, or discomfort or trauma around engaging with the medical system (de Sá Dias, Resende, & Diniz, 2015). Some of the many causes of stress in medical facilities include: not understanding treatment or diagnosis, pain from procedures, not being in control, invasive medical devices such as intravenous medication delivery systems or breathing tubes, disruption of circadian rhythm, lack of privacy, and perceiving other patients' discomfort or pain (de Sá Dias et al., 2015; Novaes et al., 1997; Novaes, 1999).

Caregivers may also experience stress before, during, and as a result of, visits to medical facilities. It has been widely documented that parents of children undergoing medical procedures experience stress. Parents with children undergoing radiological procedures exhibit short-term physiological stress such as elevated anxiety, heart rate, and blood pressure (Alexander, 2012). Their stress has been found to exacerbate their child's stress level (Alexander, 2012).

Many patients and caregivers also experience chronic stress (Von Känel et al., 2011). For the purposes of this dissertation, we distinguish chronic stress as separate from state-level anxiety; when considering chronic stress, we are interested in the stress experienced as a result of prolonged difficult psychological and/or physical circumstances. Surprisingly, we were unable to locate any existing work on the impact of chronic stress and emotion perception.

Given that perception of emotions has been found to be affected by mood and temperament, those scoring higher on measures of negative affect may perceive negative body expressions as more salient, frequent, or threatening than intended by designers, leading to distrust or misuse of assistive medical robots. Extraverts may interpret robot emotional expression as more positive in affect, causing miscommunication. It is additionally possible that there are interactions between acute stress, chronic stress, and temperament, which will cause shifts in perception of emotions and alter human-robot interaction.

Designing for the Real World

To explore how people's acute and chronic stress levels, attitudes, perceptions, and temperament correlated with their interaction with a robot, we first created and normed a set of body poses for Nao, a humanoid social robot. Poses were designed to vary along both valence and arousal scales such that five distinct categories were created: negative valence-high arousal (NH), negative valence-low arousal (NL), neutral (NE), positive valence-low arousal (PL), and positive valence-high arousal (PH). We then presented the most consistently identified poses from each category to participants in either an acute or low-stress condition. Participants were asked to rate each pose on the degree of emotional valence – negative/unhappy to positive/happy – and level of arousal – calm/relaxed to animated/excited – the robot conveyed. Consistently identified poses were selected and used in Study 2, which introduced an acute stress vs. low stress manipulation. Acute physiological and cognitive stress was induced using the Maastricht Acute Stress Test, a well-validated task shown to activate both the sympathetic-adrenal-medullary system and the hypothalamic-pituitary-adrenal axis (Smeets et al., 2012). Participants then completed the Adult Temperament Questionnaire (ATQ) Short Form (Evans & Rothbart, 2007), a modified Chronic Stress Inventory (CSI) (Turner, Wheater, & Lloyd, 1995), and a robot attitudes questionnaire.

In Study 3 we introduced the virtual reality robot condition. Extensive research has found differences in human response to physically embodied agents (i.e. an agent that has a visible body, which may only be visible on a screen) versus physically present agents (i.e. an agent that is co-located with the user). Research has found that physical robots are more compelling than are telerobots or 2D virtual agents. People rate their interaction with physical robots to be overall more positive than engaging with the same robot over live video feed (Bainbridge, Hart, Kim, & Scassellati, 2011). They are also more likely to fulfill a strange request and to give the robot more personal space if it is physically co-located (Bainbridge et al., 2011).

In a survey of 33 studies exploring reactions to virtual agents versus physical robots, physically present agents were considered more favorable in 83% of the studies

(Li, 2015). On behavioral measures, seven out of ten studies found significant effects for co-present versus tele-present robots, including reports of main effects of trust and a greater sense of the physical robot’s utility. Significant effects were stronger for a physical versus virtual agent in 19 out of 27 studies (Li, 2015). Among other findings, participants paid more attention and were more engaged with a co-present as compared to virtual agents. Little to no difference was found between responses for a physical robot presented digitally on a screen versus a virtual avatar, highlighting the importance of co-location for optimal interaction (Li, 2015).

None of the studies in the survey investigated human performance and attitude towards robots presented in virtual reality. VR robots are embodied, and may be considered co-located, because they are present in the same (virtual) environment as the perceiver (Li, 2015). If behavioral and perceptual performance as well as attitude towards robots is similar for physical and virtual reality robots, virtual reality could potentially be employed as a low cost "sandbox" for robot design. As we develop a more nuanced understanding of how factors such as stress level and individual traits impact human-robot interaction in healthcare contexts, it will be important to experiment with adjusting robot design parameters to optimize engagement. While altering parameters in physical robots can be cost prohibitive, virtual models of the same robots can be generated more economically – and more parameters can be changed in shorter periods of time.

There are some indications that virtual reality may be a viable option as a design tool. In a norming study of virtual reality avatar facial emotional expressions, traditionally developing adults identified the stimuli at levels significantly above chance: sadness (97%), joy (92%), anger (86%), surprise (81%), disgust (72%), contempt (58%), and fear (53%) while viewing them in a virtual environment (Bekele et al., 2014).

Similarly, 3D avatar faces presented on a flat screen computer have been found to reach recognition rates greater than 80% for disgust, happiness, sadness, and surprise (Wallraven, Breidt, Cunningham & Bülthoff, 2008). These rates were as good as or better than recognition of video clips of the same actor performing the facial expressions. Although these studies do not investigate robot facial expressions or body language, they do suggest that humans are able to perceive emotional expressions in virtual reality (or 3D) stimuli. If agents rendered in virtual reality can also convey information about robot emotional expression that tracks closely with human perception of physical robots, then virtual reality could become a powerful tool for robot design.

Together, the three studies reported here ask: How is an individual's perception of a robot's emotional body language affected by acute cognitive and physiological stress? Is perception when under stress dominated by emotional valence, emotional arousal, or the combination? Does acute stress draw attention or, instead, repel attention to emotional body language poses that are negative or highly arousing? Do any of the same patterns hold true for chronic stress? And what effect do individual differences have on robot emotional body language? Does temperament impact perception of valence or arousal? Can familiarity and/or acceptance of robots influence human-robot communication? How do these factors impact human’s ability and willingness to engage with robots? And lastly, how are behavioral measures and attitudes towards robots affected by whether the robot is physically present or a virtual-reality rendition? Answering these questions will increase our understanding of the dynamics of emotion perception under high and low acute and chronic stress and considering individual differences, with broad implications for how we communicate with and design robots.

Chapter 2: Study 1 – Norming Study

Study 1 - Stimulus Norming of Robot Emotion Poses

To test how well people under acute stress can judge the emotions conveyed by a robot's body postures, we first developed and normed a set of body poses for the humanoid robot known as Nao (Aldebaran Softbank Group, 2016). Five sets of emotional categories were created that conveyed negative or positive emotional valence each accompanied with high or low arousal, as well as neutral body poses.

Methods

Participants

Thirty participants (16 female, M age=21.27, SD=3.23), recruited through campus posters and a university online research participation website, took part in the norming study. To be eligible, participants were required to be between the ages of 18-35 years and have lived in the U.S. for at least 5 years. Four additional participants were omitted from the final analysis: 1 for not following directions, 2 due to equipment failure, and 1 person who was later determined not to meet the selection criteria. Participants were offered either extra credit or $10 for a 60-minute session. The study protocol was approved by the University of Minnesota Institutional Review Board.

Robot pose generation

The emotional body language poses were developed using a physical robot: the autonomous programmable humanoid robot known as Nao. Nao is 58 centimeters tall, weighs 4.3 kg, has 25 degrees of freedom of movement, but no capacity to change facial expression. The robot's eyes are able to change color, but remained gray/blue for all poses. The body of the robot was white plastic, with orange accents.

Ten poses were designed for each emotion category: negative valence-high arousal (NH), negative valence-low arousal (NL), neutral (NE), positive valence-low arousal (PL), positive valence-high arousal (PH), for a total of 50. Poses were designed based on information garnered from several sources: professional puppeteers, previous work in robot body language (Beck et al., 2013; Erden, 2013; Häring et al., 2011) and animation best practices.

Robot pose ratings

Participants viewed 50 distinct poses, 10 from each emotion category. Each pose was presented three times using block randomization, for a total of 150 presentations. Participants used the Self-Assessment Manikin 9-point scales to rate the valence and arousal for each pose (Bradley & Lang, 1994). After rating valence and arousal, participants assigned an open-ended emotion descriptor to each pose (e.g. happy, sad, perplexed). Responses were captured using E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburg, PA). The transition from one robot pose to the next was occluded by the experimenter holding a black foam board in front of the robot.

As a control, participants rated a validated set of still images of 50 body language poses performed by humans (de Gelder & Van den Stock, 2011). The set was developed using forced-choice emotion categories (anger, fear, sadness, disgust, happiness), but in keeping with the procedure for the robot pose identification task, participants were asked to rate the human poses on the valence and arousal Self-Assessment Manikin 9-point scales. No emotion descriptors were recorded. Participants also completed the Negative

Attitudes towards Robots Scale (NARS) and answered questions about previous experience with robots and virtual reality.

Study 1 Laboratory Configuration

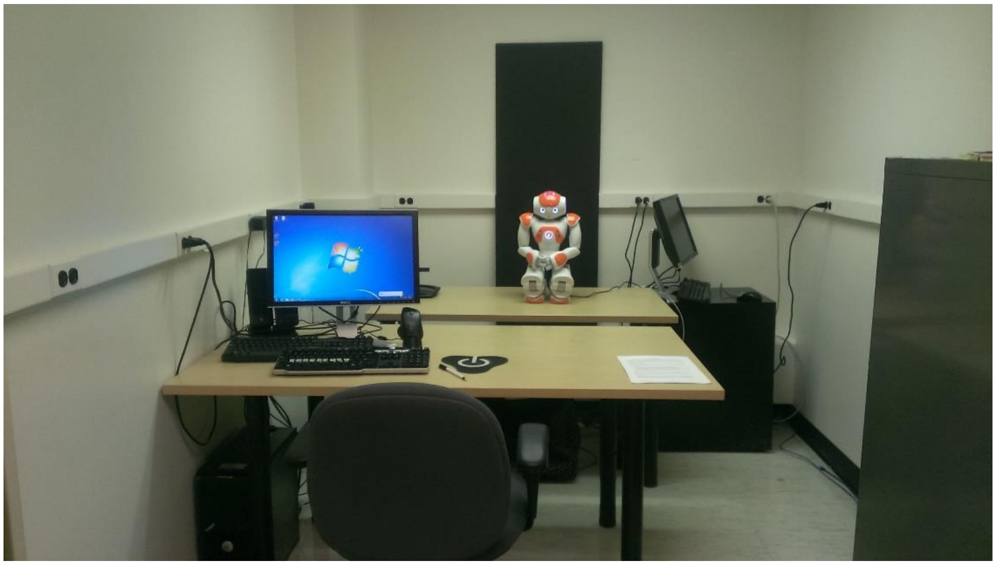


Fig 1. Experimental room set-up. Participants sat at the computer in the foreground and entered pose ratings into E-Prime 2.0. Experimenter stood next to the robot and advanced the poses with keystrokes on the computer to the right of the robot. While the robot moved from pose to pose the experimenter held the black foam board between the participant and robot.

Results

Robot pose selection

An iterative process was used to identify the poses that were most consistently rated for both their valence and arousal.

Qualitative analysis

Scatterplots were generated for Valence versus Arousal, Valence, and Arousal for all three blocks separately and the mean of all blocks for each pose. Bar graphs were generated for the descriptors (names for emotions) for each trial of each pose. Two experimenters independently coded the descriptor data. Differing forms of the same root word (e.g. excite, excitement, excited and proud, pride) were combined. Words that were similar but did not share a root were left separate (e.g. victorious and winning). Then:

- Two experimenters reviewed the graphs together. We looked for poses that most cohesively fit in the intended quadrant, and had the least variation in the types of descriptors that were given. The top performers for each category were recorded, with no specified minimum or limit on the number of poses to be identified.
- The PI analyzed the graphs, blind to pose, for performance based solely on Valence. Poses with the most conclusive responses for valence in each category were recorded.
- The PI analyzed the graphs, blind to pose, for arousal and identified poses that were consistently rated as low or high arousal. For this pass on the data, low arousal was defined as less than 5 and high arousal as over 5.

Next Valence Mean, Valence Median, Arousal Mean, Arousal Median, and Reaction Times for Valence, Arousal, and Names were calculated. Pairwise two-tailed *t*-tests were run on Valence Mean and Arousal Mean to determine possible effects of gender. Any pose with $p < 0.05$ for the effect of gender was rejected. Variance and trial-to-trial correlations for valence were calculated for each pose. Reaction Time data were also collected, but were deemed too noisy (within subject) for any conclusive information to be gleaned.

Quantitative analysis

Valence and arousal ratings for each pose were analyzed and the decision to include a pose was made based on the following considerations (listed here in order of priority):

- Effects of gender: all poses with $p<.05$, indicating a significant gender-based difference in the rating of the poses, were disqualified.
- Valence and Arousal: Poses with consistently corresponding mean and median in the target range (depending on emotion category) were prioritized.
- Valence: positive valence was defined as any rating >5.0, negative was defined as any rating <5.0, with 5.0 defined as neutral.
- Arousal: high arousal was defined as any rating >6.0, low arousal was defined as any rating <6.0, based on the distribution of ratings in the norming study.
- Variance: High variance in the ratings between blocks was avoided. For the given results, high variance was defined as >3.0.
- Correlation: Low pairwise correlation of the ratings across different presentations of the same pose were avoided. Given the correlation results, low correlation was defined as any value <0.30.

The poses selected for inclusion via the quantitative and qualitative analysis were compared and the five most consistent poses in each category were selected for inclusion in Study 2. For the Negative Valence-High Arousal category, the mean valence for the selected poses was 3.69 with arousal mean=7.31; Negative Valence-Low Arousal, valence mean=3.04, arousal mean=5.63; Neutral, valence mean=5.08, arousal mean=4.10; Positive Valence-Low Arousal, valence mean=5.67, arousal mean=4.51; Positive Valence-High Arousal, valence mean=7.16, arousal mean=7.22.

Table 1 provides the poses that were identified by each method. Examples of the selected poses and scatterplots of their accompanying valence/arousal ratings for each of the five categories of stimuli are shown in Figure 2. See Appendix A for images of the 25 selected poses. Quantitative values for the selected poses are provided in Table 2. See Appendix B for emotion ratings for all 50 poses.

Selected Robot Poses by Analysis Type

Pose Category	Joint Experimenter Qualitative Graph	PI Blind Valence Qualitative Graph	PI Blind Arousal Qualitative Graph	Quantitative Analyses	**Selected Poses**
NH	43, 44, 45, 46, 47	43, 44, 46, 47	42, 43, 44, 45, 46, 47, 48, 50	44, 45, 46, 47	**43, 44, 45, 46, 47**
NL	11, 12, 13, 14, 15, 17, 19, 20	12, 13, 14, 16, 17	None (Low <5)	12, 14, 15, 19	**12, 14, 15, 17, 19**
NE	25, 29, 38, 39, 40	32, 36, 39, 40	31, 32, 33, 34, 37, 38, 39, 40	31, 32, 33, 34, 35, 36, 37, 38, 39, 40	**31, 32, 33, 39, 40**
PL	21, 22, 24, 27, 31, 32, 35	21, 22, 24, 25, 27	24, 25, 28, 29, 31, 32	21, 22, 24, 27, 35	**21, 22, 24, 27, 35**
PH	1, 5, 6, 7, 8	1, 5, 6 ,7, 8	1, 2, 3, 4, 5, 6, 7, 8	1, 5, 6, 7, 8	**1, 5, 6, 7, 8**

Tab 1. Ten robot body language poses were designed for each emotion category – Negative Valence-High Arousal (NH), Negative Valence-Low Arousal (NL), Neutral (NE), Positive Valence-Low Arousal (PL), and Positive Valence-High Arousal (PH). This table identifies the poses that were selected through multiple methods of analysis, and those that were ultimately chosen for use in subsequent studies.

Examples of Selected Poses from Each Emotion Category

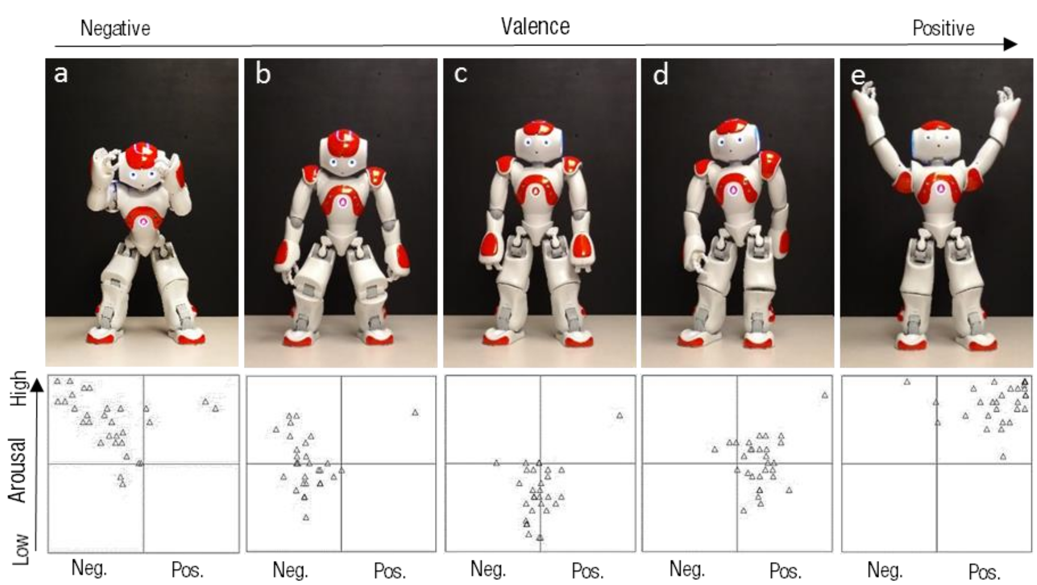


Fig. 2. Example of poses selected for Study 2. (a) negative valence-high arousal, (b) negative valence-low arousal, (c) neutral, (d) positive valence-low arousal, (e) positive valence-high arousal.

Ratings for Poses Selected for Future Studies

Selected Negative Valence-High Arousal (NH) Robot Pose Ratings											
Study 1 Pose #	Study 2&3 Pose	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
43	1	4.06	4.00	6.44	6.50	4.56	3.84	4.05	0.79	0.69	0.68
44	2	3.51	3.00	6.98	7.25	3.13	4.18	3.73	0.65	0.76	0.72
45	3	3.96	3.00	8.09	8.50	6.72	6.55	7.76	0.36	0.75	0.62
46	4	3.74	3.00	7.56	8.00	5.43	5.69	7.52	0.71	0.71	0.65
47	5	3.20	2.00	7.49	8.00	8.19	3.34	6.63	0.36	0.34	0.59

Selected Negative Valence-Low Arousal (NL) Robot Pose Ratings											
Study 1 Pose #	Study 2&3 Pose	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
12	6	2.57	2.25	5.81	6.50	0.96	3.10	1.52	0.26	0.35	0.45
14	7	2.33	2.00	6.68	7.50	0.96	1.89	0.66	-0.16	0.00	0.38
15	8	3.39	3.00	5.30	5.50	2.69	2.67	2.99	0.66	0.72	0.58
17	9	3.52	3.00	5.05	5.25	1.84	1.62	1.87	0.50	0.57	0.68
19	10	3.39	3.00	5.30	5.50	2.69	2.67	2.99	0.66	0.72	0.58

Selected Neutral (NE) Robot Pose Ratings											
Study 1 Pose #	Study 2&3 Pose	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
31	11	5.37	5.00	3.80	4.00	0.87	0.87	1.07	0.56	0.76	0.50
32	12	5.15	5.00	3.66	4.00	1.41	1.73	1.70	0.52	0.44	0.58
33	13	4.81	5.00	5.28	4.50	1.83	2.12	1.95	0.06	0.40	0.00
39	14	5.13	5.00	4.19	4.50	1.56	1.96	1.62	0.35	0.35	0.24
40	15	4.95	5.00	3.57	3.13	0.79	1.58	1.74	0.49	0.45	0.50

Selected Positive Valence-Low Arousal (PL) Robot Pose Ratings											
Study 1 Pose #	Study 2&3 Pose	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
21	16	5.71	6.00	5.14	5.75	2.04	1.61	1.66	0.44	0.63	0.44
22	17	5.63	6.00	4.34	4.25	1.11	1.06	1.89	0.10	0.38	0.11
24	18	5.42	5.00	3.55	3.00	1.50	1.70	0.90	0.32	0.67	0.36
27	19	6.04	6.00	5.06	5.00	1.09	1.64	2.03	0.30	0.35	0.28
35	20	5.53	5.00	4.44	5.00	1.84	1.15	1.67	0.74	0.69	0.69

Selected Positive Valence-High Arousal (PH) Robot Pose Ratings											
Study 1 Pose #	Study 2&3 Pose	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
1	21	7.82	8.00	7.99	8.00	2.38	2.89	3.79	0.10	0.43	0.74
5	22	6.65	7.00	6.38	7.00	1.36	1.29	1.96	0.11	0.29	-0.02
6	23	7.38	7.50	7.53	8.00	1.36	2.00	3.73	0.71	0.64	0.50
7	24	7.18	8.00	7.73	8.00	4.97	3.45	3.05	0.32	0.39	0.17
8	25	6.75	7.00	6.47	7.00	1.32	1.29	1.22	0.55	0.61	0.69

Tab 2. Ratings of robot body language poses selected for inclusion in future studies.

Discussion

It was possible to select five poses that met the criteria for each emotion category. By measuring emotion perception on the 9-point valence and arousal scales of the Self-Assessment Manikin, as opposed to providing emotional category options for forced-choice selection, we ensured that the poses were more accurately calibrated.

It is interesting to note that the average valence rating for the NL poses (3.04) was lower than that for the NH poses (3.69). We had expected the opposite as Negative Valence-High Arousal emotions (e.g. anger, rage, fear) are always coded as more negatively valenced than Negative Valence-Low Arousal emotions (e.g. sadness, depression) in emotion matrices. However, the relationship between the average arousal scores was in the anticipated direction, with the NH poses having a higher arousal (7.31 vs 5.63) rating. Analysis of the descriptors assigned to poses in the categories confirmed that participants perceived the types of emotions we had endeavored to design.

These outcomes suggest that, when perceiving robot body language, people likely do not perceive valence and arousal as mutually exclusive measures, which aligns with findings for human perception of human emotional facial expression (Russell, 1980), as well as with broader research findings discussed in the Introduction. These results also imply that there may be similar patterns between perception of human and robot emotional expression that go beyond the basic ability to detect an expressive pose's general emotional category.

Chapter 3: Study 2 – Stress Induction

Study 2 – Stress Effects on Perception of Emotional Robot Body Language

Having developed a normed set of stimuli, we introduced acute and low stress conditions. The aim of Study 2 was to answer our key questions: How is an individual's perception of a robot's emotional body language affected by acute cognitive and physiological stress? Is perception when under stress dominated by emotional valence, emotional arousal, or the combination? Does acute stress draw attention or, instead, repel attention to emotional body language poses that are negative or highly arousing? We presented each pose three times in block randomized fashion. Repeated presentations increased our power for data analysis and provided an opportunity to detect effects of increasing familiarity with the robot by allowing comparisons of ratings for the first vs. last pose presentation.

Methods

Participants

Participants were recruited through posters on campus and a university research participation website. To be eligible, participants were required to be between the ages of 18-35 years, in good health, and have lived in the U.S. for at least five years. Participants were offered either extra credit or $12 for a 65-minute session.

Participants were pre-screened for medical conditions (cardiac and neurological) that are contraindications for acute stress induction procedures. If they had any of these medical conditions they were not allowed to proceed. Participants were also screened for depression and anxiety with the Brief

Symptom Inventory (BSI) (Derogatis & Melisaratos, 1983) and were not included in the study if they scored 11 or higher for either measure. One potential participant was screened out due to a high BSI score.

Eligible participants were randomly assigned to stress conditions. We planned for, and tested, a total of 60 participants (24 male, *M* age=20.94, *SD*=2.95), with 30 participants per condition. Six additional participants withdrew from the study due to physical discomfort related to the stress induction.

Procedure Summary

Participants gave informed consent, completed screening documents, and were randomly assigned to a condition (high vs. low stress) (10 minutes). Next, participants completed the stress induction task (20 minutes), the robot pose identification task (35 minutes), a human emotional body language pose identification task as a control (5 minutes), and a series of questionnaires, including the Adult Temperament Questionnaire, attitudes towards robots, and a modified version of the Chronic Stress Inventory (15 minutes). Last, they were debriefed (5 minutes), for a total session time of 90 minutes.

Stimuli

The five most consistently identified poses in each emotion category from Study 1 comprised the set of 25 stimuli in Study 2. These were presented using the same autonomous robot Nao as in Study 1. However, participants completed Study 2 in a specially created test environment. The experiment took place in a windowless lab measuring 3.67m × 3.67m × 2.13m. Inside the main room, a smaller 2.44m × 2.44m × 1.83m lab was built. The walls were painted matte black. To reduce reflection and glare, the overhead florescent lighting was masked with white paper. There were two tables in the inner room. Participants initially sat at the smaller table to complete the paperwork and stress induction. They then moved to a larger one where the main task was completed. A large physical version of the valence and arousal SAM rating scales was affixed to the large table. The robot was placed at the far end. During the main task, the experimenter entered the subject's responses into the computer at the smaller table. When seated at the larger table, the participant's view was limited to the black walls, ceiling, table, SAM scales, and Nao. This test environment was created to enable a photorealistic virtual model to be built for the addition of the virtual reality condition in Study 3. Figure 3 illustrates the layout.

Laboratory Configuration for Study 2

Fig. 3 Participants completed the study in a specially constructed room. The walls were painted matte black to reduce glare and simplify the visual surroundings in preparation for modeling the environment in virtual reality for the Study 3 protocol. Participants completed the initial paperwork and stress induction task at the smaller desk in the foreground before moving to the larger table with the robot.

Stress Induction and Control

The Maastricht Acute Stress Test (MAST) was used to induce stress in the acute stress condition (Smeets et al., 2012). The MAST consists of two tasks: math and cold water hand immersion. The mental math task required participants to subtract two-digit numbers progressively from 2043. The participant was given a different number to subtract for each trial, and had to begin with the number they had ended on in the previous trial. If they made an error they had to start over at 2043 and received negative feedback from the experimenter. The task instructions were modified from the original protocol such that participants were told to continue subtracting into negative numbers when they reached zero. Given that no participant ever got lower than the 1600s, this prompt provided implicit negative feedback regarding their performance.

During hand immersion, participants were instructed to submerge their non-dominant hand in ice water (2°C) for durations of 60 to 90 seconds. They were not informed how long each immersion trial would last, and the duration alternated with no pattern that would be discernable for the participant. Initially participants were told that withdrawing their hand from the water would result in automatic withdrawal from the study. After a consistent withdrawal rate of 25% in the acute stress condition, we modified our design. In the initial validation studies for the MAST, participants were recruited with posters that advertised a study "examining individuals' resilience to physical and mental challenges" (Smeets et al., 2012). We hypothesized that this language may have led to a natural selection bias such that the participants in the original protocol had higher thresholds for physical and psychological pain. As the purpose of the MAST in our study was to induce stress, we felt that participants who were so uncomfortable that they needed to remove their hand from the ice water prematurely were demonstrating signs of acute stress. Further, a review of their self-reported mood assessments indicated that their ratings were statistically similar to participants who did not withdraw their hand early. This rationale resulted in a modification to the protocol midway through the study. In the modification, participants were allowed to remove their hand early if they could not withstand the pain. However, they were instructed that they were expected to be able to keep their hand in the water for the length of every trial, and that removing it early would count against their performance score. (In reality no score was kept, but giving this instruction ensured that participants would attempt to keep their hand in the water as long as possible, and added to their experience of stress if they "failed" by having to remove it.)

Participants in the low stress condition completed a control version of the MAST. The protocol was identical except for two modifications: instead of performing arithmetic they were instructed to count repeatedly from 1 to 25 using counting numbers (i.e., 1, 2, 3, 4,…) and the water for hand immersion was lukewarm (35–38°C).

In order to assess the efficacy and duration of the stress induction, participants were asked to rate their own valence and arousal on the SAM scales at seven timepoints: before MAST, twice during MAST, once after MAST (subjects did not know that the main task had ended), and after each of the three randomized blocks of robot pose presentations. As a biometric measure, participants also wore a Garmin Heart Rate monitor to record their heart rate over the course of the stress induction and main task.

Robot Pose Identification

All participants gave verbal ratings, stating first the emotional valence and then the arousal level they perceived in each robot pose, using the 9-point Self-Assessment Manikin scales. Participants viewed 25 distinct poses, five each from the categories: negative valence-high arousal, negative valence-low arousal, neutral, positive valence-low arousal, positive valence-high arousal. Each pose was presented three times in block randomized order.

The transition from one pose to the next was masked by the experimenter cueing the participant to close their eyes.

Control measures

Once the stress induction and robot pose identification tasks were complete, participants returned to the table with the computer. As a control, participants rated a validated set of still images of 50 body language poses performed by humans (de Gelder & Van den Stock, 2011). Participants also completed additional questionnaires, including questions about how familiar they felt with the robot across the course of the experiment, and then were debriefed. Results from the questionnaires will be reported in a later chapter.

Study 2 Protocol

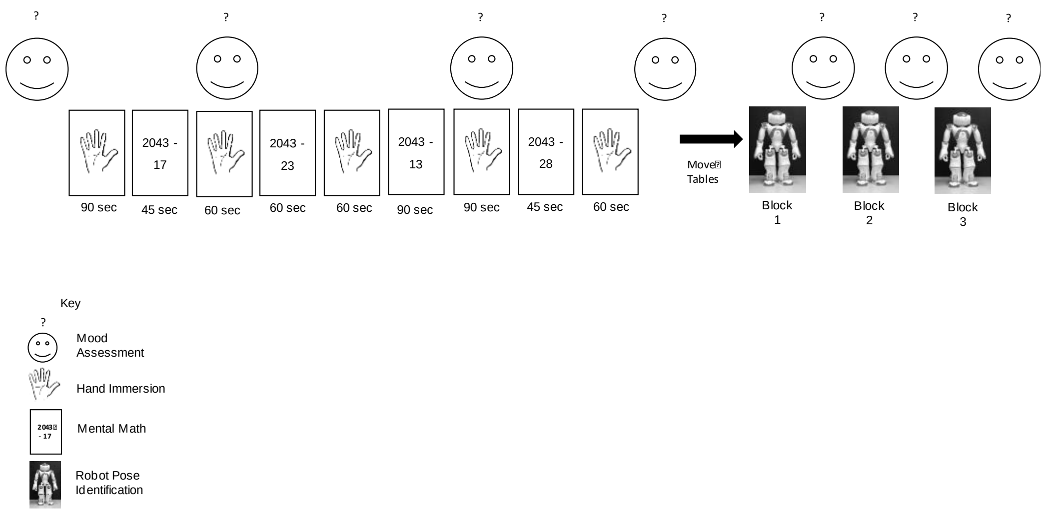


Fig. 4 Protocol for the stress induction (MAST) and robot pose identification (main task). Participants rated their mood prior to beginning the MAST, then twice during the MAST (during hand immersion), and once at the end of the task before being informed that they had completed it. Participants then moved to a different table where they completed the robot pose identification task, rating their mood after each randomized block.

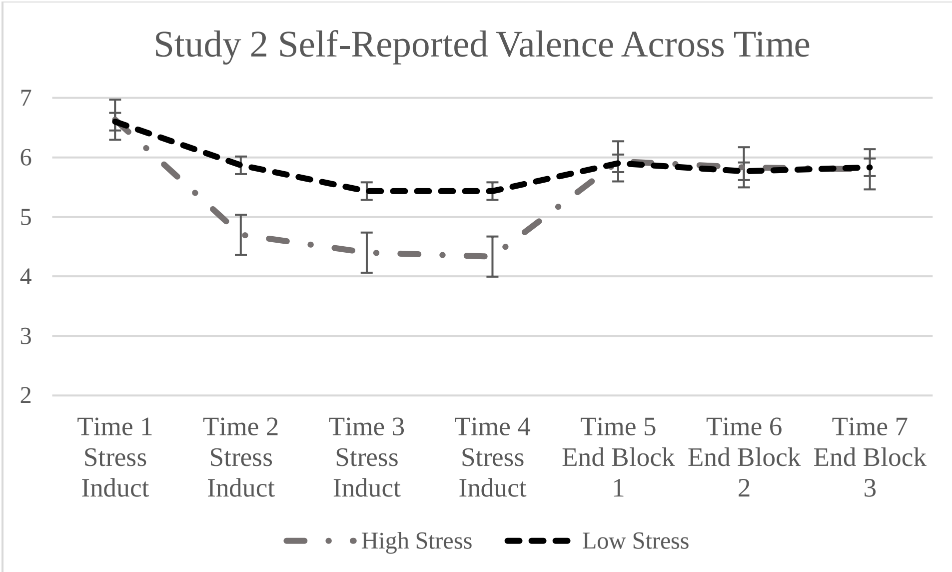
Results

Stress Induction - Self-Ratings of Emotional Valence and Arousal

We began by examining the effects of the stress induction on participants' self-rated emotional valence and then their self-rated arousal across the seven time-points for the two stress conditions.

Looking first at self-rated valence, a 2 (high/low stress) × 7 (time-point) mixed factor analysis of variance (ANOVA) revealed a significant effect of stress condition, $F(1,58)=21.94$, $MSE=4.99$, $p=.04$, Cohen's $d=1.23$, with subjects in the high stress condition reporting an average valence rating of 5.38 and those in the low stress condition reporting an average of 5.83, $M_{\text{diff}}=0.45$, 95% CI[0.27, 0.64]. There was also a significant effect of time-point $F(3.73, 216.24)=41.71$, *MSE*=0.89, *p*<.001 (results reported with Greenhouse-Geisser correction), with a general pattern of decreasing valence across the first four time-points, and a significant time-point × stress interaction $F(3.73,216.24)=9.90$, $MSE=0.89$, $p<.001$ (results reported with Greenhouse-Geisser correction). Within-subject contrasts in the time-point × stress interaction showed a significant linear $F(1,58)=5.84$, $MSE=1.02$, $p=0.02$, $d=0.63$; quadratic $F(1,58)=13.46$, $MSE=.73$, $p=.001$, $d=0.96$; cubic $F(1,58)=27.23$, $MSE=.51$, $p<.001$, $d=1.37$; and 6th order $F(1,58)=7.14$, $MSE=.42$, $p=.01$, $d=0.70$, effect, reflecting a significant decrease in valence from the first (baseline) measurement to the fourth measure (end of stress induction). By the end of the first block of assessing robot poses there was a dramatic upturn for the high stress group, such that there were no significant differences in self-related valence ratings between stress conditions for the fifth, sixth, and seventh time-points.

Looking next at self-rated arousal, a parallel ANOVA showed no main effect of stress condition nor time-point, *F*s<2.0. There was a significant interaction of time-point × stress, $F(3.24,184.84)=3.78$, $MSE=2.47$, $p=.01$ (reported with Greenhouse-Geisser correction), which showed significant quadratic $F(1,57)=5.86$, $MSE=1.65$, $p=.02$, $d=0.64$, and cubic $F(1,57)=10.43$, $MSE=1.14$, $p=.002$, $d=0.86$ effects. This reflected a steeper initial elevation in arousal for the high stress group and a downturn in arousal by the fifth measurement, whereas the level of arousal for the low stress groups showed shallower changes. Figure 5 shows self-ratings for valence and arousal across time and by condition.



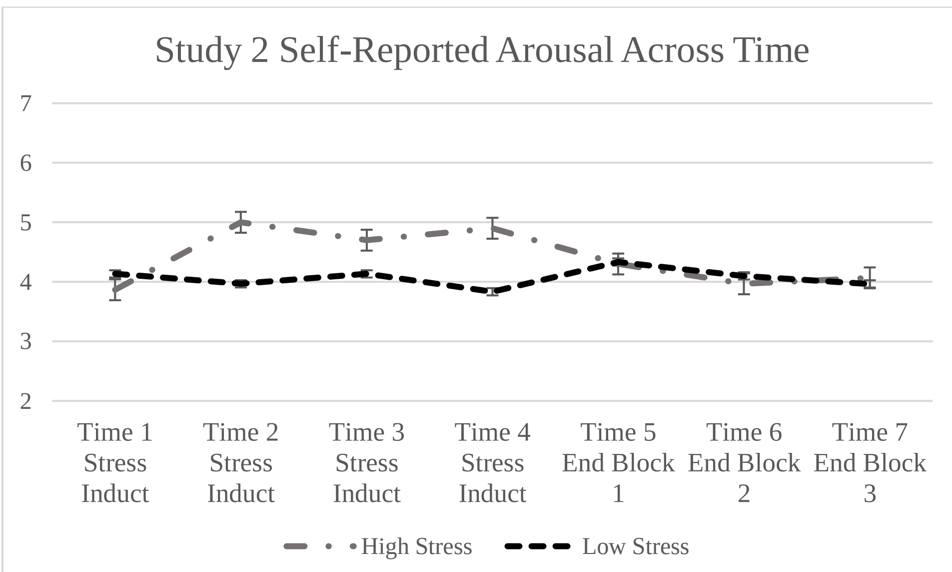


Fig. 5 Participants rated their own mood over the course of the stress induction and main task at each of four time-points. The upper panel depicts participants' self-ratings for valence (1=very negative, 9=very positive) separately by stress condition (high stress vs. low stress). The lower panel shows corresponding self-ratings for arousal (1=very low, 9=very high). Note that for both valence and arousal the range of the y-axis is 2 to 7 to increase visibility.

Stress Induction – Measures of Heart Rate

The output from the Garmin HR heart rate monitor was not automated, necessitating manual transcription and dependence on the Garmin website for the data. The experimenters could not obtain raw data and were required to log into the Garmin website and hover over a line graph to see a pop-up of the HR values. Toggling to the same time-points for each participant was extremely difficult as there were no standardized intervals on the line graphs. Inter-rater reliability was low – at least 30% of all data points that were manually transcribed were inaccurate when compared. Rigorous training of research assistants and the implementation of more precise methods of data entry did not resolve the low reliability. It was not clear whether this was due to inconsistent data output on the Garmin website or simply a result of the sheer mass of data generated from a heart rate sample rate of 60 readings per minute for approximately 60 minutes for each participant. Because we were not able to increase the concordance rate to acceptable levels, we did not include heart rate data in our analyses.

Ratings of Robot Pose Emotionality

Examination of the means shows that the five pose categories [negative valence-high arousal (NH), negative valence-low arousal (NL), neutral (NE), positive valence-low arousal (PL), and positive valence-high arousal (PH)] were perceptually differentiated from one another as expected, for both the emotional valence and arousal dimensions.

We performed separate 2 (high/low stress) × 3 (first/second/third presentation) mixed-factor ANOVAs on the mean valence and mean arousal ratings given for each of the five robot pose categories. We here report the results, organized by effects of stress condition and repeated presentation. The corresponding cell means together with the relevant means from Study 1 and Study 3 are reported in Appendix C.

Effects of stress condition

There were no main effects of stress on valence and arousal ratings of the robot poses for any of the emotion categories, all *F*s<1.0, except NH arousal, $F(1,56)=2.01$, $MSE=1.96$, $p=0.16$, $d=0.38$.

Effects of presentation (first vs. second vs. third presentation)

Whether participants were seeing a specific robot pose for the first, second, or third time influenced valence ratings in the negative and neutral categories. However, the direction of those changes was not consistent across categories. For NH poses, the valence rating decreased from first to second presentation, and increased from second to third, ($M1=4.37$, $M2=4.08$, $M3=4.16$), $F(2,112)=4.31$, $MSE=.32$, $p=.02$, $M_{\text{diff1}}= -.29$, 95% CI[-0.57,-.01], $M_{\text{diff2}}=.08$, 95% CI[0.00,0.16], with significant linear $F(1,56)=4.01$, $MSE=.35$, $p=.02$, $d=0.54$, and quadratic $F(1,56)=4.66$, $MSE=.30$, $p=.04$, $d=0.58$, effects. For NL poses, the valence rating increased consistently from first to third presentation, ($M1=3.31$, $M2=3.56$, $M3=3.81$), $F(2,112)=23.19$, $MSE=.16$, $p<.001$, $M_{\text{diff1}}=.25$, 95% CI[0.15,0.35], $M_{\text{diff2}}=.25$, 95% CI[0.15,0.35], with a significant linear effect,

F(1,56)=39.48, MSE=.19, $p < .001$, d=1.68. For NE poses, there was a significant presentation × stress interaction, F(2,112)=4.04, MSE=.13, p=.02, with a significant linear effect F(1,56)=7.03, MSE=.14, p=.01, d=0.71. Neutral ratings in the low stress group showed an increase across presentations ($M1$=4.98, $M2$=5.21, $M3$=5.30), M_{diff1}=.22, 95% CI[0.06,0.40], M_{diff2}=.09, 95% CI[0.02,0.16]. In comparison, the high stress group showed an increase, then decrease, ($M1$=5.13, $M2$=5.19, $M3$=5.09), M_{diff1}=.06, 95% CI[0.01,0.11], M_{diff2}= -.10, 95% CI[-0.18,-0.02]. There were no effects of presentation on the positive valence categories, Fs<2.2.

For arousal ratings, there were no effects of presentation on negative or neutral categories (Fs<1.2), while both positively valenced categories did show effects. PL arousal ratings increased across presentation (M1=4.57, M2=4.84, M3=4.91), F(1.75,112)=2.21, MSE=.39, p=.006, M_{diff1}=.27, 95% CI[-.09,0.63], M_{diff2}=.07, 95% CI[-0.02,0.16], with a significant linear effect F(1,56)=9.03, MSE=.38, p=.004, d=0.80 (reported with Greenhouse-Geisser correction). For PH ratings of arousal there was also an increase across presentation (M1=6.31, M2=6.58, M3=6.67), F(2,112)=5.97, MSE=.36, p=.003, M_{diff1}=.27, 95% CI[0.48,0.49], M_{diff2}=.09, 95% CI[0.2,0.16], with a significant linear effect F(1,56)=9.58, MSE=.42, p=.003, d=0.83.

Control Task: Assessment of Human Emotional Body Language

There was no evidence of effects of stress condition or robot condition on the human body language poses viewed five minutes after the main task, Fs<2.1.

This indicates that participants in the different conditions possessed comparable abilities for perceiving human emotional body language.

Participant Ratings of Familiarity with Robot

Participants were asked to retrospectively indicate on a five-point scale their level of familiarity with the robot at the beginning, two points during, and the endpoint of the main task. There was a main effect of time, $F(1.84,106.81)=69.39$, $MSE=0.80$, $p<.001$ (reported with Greenhouse-Geisser correction), with significant linear $F(1,58)=98.47$, $MSE=1.02$, $p<.001$, $d=2.61$, and quadratic effects $F(1,58)=4.39$, $MSE=.32$, $p<.04$, $d=0.55$. Participants reported experiencing greater familiarity with the robot by the end of the main task (Start(*T1*)=2.09, End(*T4*)=3.83). There was no main effect of stress, $F<1.0$. Figure 6 illustrates the change in familiarity over time.

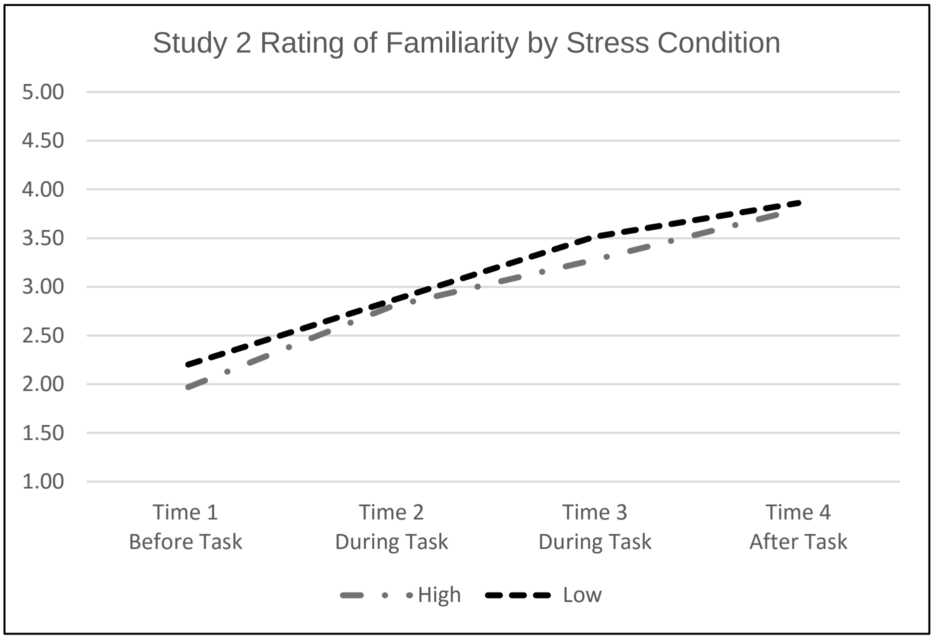


Fig. 6 Participants were asked to retroactively report their degree of familiarity with the robot on a 5-point scale (1=minimal familiarity, 5=maximum familiarity) at four time-points during the main task. There were no significant differences between stress groups on familiarity ratings.

Discussion

Participants in the acute stress condition did report experiencing stress — as shown in both initial decreases in their self-reported emotional valence and initial increases in their self-reported arousal level — but they indicated that the effects had worn off by the end of the first block of the robot identification task. We also attempted to record a biometric measure (heart rate) of stress, but complications with the equipment rendered the data unusable. Between the self-reported data and the lack of biometric measurements, we cannot confirm that participants in the acute stress group did not experience stress while identifying the robot poses. This is because self-report measures may not be accurate; participants may have a desire for privacy and therefore not report their actual valence and arousal, others may tell experimenters what they think the experimenters want to hear, and some may get bored with the activity of self reporting and stop allocating attention to the task.

On the other hand, it is possible that the stress induced by the MAST was not robust enough to last through the first block of robot emotional body language poses. One explanation for the rapid recovery of the high stress group is that the induction was not sufficiently potent to sustain acute stress for the length of the main task. Each block of the main task lasted 15 minutes on average, such that approximately 45 minutes passed between the stress induction and the last pose rating. At the end of the stress induction participants were told that they had completed the task. They were then asked to move from one table to another to complete the main task. Together, the information that the stress induction had ended and the physical disruption of moving could have contributed to the reduction in stress by shifting attention and signaling that the participant's new context would not include similar stressors. Another possible explanation is that the presence of the robot itself reduced the participants' stress level. It may have been an effect of distraction, novelty, or there could be a quality inherent to the robot's presence that resulted in reduced stress.

Although it is not possible to verify with certainty whether the acute stress group was highly stressed during the robot pose identification task, the fact that there were no main effects of stress between conditions, and the close parallels in the self-reported valence and arousal ratings for the two conditions during the later robot pose identification phases, suggests that the stress was not sufficiently long lasting. However, this conclusion presupposes that our hypothesis is true and that there are differences in the ability to perceive emotion between people in high and low acute stress states, which the findings from this study cannot inform.

The effects of repeated robot pose presentation do offer some insight into how experience with robots may impact perception of robot emotional body language. The increase of arousal ratings for positive poses across presentation may be related to the increase of familiarity with the robot that was reported by participants across the two stress conditions. Increased arousal can be interpreted as a sense of increased animacy, which may explain why Lohani et al. (2016) found that engaging with socioemotional robots increases human trust of robots. This finding suggests that experience interacting with robots may lead to more positive feelings towards them as perception of their emotional body language shifts.

While this is an intriguing finding that can inform design of robots for high stress contexts such as healthcare, the lack of differentiation between stress groups rendered it impossible to determine whether or not stress impacts human perception of robot body language, which is a key focus of our research. In order to answer this question, the protocol had to be redesigned to ensure that the effects of stress induction endure over the course of the robot pose identification task.

Chapter 4: Study 3 – Stress Induction and Robot Type

Study 3 – Effects of Stress Induction and Robot Type on Perception of Robot Emotional Body Language

The stress induction in Study 2 was not sufficiently long-lasting to ensure a high level of stress in the acute stress condition while participants completed the robot body language pose identification task. In Study 3 the stress induction and pose identification task were interleaved to prevent substantial diminishment of the induced stress during the key pose-identification task. In addition, a robot type condition was added (physical versus virtual reality) to determine whether virtual reality may be a viable testing tool for robot design. The virtual reality robot was presented in a photorealistic virtual environment that mirrored the lab in which the experiment was performed.

The design of Study 3 closely paralleled Study 2 with the following exceptions: the MAST was modified, the number of presentations of each pose was decreased to two, and a robot (physical/virtual) condition was introduced.

Methods

Participants

Participants were recruited through posters on campus and a university research participation website. To be eligible, participants were required to be between the ages of 18-35 years, in good health, and have lived in the U.S. for at least five years. Participants were offered either extra credit or $15 for a 90-minute session.

Participants were prescreened for medical conditions (cardiac and neurological), anxiety, and depression. Two potential participants were excluded due to BSI score.

We planned for, and tested, 96 participants (*M* age=21.15, *SD*=3.13), with 24 participants (10 male) per condition.

Procedure Summary

Participants gave informed consent, completed screening documents, and were randomly assigned to a condition (high vs. low stress, physical vs. virtual reality robot) (10 minutes). Next, participants completed an interleaved stress induction and robot pose identification task (30 minutes), a human emotional body language pose identification task as a control (5 minutes), and then a series of questionnaires, including the Adult Temperament Questionnaire, attitudes towards robots, and a modified version of the Chronic Stress Inventory (15 minutes). Last, they were debriefed (5 minutes), for a total session time of 65 minutes.

Stimuli

The robot and testing environment remained very similar to Study 2 for the physical robot condition. The one difference is that the large table, where the participant sat to identify the robot poses, was moved 0.60 m toward the center of the testing room. This change was made to accommodate the length of the cable that connected the headset to the computer for the virtual reality condition.

For the virtual robot, three-dimensional (3D) models of the poses were created through Autodesk 123D Catch for PCs and refined in MeshLab 1.3.3. The virtual robot was presented in a virtual testing room environment that was a photorealistic rendition of the room used for the physical robot. The virtual environment was built in Unity 5.3.3f1, with components created in SketchUp 2016. The model for the table was downloaded from the SketchUp 3D warehouse and modified to match the color of the tables in the physical environment. Figure 7 illustrates the view of the participant in the two robot conditions.

Comparison of Physical Laboratory and Virtual Reality Environment

Fig. 7 Photograph of the testing space (left). Screen shot of the virtual reality environment (right). The plastic hand immersion container was not present in the virtual environment. Participants removed the headset when completing the hand immersion and math tasks.

Participants in the virtual reality condition wore the Developer Kit 2 Oculus Rift (DK2) headset when identifying the robot poses (resolution: 960 x 1080 per eye, field of view: 100°, weight: 440g). The computer used for rendering was an ASUS ROG (CPU: Intel Core i7 (6^{th} Gen) 6700/3.4GHz, computer memory: 16GB, graphics: NVIDIA GeForce GTX 970), video memory: 4GB. Audio sounds recorded from the physical robot moving between poses was played on Dell A215 speakers to avoid the burden of additional equipment to put on and remove during the main task. The virtual recreation of the lab space had the same dimensions as the physical environment for all objects. Participants were instructed not to turn their head more than 90° to avoid motion sickness. This instruction also ensured that the participant did not look in the direction of the experimenter, whose body was not modeled in the virtual lab environment.

Robot Pose Identification and Stress Induction (and Control)

In Study 2, analysis of the participants' self-reported valence and arousal levels indicated that participants in the high stress condition had not sustained a mood reflecting stress by the end of the first block of robot poses. To counteract such diminishment of the effects of the stress induction, the hand immersion and math tasks in Study 3 were interleaved with the robot pose identification task to ensure sustained stress for the acute stress condition while the participants responded to the robot body language. (See Chapter 3 for full description of the MAST stress induction task.) Each pose was presented two times in block randomized order. We chose to omit a third presentation to reduce the time intervals between stress inducing trials. In keeping with the modified instructions for the cold water immersion, participants were told that they were expected to keep their hand in the water for the entire trial and that removing it early would result in a performance penalty. With this modification in place, there were no withdrawals from Study 3 during the stress induction. The MAST and robot pose identification tasks were also interleaved for the low stress (control) condition. Figure 8 illustrates the interleaved robot pose identification and stress induction task.

Participant response to the stress induction were assessed on the same nine-point Self Assessment Manikin (SAM), but with a different interval. Participants were asked to assess their own valence and arousal levels at four timepoints: before MAST/robot identification, twice during MAST/robot identification, and once at the end of the MAST/robot identification (participants were not aware that they had completed the task when making the final self assessment). Participants wore the Garmin Heart Rate monitor over the course of the robot identification and stress induction task, in case Garmin changed their policy and enabled access to the raw data. A change in policy did not occur during the course of this dissertation, so heart rate will not be reported for Study 3.

Study 3 Protocol

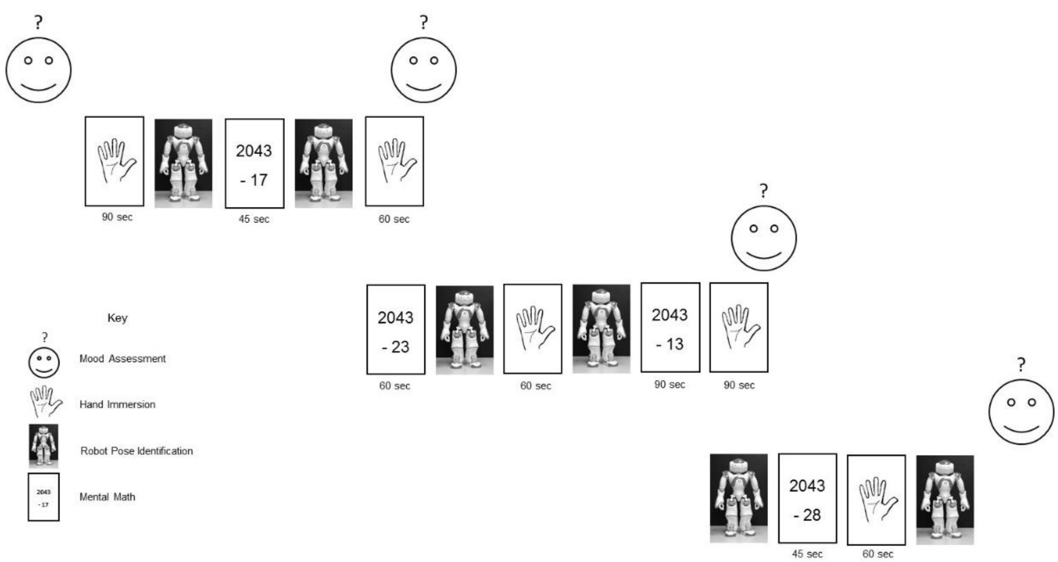


Fig. 8. Main task progression (left to right, top to bottom). The MAST stress induction was interleaved with robot pose identification to ensure sustained stress. All participants received the same ordering, with no discernable pattern to the participant.

Control Measures

As in Study 2, once the main task was completed, participants moved to the large table to identify the validated set of 50 still images of human emotional body language poses (de Gelder & Van den Stock, 2011), and to complete the Adult Temperament Questionnaire, the robot attitudes questionnaire, and the Chronic Stress Inventory on the computer. Results from the questionnaires will be presented in Chapter 5.

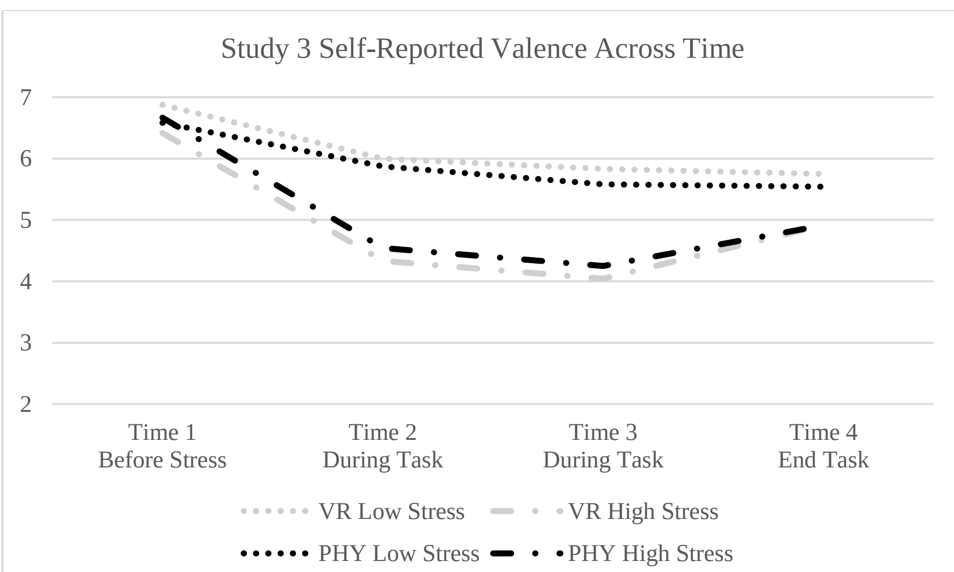
Results

Stress Induction - Self-Ratings of Emotional Valence and Arousal

We began by examining the effects of the stress induction on participants' self-rated emotional valence and then their self-rated arousal across the four time-points for the two robot conditions.

Looking first at self-rated valence, a 2 (high/low stress) × 2 (physical/virtual robot) × 4 (time-point) mixed factor analysis of variance (ANOVA) revealed a large and significant effect of stress condition, $F(1,92)=21.96$, $MSE=4.33$, $p<.001$, Cohen's $d=0.98$, with subjects in the high stress condition reporting an average valence rating of 5.01 and those in the low stress condition reporting an average of 6.01, $M_{diff}=1.00$, 95% CI[.58, 1.42]. There was also a significant effect of time-point $F(2.60, 239.48)=107.96$, $MSE=0.60$, $p<.001$, with a general pattern of decreasing valence across the experimental session, and a significant time-point × stress interaction $F(2.60,239.48)=19.90$, $MSE=0.60$, $p<.001$ (results reported with Greenhouse-Geisser correction). Within-subject contrasts in the time-point × stress interaction showed a significant linear $F(1,92)=5.11$, $MSE=0.67$, $p=0.03$, and pronounced quadratic effect $F(1,92)=46.17$, $MSE=.60$, $p<.001$, $d=1.42$, reflecting a significant decrease in valence from the first (baseline) measurement to the second measure (during the stress induction) with an upturn after the stress induction (fourth measurement).

Looking next at self-rated arousal, a parallel ANOVA showed a significant effect of time-point, $F(2.18,200.48)=4.00$, $MSE=1.87$, $p=.02$. There was no main effect of stress condition and no interaction of robot condition with time, Fs<1.5, and no time × stress interaction, $F(2.18,200.48)=2.18$, $MSE=1.87$, $p=.11$, though within-subject contrasts in the time × stress interaction showed a significant quadratic effect, $F(1,92)=4.42$, $MSE=1.53$, $p=.04$, $d=0.44$. This reflected a steeper initial elevation in arousal for the high stress groups and a downturn in arousal by the fourth measurement, whereas the level of arousal for the low stress groups showed shallower changes. There was no effect of robot condition, $F(1,92)=1.80$, $MSE=7.29$, $p=.18$, and no robot × stress interaction, $F<1$. Figure 9 shows self-ratings for valence and arousal across time and by condition.



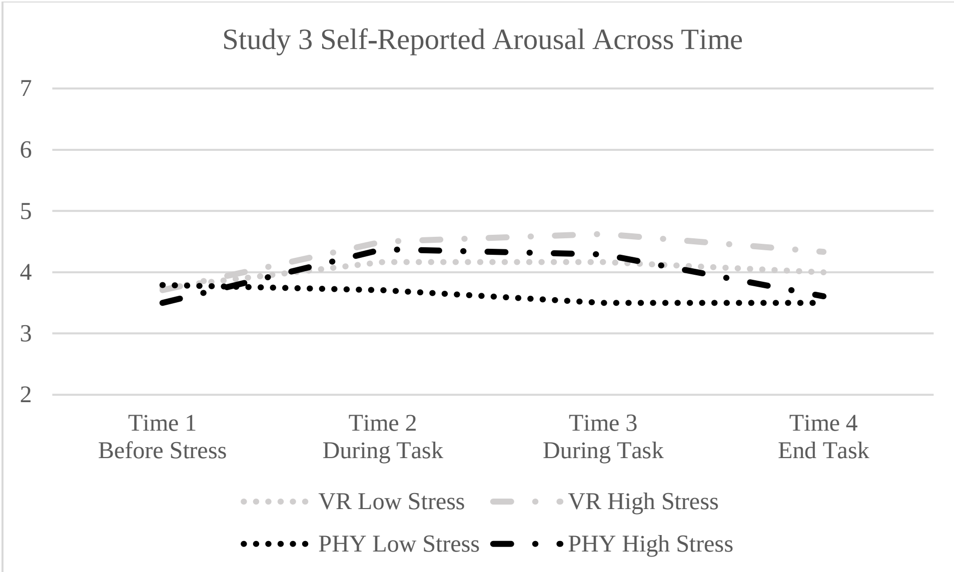


Fig. 9 Participants rated their own mood over the course of the stress induction and main task at each of four time-points. The upper panel depicts participants' self-ratings for valence (1=very negative, 9=very positive) separately by stress and robot condition. The lower panel shows corresponding self-ratings for arousal (1=very low, 9=very high). Note that for both valence and arousal the range of the y-axis is 2 to 7 to increase visibility.

Ratings of Robot Pose Emotionality

Examination of the means shows that the five pose categories [negative valence-high arousal (NH), negative valence-low arousal (NL), neutral (NE), positive valence-low arousal (PL), and positive valence-high arousal (PH)] were perceptually differentiated from one another as expected, for both the emotional valence and arousal dimensions.

We performed separate 2 (high/low stress) × 2 (physical/virtual robot) × 2 (first/second presentation) mixed-factor ANOVAs on the mean valence and mean arousal ratings given for each of the five robot pose categories. We here report the results, organized by effects of stress condition, robot condition, and repeated presentation. The corresponding cell means together with the relevant means from Study 1 are reported in Appendix B.

Effects of stress condition

Stress especially influenced participants' perception of high-arousal robot poses. There was a significant effect of stress condition for NH poses, both for perceived valence, $F(1,92)=4.56$, $MSE=3.46$, $p=.04$, $d=0.45$, and perceived arousal, $F(1,91)=8.03$, $MSE=1.17$, $p=.006$, $d=0.59$. Participants in the high-stress condition perceived lower valence, that is, less positive emotion (mean valence rating of 4.25) than those in the low-stress condition (4.82, $M_{diff=}0.57$, 95% CI[.04, 1.10]) and also perceived lower arousal (6.99) than those in the low-stress condition (7.43) for NH poses ($M_{diff}=0.44$, 95% CI[.13, .75]). For PH poses, there was a similar but weaker effect of stress for perceived valence, $F(1,92)=3.05$, $MSE=1.86$, $p=.08$ (high stress=6.58, low stress=6.93), $d=0.36$ and a significant effect for perceived arousal, $F(1,92)=3.95$, $MSE=2.79$, $p=.05$, $d=0.41$, with high stress participants again perceiving lower arousal in PH poses (6.46) than did participants in the low-stress condition (6.94, $M_{diff}=0.48$, 95% CI[.0009, .96]).

Effects of robot condition

Whether the robot was physically present versus virtually instantiated was influential primarily for robot poses that conveyed positive emotions. There was a significant effect of robot condition on valence ratings for both PL poses, $F(1,91)=7.47$, $MSE=1.47$, $p=.01$, $d=.57$, and PH poses, $F(1,92)=4.25$, $MSE=1.86$, $p=.04$, $d=0.43$. In both cases, participants in the physical robot condition perceived higher (more positive) valence than did those in the virtual robot condition (5.31 vs. 4.83 for PL, M_{diff}=0.48, 95% CI[.13,.83]; 6.96 vs. 6.55 for PH, M_{diff}=0.41, 95% CI[.01,.81]). There was also a significant effect of robot condition on arousal ratings for both PL poses, $F(1,92)=6.72$, $MSE=2.98$, $p=.01$, $d=0.54$, and PH poses, $F(1,92)=10.23$, $MSE=2.79$, $p=.002$, $d=0.67$. In both cases participants in the physical robot condition perceived higher arousal (more animation) than did those in the virtual robot condition (4.48 vs. 3.83 for PL, M_{diff}=0.65, 95% CI[.09,.1.21]; 7.08 vs. 6.31 for PH, M_{diff}=0.77, 95% CI[.29,1.25]).

Effects of presentation (first vs. second presentation)

Whether participants were seeing a specific robot pose for the first or second time predominantly influenced ratings of neutral and low-arousal poses. For neutral poses, there was a mere-exposure-like effect on valence ratings, with higher perceived valence on the second presentation (4.87) than on first presentation (4.70), $F(1,91)=6.22$, $MSE=0.22$, $p=.01$, $d=.52$, M_{diff}=0.17, 95% CI[.03,.31]. A parallel pattern was apparent for arousal ratings, with higher perceived arousal on second presentation (3.77) than on first presentation (3.54), $F(1,92)=5.32$, $MSE=0.47$, $p=.02$, $d=0.48$, M_{diff}=0.23, 95% CI[.03,.43]. Similar patterns were observed for low-arousal poses. Although presentation did not reach significance for perceived valence, $F(1,92)=2.21$, $MSE=0.34$, $p=.14$ (NL) and $F(1,91)=2.08$, $MSE=0.27$, $p=.15$ (PL), it was significant for perceived arousal. There was higher perceived arousal on second than on first presentation for both NL (second=3.91 vs first=3.68), $F(1,92)=6.67$, $MSE=0.38$, $p=.01$, $d=0.54$,

M_{diff}=0.23, 95% CI[.05,.41] and PL (second=4.28 vs first=4.03), F(1,92)=5.28, MSE=0.57, p=.02, d=0.48, M_{diff}=0.25, 95% CI[.03,.47] poses.

Interactions of Stress or Robot Condition with Presentation

For PL poses on valence ratings there was a three-way interaction of presentation × robot × stress, F(1,91)=5.16, MSE=0.27, p=.03. As this pattern was not observed for any other robot-pose conditions, it may be spurious and will not be further discussed.

Control Task: Assessment of Human Emotional Body Language

There were no effects of stress condition or robot condition on the human body language poses viewed five minutes after the main task, Fs<2.1. This indicates that participants in the different conditions possessed comparable abilities for perceiving human body language.

Participant Ratings of Familiarity with Robot

Participants were asked to retrospectively indicate on a five-point scale their level of familiarity with the robot (physical/virtual) at the beginning, two points during, and the end point of the main task. There was a main effect of time, F(1.92,176.94)=126.55, MSE=0.71, p<.001. Participants reported experiencing greater familiarity with the robot (physical/virtual) by the end of the main task (Start($T1$)=1.97, End($T4$)=3.77). There was also a main effect of stress, F(1,92)=7.43, MSE=2.30, p=0.008, d=0.57, with acute-stress participants reporting overall less familiarity (2.71) with the robot than low-stress participants (3.13), M_{diff}=0.42, 95% CI[.11,.73]; for the four time-points, high-stress: $T1$=1.71, *T2*=2.50, *T3*=3.02, *T4*=3.60; low-stress: *T1*=2.23, *T2*=2.88, *T3*=3.48, *T4*=3.94. There were no interactions, *F*s<2.1. Figure 10 shows retrospective ratings of familiarity with the robot over time separately by stress and robot condition.

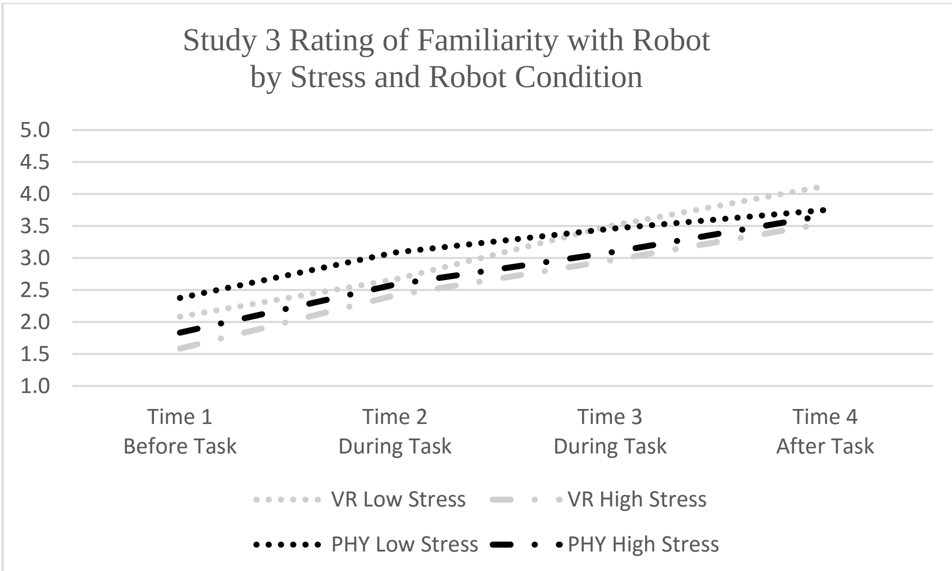


Fig. 10 Participants were asked to retroactively report their degree of familiarity with the robot at four time-points during the main task. Ratings were 1 (minimal familiarity) to 5 (maximum familiarity).

Discussion

Human interaction with robots under diverse conditions, ranging from the mundane to the extreme, will increase exponentially in the coming years. Yet we know little about how acute stress influences our perceptions of robots. Using a novel experimental paradigm, we have shown that acute cognitive and physiological stress influences an individual's judgment of emotion when robot body language conveys a state of high emotional arousal (e.g., excitement, energy, or animation). One aspect of this finding was anticipated: Individuals under high stress rated negative high-arousal poses (e.g., poses that might convey anger) more negatively than individuals under low stress. This appears congruent with past findings that acutely stressed individuals are hypervigilant to threat (Weymar et al., 2012). However, the influence of acute stress on the perception of arousal intensity was counterintuitive. Acutely stressed participants perceived *lower* — not higher — arousal in the high arousal robot poses. This was true regardless of whether the poses were negative or positive in valence.

Why might this be? High stress participants have been found to characterize emotion more rapidly, but less accurately (DeDora et al., 2011), and participants in a negative mood have been found to shift attention away from emotionally negative stimuli (Sanchez et al., 2014). Perhaps participants in the acute stress condition, after perceiving and rating negativity in the pose, rapidly shifted their attention away and so were less attuned to the arousal level. Yet, this does not explain why individuals under acute stress also perceived lower arousal, albeit to a smaller extent, in the positively valenced high-arousal poses. Unless, perhaps, they misperceived the valence and saw what was positive as negative, or conflated the two emotion scales. Although participants were asked to rate valence and arousal on separate scales, the two measures are known to influence one another (Robinson et al., 2004). It is possible that high stress participants viewed the poses as overall more negative (threatening) and felt that the valence rating accounted for this perception.

This interpretation emphasizes adaptive avoidance-related responding — where attention is diverted away from the highly arousing stimuli. Our findings for the arousal ratings appear to contrast with the reported influence of pharmacologically induced cortisol elevation on the perception of emotional arousal. At dosages within the elevated physiological range that might be observed during trauma, a marathon run, or surgery, these pharmacological studies have shown increased — not decreased — arousal ratings in response to neutral or unpleasant pictures (Abercrombie, Kalin, & Davidson, 2005; Wirth, Scherer, Hoks, & Abercrombie, 2011). However, situational factors moderate these effects of cortisol on arousal ratings (Wirth et al., 2011). Summarizing the mixed findings relating to this technique, Putnam and Roelofs (2011) concluded that cortisol administration tends to facilitate active coping behavior when emotional processing is relevant for task performance. They further underscored that whether such active coping involves approach-related versus avoidance-related behaviors appears to be context-dependent. Applying these findings to the current study, after rating emotional valence, participants in the high-stress conditions may have engaged in avoidance-related responding for the high-arousal stimuli, which then attenuated the arousal they perceived in those poses.

Notably, however, this pattern did not hold true for the low-arousal and neutral-arousal poses. For these poses, regardless of stress level, encountering the poses a second time increased participants' perception of arousal. Thus, the participants' ratings of arousal level significantly increased under the influence of one factor (repeated presentation) but decreased under the influence of another factor (stress level). These divergent patterns may be explained by a difference in immediate attention allocation in which high-arousal poses are prioritized for processing.

An alternative, but related, explanation might be that the task of rating the emotional arousal level of a stimulus is more difficult to characterize than emotional valence, and so is more subject to misattribution. This may explain why an effect of mere exposure (repeated presentation) was found in the shift in arousal ratings between pose presentations for low-arousal and neutral poses, with the second presentation rated as higher in arousal (cf. Alter & Oppenheimer, 2009; Reber, Winkielman, & Schwarz, 1998). In other words, participants perceived the robot's low arousal and neutral poses as being more animate in the second presentation. Participants retrospectively reported a linear increase in their sense of familiarity with the robot over the course of the main task. The arousal rating may reflect a misattribution of this sense of familiarity.

We compared participants' ability to perceive emotion of the physically present robot with a virtually rendered robot, situated in the same three-dimensional environment as the participant. There were few differences in our study in how participants rated emotion for the physically present versus virtually present robot. One key exception was that, regardless of whether they were in a high or low-stress condition, for the positively-valenced poses, participants rated the physical robot as significantly more positive and more animate for both low and high-arousal positive pose categories. These results indicate that (for positive poses) virtually-present robots — similar to two-dimensional agents — are perceived as less positive and less arousing than physical robots. However, the otherwise broad similarity in how participants perceived emotion for the non-physically present versus physically present robot suggests that it may be feasible to use virtually-present robots as a "sandbox" for developing and testing alternative robot designs.

Chapter 5: Study 2 and 3 Questionnaires –

Effects of Temperament, Attitudes toward Robots, and Experience with Robots

Study 2 and 3 Combined – Effects of Temperament, Attitudes toward Robots, and Experience on Perception of Robot Emotional Body Language

The previous chapters have primarily reported the impacts of acute stress on perception of robot body language, which was the focus of the stress induction and robot pose identification tasks. However, in high stress settings we hypothesized that perception of robot emotional body language would be influenced by additional factors; namely that traits specific to each human would impact one's ability to communicate and willingness to engage with the robot. To investigate these factors, we asked participants in Studies 2 and 3 to rate a set of human body language poses, and to complete the Adult Temperament Questionnaire Short Form (ATQ) (Evans & Rothbart, 2007), a questionnaire assessing attitudes towards robots, and a modified version of the Chronic Stress Inventory after they completed the stress induction and robot pose identification tasks. The data from these questionnaires allow us to explore what human traits impact people's ability to *communicate* with a robot, and what human traits impact people's willingness to *engage* with a robot.

The set of 50 human body language poses were presented as a control and point of comparison to explore whether similar patterns emerged for ratings of robot and human emotional stimuli. The images were a subset of a larger validated set of human emotional body language, known as BEAST (de Gelder & Van den Stock, 2011). The earlier validation study used a forced-choice rating paradigm. The options presented to participants in that study were: anger, fear, sadness, neutral, and happiness. In our studies we asked participants to rate the stimuli on the same emotional valence and arousal scales as they had used for making their robot pose ratings. We presented 10 images (5 female, 5 male) from each emotion category. The specific images included in Studies 2 and 3 were selected because they had been identified with 100% accuracy in the BEAST validation study.

The ATQ Short Form contained 77 questions covering the same temperament constructs as the full version. We chose the short form due to concerns about the length of the study protocol, especially given that half of the participants underwent acute stress induction during the same session. We selected eight of the sub-constructs for analysis due to their relevance to the emotional body language identification tasks. From the Negative Affect construct we investigated: *fear* (unpleasant affect related to anticipation of distress), *discomfort* (unpleasant affect resulting from the sensory quality of stimulation), and *sadness* (unpleasant affect and lowered mood and energy related to object or person loss, disappointment, and exposure to suffering). From the Extraversion/Surgency construct we analyzed: *sociability* (enjoyment derived from social interaction and being in the presence of others), *positive affect* (latency, threshold, intensity, duration, and frequency of experiencing pleasure), and *high intensity pleasure* (pleasure related to situations involving high stimulus intensity, rate, complexity, novelty, and incongruity). From the Orienting Sensitivity construct we investigated: *affective perceptual sensitivity* (spontaneous emotional cognitive content associated with low intensity stimuli) and *associative sensitivity* (spontaneous cognitive content that is not related to standard associations with the environment). We did not analyze any sub-contructs in Affiliativeness nor Effortful Control.

The robot attitude questionnaire was comprised of multiple existing tools that have been validated in previous research on robots and a series of original questions. We utilized three Godspeed scales (each scale is comprised of five dichotomous word pairings with a five-point scale): *anthropomorphism* (fake-natural, machinelike-humanlike, unconscious-conscious, artificial-lifelike, moving rigidly-moving elegantly), *likeability* (dislike-like, unfriendly-friendly, unkind-kind, unpleasant-pleasant, awful-nice), and *perceived intelligence* (incompetent-competent, ignorant-knowledgeable, irresponsible-responsible, unintelligent-intelligent, foolish-sensible) (Bartneck, Kulić, Croft, & Zoghbi, 2009). Selected questions from the Negative Attitudes about Robots (NARS) included: "I feel that in the future robots will be commonplace in society," "I feel that if I depend on robots too much something bad might happen," "If robots developed into living beings something bad might happen," and "I am concerned that robots would be a bad influence on children" (Nomura, Suzuki, Kanda, & Kato, 2006). To gauge familiarity with the robot over the course of the study we utilized a modified version of the Baddoura, Venture, & Matsukata (2012) questionnaire. The questions that we generated specifically for this series of studies assessed participants' concern for the physical and psychological wellbeing of the robot, comfort level in engaging with robots in healthcare settings and previous experience with robots and virtual reality. The full text for the robot attitude question can be found in Appendix D.

Results

In order to increase the power to detect small effects (to inform future directions for research), we combined the data sets from Study 2 and Study 3. However, there were substantitive differences between the two protocols, namely: stress induction before the main task in Study 2 and interleaved in Study 3, and the addition of a virtual robot in Study 3. To account for any effects of the different protocols, we also analyzed any significant findings from the combined data set separately for Study 2 and Study 3 to determine the strength of the relationship. All analyses reported in this chapter are correlations calculated with Goodman and Kruskal's gamma, which is a conservative measure for ordinally ranked variables. To characterize the strength of the relationship and especially the consistency across the two studies, we pre-defined a difference of .03 or less between gammas for the same measure on Study 2 compared to Study 3 to indicate strong directionality, .06 or less as indicating moderate directionality, and a difference of more than .06 as weak directionality.

Robot Pose Identification and Human Body Language Correlation

As a control, participants viewed a validated set of human body language poses after completing the robot pose identification task. For both studies, analysis of variance tests confirmed that all conditions within the respective study performed similarly, suggesting that all groups were equally capable of distinguishing emotional human body language poses. There was an effect of study for the sadness valence rating, $p = .04$, with participants in Study 3 identifying the human poses as lower in valence (M=2.58) than Study 2 (M=2.97). All other measures were statistically similar.

The set of human body language poses was generated through a forced-choice task, with anger (AN), fear (FE), sadness (SA), neutral (NE), and happiness (HA) as the emotion category options. In order to investigate whether there were correlations between ratings on the robot and human body language poses, we paired the categories in the following way: anger and negative valence-high arousal, fear and negative valence-high arousal, sadness and negative valence-low arousal, neutral and neutral, happiness and positive valence-high arousal. This left us without a pair for positive valence-low arousal. In order to compare the human body pose judgments with the robot body pose judgments under parallel conditions, we only analyzed the values for the first presentation of the robot pose identification task as participants only rated each human body pose once.

On valence, all categories were significantly correlated except neutral: AN:NH1 (G=.14, p=.02), with weak directionality between studies; FE:NH1 (G=.15, p=.006), with weak directionality; SA:NL1 (G=.32, $p < .001$), with strong directionality; HA:PH1 (G=.22, p<.001), with weak directionality. For arousal, all categories were significantly correlated: AN:NH1 (G=.23, $p < .001$), moderate directionality; FE:NH1 (G=.19, p=.003), weak directionality; SA:NL1 (G=.18, $p < .006$), weak directionality; NE:NE1 (G=.39, $p < .001$), weak directionality; HA:PH1 (G=.49, p<.001), strong directionality. These results suggest that there was a correlation between the ratings participants gave for the robot and human emotional body language poses. The correlations are quite strong, especially given that the robot was viewed in three dimensions and the human stimuli were viewed in two dimensions. Table 3 shows the gamma and *p*-values for the combined and individual studies that were significantly correlated for human and robot pose ratings.

Correlation of Robot and Human Emotional Body Language

Study	Human Emotion	Robot Emotion	Valence Gamma	*P*-Value	Study Directionality
Combined	Anger	NH1	0.14	0.02*	Weak
Study 2	Anger	NH1	0.30	.005**	
Study 3	Anger	NH1	0.08	0.31	
Combined	Fear	NH1	0.15	.006**	Weak
Study 2	Fear	NH1	0.30	.002**	
Study 3	Fear	NH1	0.10	0.14	
Combined	Sadness	NL1	0.32	<.001**	Strong
Study 2	Sadness	NL1	0.28	.002**	
Study 3	Sadness	NL1	0.29	<.001**	
Combined	Happiness	PH1	0.22	<.001**	Weak
Study 2	Happiness	PH1	0.25	.01*	
Study 3	Happiness	PH1	0.18	.01*	

Study	Human Emotion	Robot Emotion	Arousal Gamma	*P*-Value	Study Directionality
Combined	Anger	NH1	0.23	< .001**	Moderate
Study 2	Anger	NH1	0.25	.01*	
Study 3	Anger	NH1	0.20	.02*	
Combined	Fear	NH1	0.19	.003**	Weak
Study 2	Fear	NH1	0.33	.001**	
Study 3	Fear	NH1	0.13	0.12	
Combined	Sadness	NL1	0.18	.006**	Weak
Study 2	Sadness	NL1	0.12	0.28	
Study 3	Sadness	NL1	0.19	.01*	
Combined	Neutral	NE1	0.39	< .001**	Weak
Study 2	Neutral	NE1	0.23	.02*	
Study 3	Neutral	NE1	0.47	< .001**	
Combined	Happiness	PH1	0.49	< .001**	Strong
Study 2	Happiness	PH1	0.50	<.001**	
Study 3	Happiness	PH1	0.50	<.001**	

Tab. 3 Significant correlations between ratings of human and robot body language poses (only first presentation for robot poses as human stimuli were only presented once). * $p \leq .05$, ** $p < .01$, study directionality between Study 2 and Study 3 is considered weak when the difference in the magnitude of the gamma correlations for the two studies is substantial, with difference in $G > .06$, moderate for $G \leq .06$, and strong for $G \leq .03$.

Robot Pose Identification

Temperament

Goodman and Kruskal's gamma was run to determine the correlation between temperament and ratings for robot emotional body language poses. Of the subgroups on the Adult Temperament Questionnaire, we pre-selected eight that we hypothesized may influence perception of robot emotion: fear, sadness, discomfort, positive affect, high intensity pleasure, affective perceptual sensitivity, and associative sensitivity.

Analyzing the combined data set for valence ratings, we found small but significant correlations between positive affect and NE2 (G=.14, p=.02), PH1

(*G*=.12, *p*=.05), and PH2 (*G*=.13, *p*=.02), indicating that on the second presentation of the neutral robot poses and both presentations of the positive valence-high arousal poses, participants who scored more highly on the ATQ subscale of positive affect, rated the valence of those poses more highly. However, when comparing directionality between Study 2 and Study 3 results, it was found that all differences were *G*>.06, indicating weak directionality.

Analyzing the combined data set for arousal ratings, we found small but significant correlations between fear and NL1 (*G*=.12, p=.04) and NL2 (*G*=.12, p=.04), indicating that the higher participants scored on the fear subscale, the more aroused they perceived the robot to be when expressing negative valence-low arousal poses. However, as with the valence findings, when Study 2 and Study 3 were examined for directionality, it was found that all differences were *G*>.06, signaling weak directionality. Table 4 shows the gamma and *p*-values for the combined and individual studies that showed significant correlation between temperament and the robot poses.

Correlation of Temperament and Robot Emotional Body Language Poses

Study	Emotion	Personality	Valence Gamma	*P*-Value	Study Directionality
Combined	NE2	Positive	0.14	0.02*	Weak
Study 2	NE3	Positive	0.06	0.58	
Study 3	NE4	Positive	0.19	0.01*	
Combined	PH1	Positive	0.12	0.05*	Weak
Study 2	PH1	Positive	0.05	0.68	
Study 3	PH1	Positive	0.14	0.06	
Combined	PH2	Positive	0.13	0.02*	Weak
Study 2	PH2	Positive	0.07	0.51	
Study 3	PH2	Positive	0.16	0.03*	

Study	Emotion	Personality	Arousal Gamma	*P*-Value	Study Directionality
Combined	NL1	Fear	0.12	0.04*	
Study 2	NL1	Fear	0.06	0.54	Weak
Study 3	NL1	Fear	0.15	0.03*	
Combined	NL2	Fear	0.12	0.04*	
Study 2	NL2	Fear	0.04	0.70	Weak
Study 3	NL2	Fear	0.19	0.02*	
Combined	PL2	Discomfort	0.11	0.051	
Study 2	PL2	Discomfort	0.23	0.02*	Weak
Study 3	PL2	Discomfort	0.05	0.48	

Tab. 4 Significant correlations between scores on selected Adult Temperament Questionnaire Short Form subscales (fear, sadness, discomfort, sociability, positive affect, high intensity pleasure, affective perceptual sensitivity, associative sensitivity) and the ratings for the robot emotional body language poses. * $p \leq .05$, ** $p < .01$, study directionality between Study 2 and Study 3 is considered weak when the difference in the magnitude of the gamma correlations for the two studies is substantial, with difference in $G > .06$, moderate for $G \leq .06$, and strong for $G \leq .03$.

The ratings for human body language poses showed fewer correlations with temperament scores. On valence, there was only one significant Goodman and Kruskal correlation, between affective perceptual sensitivity and anger (G= -.17, p=.005), indicating that participants who score higher on emotional sensitivity rate human anger poses as more negative. Study directionality was strong. Those with higher discomfort scores rated the angry human poses as more aroused (G= .19, p=.001), suggesting that those who experience more discomfort perceived the angry human poses as more animated. Directionality was weak. High intensity pleasure was correlated with lower arousal scores for neutral poses (G= -.12, p=.03), suggesting that those who take pleasure in intense sensation perceived the neutral human poses as less aroused. Table 5 shows the gamma and *p*-values for the combined and individual studies that showed significant correlation between temperament and human body language poses.

Correlation of Temperament and Human Emotional Body Language

Valence

Study	Emotion	Temperament	Gamma	*P*-Value	Study Directionality
Combined	Anger	Affective Perceptual Sensitivity	-0.17	.005**	
Study 2	Anger	Affective Perceptual Sensitivity	-0.16	.08	Strong
Study 3	Anger	Affective Perceptual Sensitivity	-0.16	.05	

Arousal

Study	Emotion	Temperament	Gamma	*P*-Value	Study Directionality
Combined	Anger	Discomfort	0.19	.001**	
Study 2	Anger	Discomfort	0.28	.001**	Weak
Study 3	Anger	Discomfort	0.12	.07	
Combined	Neutral	High Intensity Pleasure	-0.12	.03*	
Study 2	Neutral	High Intensity Pleasure	-0.16	.09	Moderate
Study 3	Neutral	High Intensity Pleasure	-0.12	.07	

Arousal

Study	Emotion	Temperament	Gamma	*P*-Value	Study Directionality
Combined	Anger	Discomfort	0.19	.001**	
Study 2	Anger	Discomfort	0.28	.001**	Weak
Study 3	Anger	Discomfort	0.12	.07	
Combined	Neutral	High Intensity Pleasure	-0.12	.03*	
Study 2	Neutral	High Intensity Pleasure	-0.16	.09	Moderate
Study 3	Neutral	High Intensity Pleasure	-0.12	.07	

Tab. 5 Significant correlations between scores on selected Adult Temperament Questionnaire Short Form subscales (fear, sadness, discomfort, sociability, positive affect, high intensity pleasure, affective perceptual sensitivity, associative sensitivity) and the ratings for the human emotional body language poses. * $p \leq .05$, ** $p < .01$, study directionality between Study 2 and Study 3 is considered weak when the difference in the magnitude of the gamma correlations for the two studies is substantial, with difference in $G > .06$, moderate for $G \leq .06$, and strong for $G \leq .03$.

Robot attitudes

Correlations between participant attitudes about robots and their ratings of the robot poses were explored. As the questionnaire was comprised of a mix of validated and original questions, we first grouped the questions and calculated their correlations to generate the following categories: anthropomorphism (Godspeed), likeability (Godspeed), perceived intelligence (Godspeed), protectiveness towards robot (original questions assessing participant concern for robot physical and psychological wellbeing), healthcare (original questions assessing participant willingness to engage with robots in healthcare settings), negativity (NARS), and prior experience (original questions assessing prior experience and familiarity with robots). Responses to one question (unconscious – conscious) were removed from the anthropomorphism category due to low correlation with the remainder of the questions within that category.

An analysis of the combined valence results found a small but significant negative correlation between prior experience and NH1 (G= -.15, p=.03) and NH2 (G= -.14, p=.04), suggesting that participants with more experience with robots rated the negative valence-high arousal poses as more negatively valenced than did their less experienced counterparts. However, a comparison of gamma values for Study 2 and Study 3 found weak directionality for both measures (G<.06).

There were small but significant positive correlations on valence for anthropomorphism and NE1 (G=.18, p=.005), PL1 (G=.15, p=.01), and PL2 (G= .10, p=.01), indicating that participants who perceived the robot as more human-like rated it higher on valence in the positive valence-low arousal (and neutral) category. A comparison of directionality found that there was a moderate ($G \leq .06$) relationship for NE1 and PL1 and a strong ($G \leq .03$) relationship for PL2.

For protectiveness, there were small but significant negative correlations in the combined data set for NL1 ($G = -.16$, $p = .01$), NE1 ($G = -.17$, $p = .006$), PL1 ($G = -.18$, $p = .006$), PH2 ($G = -.13$, $p = .05$), suggesting that participants who felt more protective of the robot perceived the robot's valence to be lower for all categories except for negative valence-high arousal. An analysis of the separate studies showed weak directionality for NE1 and PH2, and a moderate directionality for NL1 and PL1.

On measures of arousal, a small but significant correlation was found in the combined data set for anthropomorphism and NE1 ($G = .14$, $p = .02$) and PL1 ($G = .15$, $p = .02$), indicating that participants who perceived the robot as more human-like also perceived the robot's arousal to be higher in the positive valence-low arousal and neutral categories. Directionality between Study 2 and Study 3 was weak for NE1 and moderate for PL1.

There was a small but significant positive correlation on arousal ratings between negativity and NH2 ($G = .13$, $p = .03$), and a trend for NE2 ($G = .12$, $p = .054$), and PH2 ($G = .13$, $p = .054$), indicating that participants who felt more negative towards robots perceived the arousal of the high arousal and neutral poses to be higher on the second presentation. An analysis of the directionality between studies found a strong relationship for NH2 and PH2, and a weak relationship for NE2.

Arousal ratings for NH2 and PH2, as well as NE1, also negatively correlated with participants' level of experience with robots; NH2 (G= -.19, p=.005), NE1 (G= -.17, p=.007), PH2 (G= -.13, p=.05), indicating that participants with more experience perceive less arousal in high arousal poses on repeated presentation. Directionality between studies was moderate for NH2 and weak for NE1 and PH2.

Table 6 shows the gamma and p-values by study for the robot attitudes and robot pose correlations discussed here.

Correlation of Robot Attitudes and Robot Emotional Body Language

			Valence		
Study	**Emotion**	**Attitude**	**Gamma**	***P*-Value**	**Study Directionality**
Combined	NH1	Experience	-0.15	.03*	Weak
Study 2	NH1	Experience	-0.03	.81	
Study 3	NH1	Experience	-0.21	.02*	
Combined	NH2	Experience	-0.14	.04*	Weak
Study 2	NH2	Experience	-0.004	.97	
Study 3	NH2	Experience	-0.21	.02*	
Combined	NE1	Anthropomorphism	0.18	.005**	Moderate
Study 2	NE1	Anthropomorphism	0.18	.10	
Study 3	NE1	Anthropomorphism	0.12	.13	
Combined	PL1	Anthropomorphism	0.15	.01*	Moderate
Study 2	PL1	Anthropomorphism	0.06	.54	
Study 3	PL1	Anthropomorphism	0.12	.14	
Combined	PL2	Anthropomorphism	0.16	.01*	Strong
Study 2	PL2	Anthropomorphism	0.11	.26	
Study 3	PL2	Anthropomorphism	0.11	.19	
Combined	NL1	Protective	-0.16	.01*	Moderate
Study 2	NL1	Protective	-0.09	.39	
Study 3	NL1	Protective	-0.15	.08	
Combined	NE1	Protective	-0.17	.006**	Weak
Study 2	NE1	Protective	-0.01	.90	
Study 3	NE1	Protective	-0.21	.009**	
Combined	PL1	Protective	-0.18	.006**	Moderate
Study 2	PL1	Protective	-0.12	.25	
Study 3	PL1	Protective	-0.18	.02*	
Combined	PH2	Protective	-0.13	.05*	Weak
Study 2	PH2	Protective	-0.02	.81	
Study 3	PH2	Protective	-0.15	.09	

Arousal

Study	Emotion	Attitude	Gamma	*P*-Value	Study Directionality
Combined	NE1	Anthropomorphism	0.14	.02*	
Study 2	NE1	Anthropomorphism	0.17	.08*	Weak
Study 3	NE1	Anthropomorphism	0.06	.44	
Combined	PL1	Anthropomorphism	0.15	.02*	
Study 2	PL1	Anthropomorphism	0.16	.15	Moderate
Study 3	PL1	Anthropomorphism	0.12	.15	
Combined	NH2	Negative	0.13	.03*	
Study 2	NH2	Negative	0.12	.27	Strong
Study 3	NH2	Negative	0.14	.07*	
Combined	NE2	Negative	0.12	.05	
Study 2	NE2	Negative	0.21	.04*	Weak
Study 3	NE2	Negative	0.09	.29	
Combined	PH2	Negative	0.13	.05	
Study 2	PH2	Negative	0.1	.35	Strong
Study 3	PH2	Negative	0.13	.11	
Combined	NH2	Experience	-0.19	.005**	
Study 2	NH2	Experience	-0.22	.05*	Moderate
Study 3	NH2	Experience	-0.17	.05*	
Combined	NE1	Experience	-0.17	.007**	
Study 2	NE1	Experience	-0.34	<.001**	Weak
Study 3	NE1	Experience	-0.05	.57	
Combined	PH2	Experience	-0.13	.05*	
Study 2	PH2	Experience	-0.19	.07	Weak
Study 3	PH2	Experience	-0.11	.20	

Tab. 6 Significant correlations between scores on robot attitude subscales (anthropomorphism, likeability, perceived intelligence, protectiveness towards the robot, healthcare, negativity, experience with the robot) and the ratings for the robot emotional body language poses. * $p \leq .05$, ** $p < .01$, study directionality between Study 2 and Study 3 is considered weak when the difference in the magnitude of the gamma correlations for the two studies is substantial, with difference in $G > .06$, moderate for $G \leq .06$, and strong for $G \leq .03$.

Analyses of the correlation between robot attitudes and human emotional body language ratings revealed a small but significant association on valence for protectiveness and anger (G= -.14, p=.03) for the combined data set, suggesting that those who feel more protective towards the robot see angry human body poses as more negative. There was a weak directional relationship between Studies 2 and 3. Experience and fear were also correlated on valence (G= -.14, p=.03), with strong directional relationships between the separate studies, indicating that participants with more experience interacting with robots see human fearful poses as more scared. Anthropomorphism was correlated with sadness SA (G= .12, p=.05) and NE (G= .13, p=.04) human poses on the combined data set, suggesting that those who see the robot as more human-like perceive human sad and neutral poses as more positive than their counterparts. There was a strong directional relationship for anthropomorphism and sadness, with a weak relationship for neutral human poses.

For arousal, there were three small but significant correlations between attitudes about robots and human body language poses. Anger and anthropomorphism were negatively correlated (G= -.16, p=.008) suggesting that the more human-like the participant perceived the robot to be, the lower in arousal they perceived the angry human poses to be. Directionality was weak. Neutral and experience were also inversely related (G= -.17, p=.01), such that participants with more experience with robots perceived human neutral poses to be less activated. Directionality was weak. Happiness and negativity were positively correlated (G= .12, p=.05), indicating that those who feel that robots will have a negative impact on society perceive happy human poses to be more activated. Table 7 shows the gamma and p-values by study for the robot attitudes and human body pose correlations discussed here.

Correlation of Robot Attitudes and Human Emotional Body Language

			Valence		
Study	**Emotion**	**Attitude**	**Gamma**	***P*-Value**	**Study Directionality**
Combined	Anger	Protective	-0.14	.03*	Weak
Study 2	Anger	Protective	-0.04	.74	
Study 3	Anger	Protective	-0.20	.01*	
Combined	Fear	Experience	-0.14	.03*	Strong
Study 2	Fear	Experience	-0.14	.15	
Study 3	Fear	Experience	-0.12	.13	
Combined	Sadness	Anthropomorphism	0.12	.05*	Strong
Study 2	Sadness	Anthropomorphism	0.09	.36	
Study 3	Sadness	Anthropomorphism	0.09	.23	
Combined	Neutral	Anthropomorphism	0.13	.04*	Weak
Study 2	Neutral	Anthropomorphism	0.23	.02*	
Study 3	Neutral	Anthropomorphism	0.07	.46	

			Arousal		
Study	**Emotion**	**Attitude**	**Gamma**	***P*- Value**	**Study Directionality**
Combined	Anger	Anthropomorphism	-0.16	.008**	Weak
Study 2	Anger	Anthropomorphism	-0.10	.36	
Study 3	Anger	Anthropomorphism	-0.20	.01*	
Combined	Neutral	Experience	-0.17	.01*	Weak
Study 2	Neutral	Experience	-0.33	.003**	
Study 3	Neutral	Experience	-0.09	.30	
Combined	Happiness	Negativity	0.12	.05*	Moderate
Study 2	Happiness	Negativity	0.09	.36	
Study 3	Happiness	Negativity	0.15	.07	

Tab. 7 Significant correlations between scores on robot attitude subscales (anthropomorphism, likeability, perceived intelligence, protectiveness towards the robot, healthcare, negativity, experience with the robot) and the ratings for the human emotional body language poses. * $p \leq .05$, ** $p < .01$, study directionality between Study 2 and Study 3 is considered weak when the difference in the magnitude of the gamma correlations for the two studies is substantial, with difference in $G > .06$, moderate for $G \leq .06$, and strong for $G \leq .03$.

Chronic Stress

There was only one significant correlation between chronic stress and robot pose ratings for the combined data set, NE2 (G= .14, p=.02), with weak directionality. There were no significant correlations between chronic stress and the human poses.

Correlation Among Categories

Personality compared with robot attitudes

We explored Goodman and Kruskal's gamma to determine whether there were any correlations between participants' attitudes towards robots and their temperament scores. Anthropomorphism was negatively correlated with discomfort, (G= -.15, p=.01), with weak directionality, and positively correlated with positive affect, (G= .15, p=.01), also with weak directionality. This suggests that those who experience more discomfort also perceive the robot to be less human-like, and that those with higher trait positive affect tend to view the robot as more human-like. Positive affect was also correlated with likeability (G= .26, p<.001) with strong directionality, and perceived intelligence (G= .16, p=.009) with strong directionality, such that those participants who scored higher on trait positive affect tended to rate the robot as more likeable and more intelligent.

Participants who were more protective of the robot tended to score higher on the subscales of fear (G= .17, p=.003) with weak directionality, and sadness (G= .16, p=.009) with weak directionality. In comparison, participants scoring higher on high intensity pleasure tended to be less protective of the robot (G= -.17, p=.009). Thus, those who experience more fear and sadness perceive the robot as more vulnerable to verbal and physical harassment and more likely to have feelings. Those who experience pleasure from high intensity sensations tend to feel less protective toward the study robot.

Negativity was correlated with fear (*G*= .13, *p*=.04) and discomfort (*G*= .17, *p*=.01), both with strong directionality, indicating that those with a negative outlook on robots' role in society also tend to experience more fear and discomfort. Table 8 shows significant correlations between temperament and robot attitudes for the combined and separate data sets.

Correlation of Temperament and Robot Attitudes

Study	Robot Attitude	Temperament	Gamma	*P*-Value	Study Directionality
Combined	Anthropomorphism	Discomfort	-.15	.01*	Weak
Study 2	Anthropomorphism	Discomfort	-.31	.001**	
Study 3	Anthropomorphism	Discomfort	-.05	.48	
Combined	Anthropomorphism	Positive Affect	.15	.01	Weak
Study 2	Anthropomorphism	Positive Affect	.26	.007**	
Study 3	Anthropomorphism	Positive Affect	.06	.45	
Combined	Likeability	Positive Affect	.26	.00**	Strong
Study 2	Likeability	Positive Affect	.24	.003**	
Study 3	Likeability	Positive Affect	.25	<.001**	
Combined	Perceived Intelligence	Positive Affect	.16	.009**	Strong
Study 2	Perceived Intelligence	Positive Affect	.18	.06	
Study 3	Perceived Intelligence	Positive Affect	.15	.07	
Combined	Protective	Fear	.17	.003*	Weak
Study 2	Protective	Fear	.06	.54	
Study 3	Protective	Fear	.22	.004**	
Combined	Protective	Sadness	.16	.009**	Weak
Study 2	Protective	Sadness	.00	.99	
Study 3	Protective	Sadness	.22	.005**	
Combined	Protective	High Intensity Pleasure	-.17	.009**	Weak
Study 2	Protective	High Intensity Pleasure	-.2	.04	
Study 3	Protective	High Intensity Pleasure	-.15	.09	
Combined	Negativity	Fear	.13	.04*	Strong
Study 2	Negativity	Fear	.14	.20	
Study 3	Negativity	Fear	.11	.12	
Combined	Negativity	Discomfort	.17	.01*	Strong
Study 2	Negativity	Discomfort	.17	.16	
Study 3	Negativity	Discomfort	.17	.02	

Tab. 8 Significant correlations between scores on selected Adult Temperament Questionnaire Short Form subscales (fear, sadness, discomfort, sociability, positive affect, high intensity pleasure, affective perception sensitivity, associative sensitivity) and robot attitude subscales (anthropomorphism, likeability, perceived intelligence, protectiveness towards the robot, healthcare, negativity, experience with the robot). * $p \leq .05$, ** $p < .01$, study directionality between Study 2 and Study 3 is considered weak when the difference in the magnitude of the gamma correlations for the two studies is substantial, with difference in $G > .06$, moderate for $G \leq .06$, and strong for $G \leq .03$.

Robot attitudes compared with other robot attitudes

We also explored the associations between the robot subscales. Anthropomorphism was significantly correlated with likeability ($G = .32$, $p < .001$) with strong directionality, and perceived intelligence ($G = .22$, $p = .001$) and healthcare ($G = .23$, $p < .001$), both with weak directionality, suggesting that participants who perceived the robot to be more human-like also found it to be more likeable, intelligent, and were more likely to be willing to engage with it in healthcare contexts.

Likeability was correlated with anthropomorphism (see above results), perceived intelligence ($G = .24$, $p < .001$), protectiveness ($G = .24$, $p < .001$), and healthcare ($G = .16$, $p = .01$), all with weak directionality, indicating that the more likeable a participant found the robot, the more human-like, intelligent, vulnerable, and potentially helpful in healthcare contexts they also perceived it to be.

Perceived intelligence was correlated with anthropomorphism (see above results), likeability (see above results), and protectiveness ($G = .17$, $p = .02$) with weak directionality, suggesting that a robot that is perceived as more intelligent is also seen as more human-like, likeable, and vulnerable. Unlike anthropomorphism and likeability, perceived intelligence was not significantly correlated with engaging with robots in healthcare contexts.

Protectiveness was significantly correlated with likeability (see above results), perceived intelligence (see above results), and healthcare (G= .16, p<.009) with moderate directionality, suggesting that when participants view the robot as more vulnerable, they also find it more likeable, intelligent, and are more likely to want to engage with it in healthcare contexts.

Healthcare was significantly correlated with anthropomorphism (see above results), likeability (see above results), protectiveness (see above results), experience (G= .20, p<.006) with weak directionality, and inversely correlated with negativity (G= -.28, p<.001) with moderate directionality. This suggests that participants who have more experience with robots tend to be more willing to engage with them in healthcare contexts, while those with a negative view of how robots will impact society are less willing to engage with the robots in healthcare settings.

Negativity was significantly, and inversely, correlated with healthcare (see above results) and experience (G= -.21, p<.001) with weak directionality, indicating that those with more negative views of robots in society have less experience with robots in general and are less likely to be interested in engaging with them in a healthcare context. Table 9 illustrates the combined study correlations between the robot subscales while Table 10 provides the results by study.

Correlation of Robot Attitudes with other Robot Attitude Categories

	Anthro	Like	Intell	Protect	Health	Neg	Exper
Anthro		**0.32**	**0.22**	0.08	**0.23**	0.00	-0.07
Like	**0.32**		**0.24**	**0.24**	**0.16**	-0.05	-0.10
Intell	**0.22**	**0.24**		**0.17**	0.07	0.05	-0.09
Protect	0.08	**0.24**	**0.17**		**0.16**	0.00	0.00
Health	0.23	**0.16**	0.07	**0.16**		**-0.28**	**0.20**
Neg	0.00	-0.05	0.05	0.00	**-0.28**		**-0.21**
Exper	0.07	-0.10	-0.09	0.00	**0.20**	**-0.21**	

Key

	NA
Regular	p > .05
Bold	p < .05
Bold	p < .01

Tab. 9 Significant correlations among robot attitude subscales: anthropomorphism (anthro), likeability (like), perceived intelligence (intel), protectiveness towards the robot (protect), healthcare (health), negativity (neg), experience with the robot (exper).

Correlation of Robot Attitudes with other Robot Attitudes by Study

Study	Robot Attitude	Robot Attitude	Gamma	*P*-Value	Study Directionality
Combined	Anthropomorphism	Likeability	.32	<.001**	Strong
Study 2	Anthropomorphism	Likeability	.3	.002**	
Study 3	Anthropomorphism	Likeability	.33	<.001**	
Combined	Anthropomorphism	Perceived Intelligence	.22	.001**	Weak
Study 2	Anthropomorphism	Perceived Intelligence	.37	<.001**	
Study 3	Anthropomorphism	Perceived Intelligence	.13	.1	
Combined	Anthropomorphism	Healthcare	.23	<.001**	Weak
Study 2	Anthropomorphism	Healthcare	.23	.04*	
Study 3	Anthropomorphism	Healthcare	.2	.01*	
Combined	Likeability	Perceived Intelligence	.24	<.001**	Weak
Study 2	Likeability	Perceived Intelligence	.3	.002**	
Study 3	Likeability	Perceived Intelligence	.2	.03*	
Combined	Likeability	Protective	.24	<.001**	Weak
Study 2	Likeability	Protective	.19	.06	
Study 3	Likeability	Protective	.3	<.001**	
Combined	Likeability	Healthcare	.16	.01*	Weak
Study 2	Likeability	Healthcare	.1	.37	
Study 3	Likeability	Healthcare	.17	0.04*	
Combined	Perceived Intelligence	Protective	.17	.02*	Weak
Study 2	Perceived Intelligence	Protective	.2	.06	
Study 3	Perceived Intelligence	Protective	.17	.07	
Combined	Protective	Healthcare	.16	.009**	Moderate
Study 2	Protective	Healthcare	.15	.13	
Study 3	Protective	Healthcare	.21	.01*	
Combined	Healthcare	Negativity	-.28	<.001**	Moderate
Study 2	Healthcare	Negativity	-.26	.03*	
Study 3	Healthcare	Negativity	-.31	<.001**	
Combined	Healthcare	Experience	.2	.006**	Weak
Study 2	Healthcare	Experience	.06	.57	
Study 3	Healthcare	Experience	.29	.002**	
Combined	Negativity	Experience	-.21	.001**	Weak
Study 2	Negativity	Experience	-.08	.48	
Study 3	Negativity	Experience	-.27	<.001**	

Tab. 10 Significant correlations among robot attitude subscales: anthropomorphism, likeability, perceived intelligence, protectiveness towards the robot, healthcare, negativity, experience with the robot. * $p \leq .05$, ** $p < .01$, study directionality between Study 2 and Study 3 is considered weak when the difference in the magnitude of the gamma correlations for the two studies is substantial, with difference in $G > .06$, moderate for $G \leq .06$, and strong for $G \leq .03$.

Robot attitudes compared with chronic stress

There were no significant correlations between the robot attitude subscales and chronic stress.

Temperament subscales compared with temperament subscales

Temperament subscales compared with chronic stress

These correlations were not analyzed or included because none of the variables contain a robot component, nor would the results make sense to compare to outcomes with robots. While this information is interesting, there are entire disciplines devoted to studying these relationships, which is out of scope for this dissertation.

Discussion

The questionnaire data allowed us to compare temperament, experience, and attitudes about robots with behavioral measures on robot and human body language pose ratings. Temperament was found to correlate with robot poses; positive affect had an association with valence ratings on neutral and positive poses while negative affect was correlated with arousal for negative valence – low arousal poses. These findings are in line with past work in which people scoring high on extraversion (i.e., positive affect) had a relationship to perception of positivity (Young & Brunet, 2011) and those scoring high on introversion (i.e., negative affect) perceive more intensity (Buchanan, Bibas, & Adolphs, 2010). While ratings for the human body language stimuli in our studies did not follow a similar pattern, the difference may be due to the fact that the robot was viewed in 3D while the human stimuli were presented in 2D.

The pattern between past findings with human stimuli and our robot stimuli is exciting because it is another parallel between performance on perception of humans and robots. The more similar perception is between the two, the more likely it will be that people will have productive relationships with robots; the same communication skills and accommodations will be required for both, meaning that by the time people encounter robots, their lifetime of interactions with humans will have prepared them. Robots offer an additional opportunity for productive communication, because if we can train them to perceive temperament, we could program specific shifts in robot communication style to ensure successful interaction.

In addition to temperament, there were many correlations between robot pose ratings and attitudes towards robots. Experience, anthropomorphism, and (possibly) negative feelings towards robots, are categories in which perception may be mutable either through design choices and/or exposure to robots. These findings suggest that there are ways to impact perception of robot emotion. Used strategically, specific interventions may be able to counteract some detrimental effects of stress or temperament in high-risk domains. For example, participants scoring high on negative attitudes towards robots perceived the high arousal robot poses to be more aroused. Since perception of high arousal can lead to hypervigilance (Weymar et al., 2012), exploring ways to mediate negative attitudes could lower perceived arousal levels, muting response to high intensity stimuli.

That temperament and attitudes impact the perception of robot emotional body language are important findings. Data sets on robot emotion identification that do not parse temperament and attitudes may lead to the faulty conclusion that people in general can accurately perceive emotional robot body language and will therefore accept robots. However, the current results indicate that there are both fixed and comparatively mutable factors that impact an individual's perception of robot emotion, which could result in different levels of engagement and trust. In low-risk domains, deciding not to engage with a robot will typically not lead to harm. Conversely, in healthcare and other high-risk contexts, understanding who may not be fully benefitting from robot interaction will be important.

Chapter 6: General Discussion

General Discussion

As we work to design autonomous robots that will become integrated into multiple domains in our lives, researchers have investigated methods for ensuring effective human-robot communication. Emotional expression offers a rapid transfer of information between humans, and designers are interested in harnessing this technique to interact with robots. Early studies confirmed that humans are able to perceive and interpret a humanoid robot's emotional facial and body expressions. More recent work has sought to investigate how specific aesthetic and cognitive robot design parameters impact human-robot interaction, such as: size, shape, color, movement pattern, degree of anthropomorphism, and the cognitive capability of the robot. However, until the research reported here, an important component that had been overlooked was the influence of humans' psychological conditions and traits on their perception of robot emotion.

Many uses have been proposed for social and assistive robots. In high-risk applications such as healthcare, it is important to understand how an individual's circumstances may affect their ability and willingness to communicate and engage with robots. Our series of studies sought to investigate multiple factors that may influence human-robot interaction in these settings. The first area of focus was how acute cognitive and physical stress would impact perception of robot emotional body language. The second area investigated individual traits, including temperament, and attitudes toward and experience with robots, that may influence human-robot interaction. Thirdly, we explored similarities and differences between how emotional body language was perceived in a physically present robot versus the same robot rendered in virtual reality.

Study Design

We began by designing and norming a set of emotional body language poses for the robot Nao. We decided against generating poses that were meant to evoke a specific emotion, which would require a forced-choice protocol. In healthcare settings humans will not be given a list of possible emotions the robot could express to use as a guide. For this reason, we wanted to determine whether humans could perceive the overall valence and arousal of each pose without prompting. We used the valence and arousal Self-Assessment Manikin scales to assess this. We designed 10 poses in the following categories: Negative Valence-High Arousal, Negative Valence-Low Arousal, Neutral, Positive Valence-Low Arousal, and Positive Valence-High Arousal.

We debated the inclusion of the neutral category as humans never "express" true neutrality in their emotions. However, we reasoned that designers may use neutral poses as a strategy to communicate that the robot is not engaged or not processing information. The inclusion of a neutral set of poses also allowed a point of comparison to ensure that there was sufficient difference between the emotion categories.

We performed a norming study to determine which poses were most consistently identified in each target category. Participants were asked to rate each pose on the Self-Assessment Manikin (SAM) valence and arousal scales. One pose that had been initially designed as neutral was rated as positive valence-low arousal, and was selected for use in the second and third studies as a PL pose. All other selected poses performed most consistently within the category they had originally been designed to represent. The neutral and positive valence–low arousal poses were more similar in ratings than poses in any of the other categories.

All categories were perceptually differentiated (even NE and PL). There was some variation in the mean values of the emotion categories across studies. However, they all remained within the target categories, suggesting that perception of the emotional body language poses that were created and selected for the robot remained stable for both valence and arousal across studies.

Impact of Acute Cognitive and Physical Stress

Utilizing the set of pose stimuli, we first investigated the impact of acute cognitive and physical stress on perception of robot emotional body language. We sought to understand whether perception of robot emotions under stress is dominated by emotional valence, emotional arousal, or a combination. We asked whether acute stress captures or repels attention to negative or highly arousing emotional body language, and whether any of the same patterns hold true for chronic stress?

In Study 2 we induced stress using the MAST protocol, a combination of cold water hand immersion and mental arithmetic with a social evaluation component. This task was completed prior to rating the robot poses. Participant self-reports of their own emotional valence and arousal displayed a significant difference between the high and low stress conditions at time-points 2 and 3 (during MAST). However, by the end of viewing and rating the first block of robot poses, participants' self-reported affect ratings revealed that there was no longer affective differentiation between the high and low stress groups.

The apparent lack of sustained stress may explain why we did not find any effect of stress on the robot emotional body pose ratings in Study 2. As discussed in Chapter 3, there are multiple explanations for the convergence of self-reported mood ratings after the first block of poses. The MAST may not have been sufficiently salient to induce sustained stress. The physical and cognitive interruption created by requiring participants to move from one table (where the MAST was performed) to another table to view the robot poses may have proved a sufficient distraction. Or the presence of the robot itself (given that Nao was often perceived as endearing or "cute") could have had a moderating effect. Our study was not designed to investigate this last possibility, but it would be worth pursuing as a future research question; if robot presence is found to reduce stress, it would be another benefit for integrating social and assistive robots into healthcare contexts.

We adjusted the stress induction protocol for Study 3 by interleaving it with the robot pose identification task. This ensured that the stress-inducing trials were closer in time to the emotional pose identification task. This circumstance is more likely to mirror the temporal and physical proximity of robot interactions and stress in medical settings, as patients will be dealing with ongoing stressors while simultaneously engaging with robots. Analysis of the self-report ratings for Study 3 confirmed that stress was maintained in the high stress condition throughout viewing the robot poses.

In Study 3 there was an effect of stress — for the high arousal pose categories (whether those poses were negative or positive in valence). Participants in the high-stress condition perceived lower valence and also lower arousal in the NH and PH poses. In terms of design, this suggests that people under stress tend to see high intensity emotions as less energized and more negative. This could mean that poses intended to communicate excitement, happiness, or positive surprise may be perceived by acutely stressed individuals as lower in positivity and animation. While there may be a tendency to program robots to be cheerful in medical settings, these findings suggest that high arousal poses, even if they are meant to be positive, may not be an effective means for communicating this category of emotions.

Likewise, angry poses may be perceived as more negative than intended. Although we are not accustomed to medical staff expressing anger, and therefore may imagine that a robot would not need to communicate this emotion, robots in healthcare settings are going to play roles beyond that of medical staff; they will provide therapeutic assistance and also companionship. It may be that a robot would need to express anger, frustration, or fear to mirror what the patient is feeling or has experienced. The creation of these emotional expressions will have to be calibrated with the understanding that the perception of their negativity will be heightened and perception of arousal blunted when their human interaction partners are experiencing high levels of stress.

If NH and PH poses are redesigned for individuals experiencing high-stress, it presents an intriguing challenge. It may be that individuals who are not under the same level of acute stress, such as a caregiver, spouse, friend, or personal care assistant, would misinterpret the robot's communication. Instead of trying to calibrate the design of high arousal poses, it may be better practice to avoid their use in healthcare settings. There were no effects of stress for the low arousal categories, suggesting that these poses may yield more stable human-robot communication across high-stress and other settings. If it is important to convey arousal level, other strategies could be employed. For example, design elements such as lights or color could denote energy or arousal level.

While there were no effects of the acute stress induction on participants' emotional perception of low arousal poses, there were effects of presentation for low-arousal and neutral poses (NL, NE, and PL), with valence and arousal rated higher on the second presentation as compared to the first. This effect may be explained by mere-exposure, or growing familiarity with the robot. In Study 3 we only presented each pose twice; it would be instructive to investigate the pattern of increase in ratings over many more presentations, and across longer time periods.

Although we did not see evidence of a similar effect of familiarity on high arousal poses, it may be that more exposure would eventually yield similar results. Would there be a point at which an acutely stressed individual with sufficient previous robot exposure would rate the robot's high arousal expressions similarly to a low-stress individual? This would mean that robots that interact regularly with a particular patient could be programmed to express a wider range of emotions without concern for misinterpretation. However, given the lack of any trend towards increasing valence and arousal scores across presentation for the high arousal poses, and the very consistent effect of presentation for NL, NE, and PL, it may be that familiarity/exposure only impacts low arousal poses.

Temperament, Attitudes, and Experience

Results from our robot attitudes questionnaire may shed some light on the possible impact of increased familiarity or exposure. Those who reported more experience with robots tended to rate the NH poses as both lower in valence and arousal, and the second presentation of the PH poses as lower in arousal. These findings are in the same direction as the high-stress group. Since there are no "correct" values for the categories, it is impossible to determine whether the high-stress and experienced robot viewers perceived the high arousal poses more accurately. However, in some situations acute stress has been found to focus attention on threatening stimuli (Weymar, Schwabe, Löw, & Hamm, 2012), which may mean that participants in the high-stress condition had a more accurate perception of the robot pose. The experienced group may have rated the poses similarly because they are more familiar with robots, and thus interpret poses more "correctly," possibly indicating that high arousal poses would not be subject to the same mere-exposure increase in ratings seen in the low arousal poses.

Negative attitudes towards robots were positively correlated with high arousal poses on measurements of arousal. NH2, NE2, and PH2 were all rated as higher on arousal when participants felt more strongly that robots would have a negative impact on society. Attention is drawn to threatening stimuli, which a highly-aroused robot would likely appear to be for a viewer who has concerns about their role among humans.

Consistent with these behavioral findings, experience with robots and negative attitudes about robots were negatively correlated. Those with less experience may have intentionally avoided robots because of their negative views, or they may have negative views because they have not been exposed. These perceptual patterns may be self-reinforcing as someone with negative attitudes will perceive the robot as more aroused/threatening, which could increase their discomfort with robots in society. Conversely, someone with more experience will tend to perceive the emotional expression as more negative, but less arousing, meaning that the robot would appear less threatening.

This pattern may seem to suggest that an analysis of the temperament subconstructs would reveal a relationship between ratings on high arousal poses and the trait negative affect factors. While negative attitudes toward robots were associated with both fear and discomfort, neither subconstruct was correlated with high arousal ratings. Instead, fear was correlated with higher arousal for NL1 and NL2, and discomfort was associated with perceived high arousal on PL2 - the low arousal categories.

Unfortunately, the Negative Attitudes Towards Robots Survey (NARS) does not investigate the reasons behind participants' negative attitudes. As the correlations between the low arousal ratings and fear and discomfort were weak, there are many factors beyond temperament that could account for negative attitudes about robots. It may be that individuals scoring high on trait negative affect perceive robots to be more intense on what are usually less intense poses because they react more strongly to novelty. Participants who perceive robots as a negative influence on society may be reacting to factors beyond novelty, such as the potential for loss of jobs, human connections, and traditions. These factors are threatening in many life domains, so participants who are concerned with these impacts may be hypervigilant to highly aroused robot expressions.

While negative attitudes and experience were correlated with high arousal poses, anthropomorphism was consistently associated with low arousal poses. Ratings for valence on NE1, PL1, and PL2 and arousal for NE1 and PL1 were positively correlated with anthropomorphism; the more human-like the robot appeared, the more positive and animated the participant perceived its emotional expression to be. Anthropomorphism is associated with positive affect from the extraversion temperament construct, and positive affect was correlated with higher valence ratings for NE2, PH1, PH2. As Yi et al. (2016) discusses, people with different temperaments allocate attentional resources differently. It may be that those high in trait positive affect perceived the positive high arousal poses as even more positive because they are drawn to those emotions in their environment.

This insight is interesting and in line with the findings of Yi et al. (2016). Equally important, it offers an important design opportunity. If robot designers wish to successfully integrate robots into healthcare settings, where patients may be unfamiliar with the technology and therefore likely to perceive its emotional expressions as more threatening, then focusing on anthropomorphism as a design parameter could counteract this disadvantageous perceptual pattern.

When comparing correlations among the robot attitude categories, two clusters appeared. First, anthropomorphism was correlated with likeability, perceived intelligence, and willingness to engage with robots in a healthcare setting. (Additionally, likeability, perceived intelligence, and willingness to engage healthcare robots were correlated with protectiveness.) Designing healthcare robots that will be perceived as human-like, likeable, and intelligent may increase willingness to engage healthcare robots directly, and could lead to perception of higher valence in positive low arousal expressions (NE and PL), which may in turn increase favorable sentiments about the robot.

Prior experience with robots was also positively correlated with willingness to engage robots in healthcare, while negativity was negatively correlated with healthcare robots; perhaps not surprisingly, negativity and experience were negatively correlated. This second cluster was not correlated with any of the other robot attitude categories.

On the one hand, this means that either prior experience, or a combination of perceiving the robot as likeable, human, and intelligent may lead to openness to utilizing the services of a healthcare robot. On the other hand, the lack of correlation between the clusters indicates that some participants rated the robot high on some or all of the factors that are correlated with healthcare robot acceptance and also felt negatively about robots in society. As the negative correlation between negativity and healthcare is stronger than any of the positive associations with healthcare, in some cases it seems that people's negative view of robots in society is a more compelling factor in deciding whether to engage healthcare robots than their perception of its animacy and vulnerability.

In terms of increasing acceptance in this group, exposure to robots may be the most effective strategy. However, the directionality of the relationship between prior experience and negative views of robots is unclear. It may be that people with negative views of robots intentionally avoid them, thereby ensuring lack of experience, or people without prior experience with robots may be predisposed to view them negatively until having the opportunity to interact.

Future work on the effects of exposure to robots on emotional expression perception, and robot attitudes – especially willingness to utilize healthcare robots – would be instructive. From Study 3 we know that valence ratings of low arousal poses increase over time, leading to the perception of the robot's emotional state as more positive. We also know that prior experience is associated with increased willingness to engage. Although it will be important to consider the integration of other design factors discussed here, healthcare settings could increase exposure to robots ahead of patient-robot interactions as another strategy for mediating risk of poor communication or lack of acceptance in human-robot interaction. As discussed earlier, this series of studies has not found conclusive evidence that perception of high arousal poses would be tempered beneficially with exposure, but designers could either avoid high arousal poses altogether, or only employ them with individuals experienced in robot interaction since they already indicate a willingness to engage with healthcare robots.

Temperament also plays a role in the perception of robot emotion. Participants scoring higher on trait positive affect tended to perceive positive valence – high arousal (presentation 1 and 2) poses and the second presentation of the neutral poses as more positively valenced, while those scoring higher on the temperamental sub-dimension of fear perceived negative valence-low arousal poses to be more aroused. These findings confirm that there will be individual variation in emotion perception. As the cognitive and expressive capacities of robots increase, developing temperament detection algorithms may be beneficial for real-time calibration of robot emotional expressions to ensure optimized communication with individual patients. However, in initial forays into integrating robots into healthcare settings, considerations of temperamental differences should be a low priority.

Likewise, chronic stress did not prove to be an impactful factor in the perception of robot emotion. Nonetheless, although there was only one significant correlation – between chronic stress and the second presentation of neutral poses – it should not be concluded that chronic stress has no influence on emotional perception of robot body language. Participants were not recruited by chronic stress level, so the range on this dimension in our sample was limited. For example, there were 98 points possible on the questionnaire and the highest score from a participant was 61, with the vast majority of participants receiving scores between 5 and 30. As such, the effect of chronic stress on perception of robot emotional expression is still an open question. It would be worthwhile to pursue, either to confirm or contradict our findings, because many patients experience chronic – in addition to acute – stress.

Utilizing Virtual Reality Robots to Design for the Real World

Based on our findings, people's willingness and ability to engage and communicate with robots in high stress medical settings is impacted by acute stress, arousal level of robot poses, anthropomorphism, likeability, perceived intelligence, and perceived vulnerability of the robot. Many studies have experimented with design parameters that increase or decrease perception of robot animacy. However, to our knowledge, none of those studies or designers have considered the impact of stress. It would be advisable to vary parameters such as shape, size, color, movement, and materials and measure reactions under stress.

Building so many variations of robots is cost prohibitive, but generating robots in virtual reality is not. As our third area of focus we investigated the similarities and differences in how emotions were perceived when conveyed through the body language of the physical robot Nao and a recreation of the same robot in virtual reality, with the aim to gather insight about the potential to use virtual reality as a "sandbox" for robot design. Study 3 was a 2 (low vs. high stress) by 2 (physical vs. virtual robot) factorial design. We found that those perceiving the robot in the virtual reality conditions performed similarly to their physical condition counterparts when rating NH, NL, and NE poses; those in the high stress group similarly perceived the NH poses to be more negative and less aroused than those in the control group. We also found a similar pattern for an increase in valence and arousal ratings across presentation for the negative valence-low arousal and neutral poses. These findings suggest that there is potential for using virtual reality robots as proxies during design. As Li (2015) hypothesized, perception of virtual reality renditions of robots may align more closely with physical as compared to virtual (2D) agents, because they are both embodied and present in a viewer's environment.

However, there were significant effects of robot type for PL and PH poses on both valence and arousal, with the physical robot rated higher on all measures. Why would robot type impact performance on positive, but not neutral or negative, poses? Will this difference in performance always exist, or are there design factors that could be adjusted to align perception on these measures?

We found an effect of robot type on two of the robot attitudes categories. There was a trend for anthropomorphism, and a significant effect for likeability, with the physical robot rated higher than the virtual reality robot for both. Anthropomorphism and likeability are correlated, and anthropomorphism is associated with higher valence and arousal ratings for neutral and PL pose categories, which suggests that some of the variation may be due to differences in perceived animacy. The virtual reality robot was generated by a sophisticated algorithm that uses discrete 2D photographs to generate a 3D model. However, there are limitations to this modeling approach and almost all of the poses had some visual artifacts present, such as pixilation and soft focus. Despite its limitations, this modeling technique was chosen due to cost considerations. More sophisticated systems now exist that utilize CAD files of existing robots to model them in 3D. This prevents artifacts and distortion and also enables lighting to be precisely rendered in the virtual environment.

It would be instructive to run the same protocol used in Study 3, but with a virtual reality model of the robot generated through these more precise means. Would the increase in fidelity be matched by increases in ratings for the degree of anthropomorphism and likeability of the virtual reality robot? Would the effect of robot type be entirely eliminated for the positive valence poses? Investing in research in this area could have long-term benefits for addressing other robot design questions if the virtual environment can function as a lower-cost but higher-efficiency and higher-flexibility testbed.

The ultimate goal for healthcare robot researchers and designers is to create autonomous tools that can improve efficacy, efficiency, and quality of healthcare. While human-robot interaction is a maturing field, studies have primarily focused on design parameters intrinsic to the robot. To make significant gains in the integration of robots into daily contexts, we must also consider how circumstances (such as situationally induced acute stress) and individual traits (such as attitudes and temperament) impact humans' ability and willingness to communicate with robots. By exploring these factors and developing design plans to mitigate opportunities for miscommunication, we can increase the likelihood of successful encounters with healthcare robots, ultimately improving care and patient quality of life.

Limitations

While this series of studies begins to explore the perception of robot emotional body language, there are factors that limit the findings' generalizability. First, we were only able to test participants' reactions to one robot design. The robot was humanoid in shape, but small. Healthcare robots will likely exist in many sizes, so it is ecologically valid to test a 54 cm robot, but with no other statures to compare with our findings, it is not possible to determine whether perception of emotion for differently-sized robots will follow similar patterns. Likewise, the humanoid shape is effective at facilitating emotional expression, but is it necessary? Would human emotion perception follow similar patterns under stress if the assistant robot were shaped like a non-human animal or in some novel form? Our robot had a face with eyes that were illuminated, but did not change color. The Nao robot also had a small mouth and a head shape that suggested ears on the sides. If body language is the salient form of expression, does the presence of a face, or even of a head, matter? Are there non-humanoid shapes that are less prone to perceptual shifts under acute stress or given a certain temperament or attitude towards robots? In the context of acutely stressful circumstances, how well would these potentially quite divergent body forms score on Godspeed's animacy scales?

Another limitation to the studies was the choice to use static body language poses. Emotional expression is comprised of many aspects: key pose frames (i.e., static poses), movement, gesture, head tilt. We chose static body poses because they are the most basic articulation of emotional expression that can be identified (Beck, 2013). However, robots typically will be moving when they are engaging with patients. Dynamically unfolding movement-related gestures are likely to provide patients with far more emotional information than static poses. Emotion is also context dependent. Due to our interest in exposing participants to poses from multiple categories, the test environment itself was kept emotion-neutral. If the robot's body language had been performed in reaction to specific circumstances or detailed scenarios, participants would have had more information about the possible motivations of the robot, possibly mediating some of the shifts that occurred due to stress, familiarity, and individual traits.

There are also some limitations to our methods for assessing participants' level of stress. We originally planned to collect self-report and biometric feedback. Unfortunately, the manufacturer of our heart rate monitor did not provide sufficiently detailed and accurate data, so we were ultimately reliant exclusively on self-report for our measurement. Participants may have intentionally or unintentionally reported inaccurate ratings; some may have intuited what the experimenter wanted them to say and answered accordingly, others may have wanted to mask their reactions to the task, while others may have exaggerated.

For those in the high stress group, regardless of their self-reported valence and arousal scores, the stress induction task had differing impacts. In the debriefing phase, some participants described the ice water baths they had to take in high school or college to treat injuries and explained that the cold-water task had not been stressful. Others claimed that the math was easy while the hand immersion was nearly impossible to sustain. A few participants stated that they eventually guessed which stress condition they had been assigned to. If they understood that the protocol was designed to cause stress, they may have been able to self-regulate to a certain degree and reduce the impact.

The effects of the stress were short-lived. Patients in healthcare environments may have much more intense and prolonged acute stress. Would this increase in severity and persistence increase the effects of stress on emotion perception? At higher levels of stress intensity, would low arousal poses also be impacted? Or will the greater richness of information available in contextual, gestural emotional body language temper any additional effects of more severe acute stress?

There are many open questions regarding the design of the virtual environment and robot. Although the virtual reality model of the robot was embodied and present in the same environment as the viewer, the viewer was required to continually leave that environment to complete other tasks. If the viewer's own body had been modeled in the virtual world so that they could complete all of the tasks in the same context, the sense of presence may have increased, possibly increasing perception of robot animacy and influencing emotion ratings. Because of the relatively low fidelity modeling of the robot, the lighting (and therefore color) was not identical in the physical and virtual environment. The constant switching between virtual and physical world, which were imperfect versions of one another, may have added to a sense of dis-ease or disorientation, especially reducing ratings for positively-valenced poses.

Additionally, the audio of the robot movement was played on external speakers to avoid another piece of equipment to don and doff. The trade-off was an aural landscape for the virtual condition that was noticeably different from the physical world (in which the audio of robot movements came directly from the robot). All these design parameters could be explored in future studies to determine which components are most important for eliciting similar behavioral responses from physical and virtual robots.

Contributions

To our knowledge this dissertation is the first exploration of the effects of acute and chronic stress on human perception of robot emotional body language. Robots are an intriguing tool because they can be deployed in circumstances and contexts that are physically and/or psychologically hazardous without worry that they will suffer. Conversely, the humans receiving robotic assistance in these contexts will most likely be experiencing some degree of acute stress. By examining how humans perceive robot communication while under stress, this dissertation begins the work of considering design factors that will facilitate human-robot interaction in real world high-risk and/or high-tension conditions.

Many studies have surveyed humans about their reactions to interactions with robots. Few have investigated correlations between experimental measures and those reactions. Our examination of how personal characteristics (i.e., temperament, attitudes toward robots) and prior experience with robots are related to perceptual processing will facilitate more sophisticated and targeted robot design.

Design recommendations based on our findings include: avoid high-arousal poses because stress and experience with robots seem to affect perception; low-arousal robot poses should be utilized in high-risk settings as they are consistently identified and people tend to find them more positive as they gain familiarity with robots; anthropomorphism may increase likelihood to engage with a robot in healthcare settings and is correlated with other positive feelings towards robots such as likeability and perceived intelligence; repeated exposure to robots prior to requiring their services may result in more positive interactions through emotion perception and attitudes towards use of robots; variation in emotion perception of robot emotional expression based on temperament mirrors patterns found in relation to human stimuli and should be considered when creating algorithms for robot interaction.

By comparing performance with physical versus virtual reality robots and documenting that human perceptions of emotional body language were largely similar for the majority of emotional poses, we have begun to create a technique for testing robot design more rapidly, iteratively, and with fewer costs. Once this sandbox technique yields consistently similar results between the two robot types, experimentation and consideration of the human factors involved in robot design should accelerate.

Our last contribution was unintended, but will nonetheless prove useful for future work. We designed a variation of the MAST stress induction protocol that facilitates broad participant engagement while maintaining high levels of stress. By allowing individuals to remove their hand from the cold-water immersion if it becomes too painful, studies using this modified acute stress induction procedure will have less participant attrition. By facilitating the inclusion of people with different pain thresholds, investigations will have greater generalizability. Additionally, the prompt to continue subtracting into negative numbers after a participant reaches zero on the math task implicitly provided sustained negative feedback regarding performance.

Taken together, by illuminating human factors that impact human-robot interaction in the real world and pioneering the development of an economical experimental technique to facilitate broader experimentation, the findings collectively reported here will contribute to effective design of robots operating in high-risk domains.

References

References

Abercrombie, H. C., Kalin, N. H., & Davidson, R. J. (2005). Acute cortisol elevations cause heightened arousal ratings of objectively nonarousing stimuli. *Emotion, 5*, 354–359. http://doi.org/ 10.1037/1528-3542.5.3.354

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, *4*(7), 267–278. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10859571

Aldebaran Softbank Group. (2016). A-Robots: Nao. Retrieved November 18, 2016, from https://www.ald.softbankrobotics.com/en/cool-robots/nao

Alexander, M. (2012). Managing patient stress in pediatric radiology. (Report). *Radiologic Technology, 83*(6), 549.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*, 219–235. http://doi.org/10.1177/1088868309341564

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–9. doi:10.1126/science.1224313

Baddoura, R., Venture, G., & Matsukata, R. (2012). The familiar as a key-concept in regulating the social and affective dimensions of HRI. *12th IEEE-RAS International Conference on Human Robots*, 234-241.

Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, *3*, 41–52. doi:10.1007/s12369-010-0082-7

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*, 71-81.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Brain Research. Cognitive Brain Research*, *17*(3), 613–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14561449

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of emotional body language displayed by a humanoid robot: A case study with children. *International Journal of Social Robotics*, *5(3)*, 325–334. http://doi.org/10.1007/s12369-013-0193-z

Bekele, E., Crittendon, J., Zheng, Z., Swanson, A., Weitlauf, A., Warren, Z., & Sarkar, N. (2014) Assessing the utility of a virtual environment for enhancing facial affect recognition in adolescents with autism. *Journal of Autism and Developmental Disorders, 44*(7), 1641-1650.

Bethel, C., & Murphy, R. (2008) Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems Mand and Cybernetics – Part C: Applications and Reviews, 38*(1), 83-92.

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass. O. (2016). What robots can teach us about intimacy: The reassuring effects of

robot responsiveness to human disclosure. *Computers in Human Behavior, 63*, 416–423. http://doi.org/ 10.1016/j.chb.2016.05.064
Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25(1)*, 49–59. http://doi.org/10.1016/0005-7916(94)90063-9
Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, *42*(3), 167-175.
Buchanan, T.W., Bibas, D., & Adolphs, R. (2010). Associations between feeling and judging the emotions of happiness and fear: Findings from a large-scale field experiment. *PLOS ONE, 5*, e10640.
Canli, T. (2004). Functional brain mapping of extraversion and neuroticism: Learning from individual differences in emotion processing. *Journal of Personality*, *72*(6), 1105-1132.
Casper, J., & Murphy, R. R. (2003). Human – Robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man, and Cybernetics--Part B: Cybernetics*, *33*(3), 367–385.
Chammat, M., Foucher, A., Nadel, J., & & Dubal, S. (2010). Reading sadness beyond human faces. *Brain Research*, *1348*, 95–104. http://doi.org/10.1016/j.brainres.2010.05.051
Chang, C., & Murphy, R. R. (2007). Towards robot-assisted mass-casualty triage. In *Proceedings of the 2007 IEEE International Conference on Networking, Sensing and Control* (pp. 15–17).
Clewett, D., Schoeke, A., & Mather, M. (2013). Amygdala functional connectivity is reduced after the cold pressor task. *Cognitive, Affective, and Behavioral Neuroscience*, *13*(3), 501-518.
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), 668-678.
Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusion and viewpoint dependence. *Journal of Nonverbal Behavior*, *28*(2), 117–139. http//doi.org/10.1023/B:JONB.0000023655.25550.be
Csala, E., Nemeth, G., & Zainko, C. (2012). Application of the NAO humanoid robot in the treatment of marrow-transplanted children. In *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 655–659). IEEE. doi:10.1109/CogInfoCom.2012.6421933
DeDora, D. J., Carlson, J. M., & Mujica-Parodi, L. R. (2011). Acute stress eliminates female advantage in detection of ambiguous negative affect. *Evolutionary Psychology*, *9(4)*, 532–542. http://doi.org/nsq097[pii]10.1093/scan/nsq097
De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, *7*(3), 242–9. doi:10.1038/nrn1872
de Gelder, B., & Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring

perception of whole body expression of emotions. *Frontiers in Psychology, 2*(August), 181. http://doi.org/doi.10.3389/fpsyg.2011.00181
Dennett, D. (1989). *The Intentional Stance*. Cambridge, MA: Bradford MIT Press.
Derogatis, LR., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*(3), 595-605.
Derryberry, D., & Rothbart, M.K. (1988). Arousal, affect, and attention as components of temperament. *Journal of Personality and Social Psychology, 55*(6), 958-966.
De Sá Dias, D., Resende, M., & Diniz, G. (2015). Patient stress in intensive care: Comparison between a coronary care unit and a general postoperative unit. *Revista Brasileira De Terapia Intensiva, 27*(1), 18-25.
Dubal, S., Foucher, A., Jouvent, R., & Nadel, J. (2011). Human brain spots emotion in non humanoid robots. *Social Cognitive and Affective Neuroscience*, *6*, 90–97. http://doi.org/doi:10.1093/scan/nsq019
Duffy, B. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*(3), 177-190.
Eger, E., Jedynak, A., Iwaki, T., & Skrandies, W. (2003). Rapid extraction of emotional expression: Evidence from evoked potential fields during brief presentation of face stimuli. *Neuropsychologia, 41*(7), 808–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12631531
Ellenbogen, M., Schwartzman, A. E., Stewart, J., & Walker, C.-D. (2002). Stress and selective attention: The interplay of mood, cortisol levels, and emotional information processing. *Psychophysiology, 39*, 723–732. http://doi.org/10.1017/S0048577202010739
Erden, M. S. (2013). Emotional postures for the humanoid-robot Nao. *International Journal of Social Robotics, 5(4)*, 441–456. http://doi.org/10.1007/s12369-013-0200-4
Evans, D.E. & Rothbart, M.K. (2007). Developing a model for adult temperament. *Journal of Research in Personality, 41*(4), 868-888.
Fink, J. (2012) Anthropomorphism and human likeness in the design of robots and human-robot interaction. *Social Robotics, Proceedings of 4th International Conference on*, 199-208.
Fridin, M., & Belokopytov, M. (2014). Robotics agent coacher for CP motor function (RAC CP Fun). *Robotica, 32*(8, SI), 1265-1279.
Gilbert, T., Martin, R., & Coulson, M. (2011). Attentional biases using the body in the crowd task: Are angry body postures detected more rapidly? *Cognition & Emotion, 25*(4), 700-708.
Gliga, T., & Dehaene-Lambertz, G. (2005). Structural encoding of body and face in human infants and adults. *Journal of Cognitive Neuroscience*, *17*(8), 1328–40. doi:10.1162/0898929055002481
Hadjikhani, N., & de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. *Current Biology*, *13*(24), 2201–2205. doi:10.1016/j.cub.2003.11.049
Häring, M., Bee, N., & Andre, E. (2011). Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In 2011 *RO-MAN 20th IEEE International Symposium on Robot and Human*

Interactive Communication. http://doi.org/10.1109/ROMAN.2011.6005263
Huang, Y., Yu, X., Ming, X., Xiang, B., & Ge, Y. (2009). Two types of flexible tactile sensor arrays of robot for three-dimension force based on piezoresistive effects. *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, 1032-1037.
Hwang, J., Park, T., & Hwang, W. (2013). The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot. *Applied Ergonomics*, *44*(3), 459–71. doi:10.1016/j.apergo.2012.10.010
Jefferies, L., Smilek, D., Eich, E., et al. (2008). Emotional valence and arousal interact with attentional control. *Psychological Science*, *19*(3), 290-295.
Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., & Hagita, N. (2009). An affective guide robot in a shopping mall. *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, 173-180.
Klein, B., Gaedt, L., & Cook, G. (2013). Principles and experiences with Paro in Denmark, Germany, and the UK. *GeroPsych: Journal of Gerontopsychology and Geriatric Psychiatry*, *26(2)*, 89–99. http://doi.org/10.1024/1662-9647/a000085
Kleinsmith, A., & Bianchi-Berthouze, N. (2011). Form as a cue in the automatic recognition of non-acted affective body expressions. *IEEE Transactions on Systems, Man, and Cybernetics--Part B: Cybernetics*, *41*(4), 155–164.
Koschate, M., Potter, R., Bremmer, P., & Levine, M. Overcoming the uncanny valley: Displays of emotions reduce the uncanniness of humanlike robots. *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, 359-365.
Kret, M. E., Stekelenburg, J. J., Roelofs, K., & de Gelder, B. (2013). Perception of face and body expressions using electromyography, pupillometry and gaze measures. *Frontiers in Psychology*, *4*(February), 28. http://doi.org/10.3389/fpsyg.2013.00028
Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50*, 372–385. http://doi.org/10.1037//0003-066X.50.5.372
Lang P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O. (2013). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 261–273.
Lange, J., & Lappe, M. (2007). The role of spatial and temporal information in biological motion perception. *Advances in Cognitive Psychology*, *3*(4), 419–28. doi:10.2478/v10053-008-0006-3
Larsen, R., Ketelaar, T., & Sarason, Irwin G. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology, 61*(1), 132-140.
Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, *77*, 23–37. http://doi.org/10.1016/j.ijhcs.2015.01.001
Lohani, M., Stokes, C., McCoy, M., et al. (2016). Social interaction moderates human-robot trust reliance relationship and improves stress coping. *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*,

471-472.
Malik, N. A., Yussof, H., & Hanapiah, F. A. (2014). Development of imitation learning through physical therapy using a humanoid robot. *Procedia Computer Science*, *42*, 191–197. doi:10.1016/j.procs.2014.11.051
Mara, M., & Appel, M. (2015). Effects of lateral head tilt on user perceptions of humanoid and android robots. *Computers in Human Behavior, 44*, 326–334. http://doi.org/10.1016/j.chb.2014.09.025
Martelaro, N., Nneji, V., Ju, W., & Hinds, P. (2016). Tell me more: Designing HRI to encourage more trust, disclosure, and companionship. *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, 8370.
Martinez, L., Falvello, V., Aviezer, H., & Todorov, A. (2015). Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cognition and Emotion,* 1-14.
Marttos, A., Kelly, E., Graygo, J., Rothenberg, P., Alonso, G., Kuchkarian, F. M., … Schulman, C. I. (2013). Usability of telepresence in a level 1 trauma center. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *19*(4), 248–51. doi:10.1089/tmj.2012.0102
Mather, M., Lighthall, N. R., Nga, L., & Gorlick, M. A. (2010). Sex differences in how stress affects brain activity during face viewing. *Neuroreport*, *21*(14), 933–7. doi:10.1097/WNR.0b013e32833ddd92
McColl, D., & Nejat, G. (2014). Recognizing emotional body language displayed by a human-like social robot. *International Journal of Social Robotics, 6*, 261–280. http://doi.org/10.1007/s12369-013-0226-7
Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16518–23. doi:10.1073/pnas.0507650102
Mohanty, A., & Sussman, T. J. (2013). Top-down modulation of attention by emotion. *Frontiers in Human Neuroscience*, *7*(April), 102. doi:10.3389/fnhum.2013.00102
Moser, E., Derntl, B., Robinson, S., Fink, B., Gur, R. C., & Grammer, K. (2007). Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *Journal of Neuroscience Methods*, *161*(1), 126–33. doi:10.1016/j.jneumeth.2006.10.016
Murphy, R. R. (2014). *Disaster Robotics*. Cambridge, MA: MIT Press.
Newman, K. R., & Sears, C. R. (2015). Eye gaze tracking reveals different effects of a sad mood induction on the attention of previously depressed and never depressed women. *Cognitive Therapy and Research, 39(3)*, 292–306. http://doi.org/10.1007/s10608-014-9669-x
Ni, J., Jiang, H., Jin, Y., Chen, N., Wang, J., Wang, Z., … Hu, X. (2011). Dissociable modulation of overt visual attention in valence and arousal revealed by topology of scan path. *PLoS ONE, 6(4).* Article e18262. http://doi.org/10.1371/journal.pone.0018262
Nomura, T., Suzuki, T., Kanda, T, & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3), 437-454.

Novaes, M., Aronovich, A., Ferraz, F., & Knobel, P. (1997). Stressors in ICU: Patients' evaluation. *Intensive Care Medicine, 23*(12), 1282-1285.

Novaes, M., Knobel, E., Bork, A., Pavão, O., Nogueira-Martins, L., & Bosi Ferraz, M. (1999). Stressors in ICU: Perception of the patient, relatives and health care team. *Intensive Care Medicine, 25*(12), 1421-1426.

Pan, X., Han, C. S., Dauber, K., & Law, K. H. (2007). A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI & Society*, *22*(2), 113–132. doi:10.1007/s00146-007-0126-1

Pate, J., Blount, R., Cohen, L., & Smith, A. (1996). Childhood medical experience and temperament as predictors of adult functioning in medical situations. *Children's Health Care, 25*(4), 281-298.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: A systematic review. *Autism Research*, *9*(2), 165-183.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*(16), 5205–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8756449

Putnam, P., & Roelofs, K. (2011). Effects of single cortisol administrations on human affect reviewed: Coping with stress through adaptive regulation of automatic cognitive processing. *Psychoneuroendocrinology, 36,* 439–448. http://doi.org/ 10.1016/j.psyneuen.2010.12.001

Quigley, L., Nelson, A. L., Carriere, J., Smilek, D., & Purdon, C. (2012). The effects of trait and state anxiety on attention to emotional images: An eye-tracking study. *Cognition & Emotion*, *26*, 1390–1411. http://doi.org/10.1080/02699931.2012.662892

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science, 9*, 45–48. http://doi.org/ 10.1111/1467-9280.00008

Recio, D. L., Segura, E. M., Segura, L. M., & Waern, A. (2013). The NAO models for the elderly. *Human-Robot Interaction, 8th ACM/IEEE International Conference on HRI*, 187–188. http://doi.org/10.1109/HRI.2013.6483564

Recio, G., Conrad, M., Hansen, L. B., & Jacobs, A. M. (2014). On pleasure and thrill: The interplay between arousal and valence during visual word recognition. *Brain & Language, 134*, 34–43. http://doi.org/10.1016/j.bandl.2014.03.009

Rennick, J., Johnston, C., Dougherty, G., Platt, R., & Ritchie, J. (2002). Children's psychological responses after critical illness and exposure to invasive technology. *Journal of Developmental and Behavioral Pediatrics, 23*(3), 133-44.

Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! that could be dangerous: Valence-arousal interactions in evaluative processing. *Personality and Social Psychology Bulletin, 30(11)*, 1472–1484. http://doi.org/10.1177/0146167204266647

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178.

Sanchez, A., Vazquez, C., Gomez, D., & Joormann, J. (2014). Gaze-fixation to happy faces predicts mood repair after a negative mood induction. *Emotion, 14(1)*, 85–94. http://doi.org/10.1037/a0034500

Sanchez, M. (2012, August). The history of medical robots! *neurochangers*. Retrieved July 18, 2015, from http://neurochangers.com/2012/08/09/the-history-of-medical-robots/

Schmid, P., & Schmid Mast, C. (2010). Mood effects on emotion recognition. *Motivation and Emotion, 34*(3), 288-292.

Seelye, A. M., Wild, K. V., Larimer, N., Maxwell, S, Kearns, P., & Kaye, J. A. (2012). Reactions to a remote-controlled video-communication robot in seniors' homes: A pilot study of feasibility and acceptance. *Telemedicine and e-Health, 18*, 755–759. http://doi.org/10.1089/tmj.2012.0026

Smeets, T., Cornelisse, S., Quaedflieg, C. W. E. M., Meyer, T., Jelicic, M., & Merckelbach, H. (2012). Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress response. *Psychoneuroendocrinology, 37*, 1998–2008. http://doi.org/ 10.1016/j.psyneuen.2012.04.012

Stoffregen, T. a., Villard, S., & Yu, Y. (2009). Body sway at sea for two visual tasks and three stance widths. *Aviation Space and Environmental Medicine*, *80*(12), 1039–1043. doi:10.3357/ASEM.2538.2009

Turner, J.R., Wheater, B., & Lloyd, D.A. (1995). The epidemiology of social stress. *American Sociological Review*, *60*(1), 104-125.

Van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2009). From specificity to sensitivity: How acute stress affects amygdala processing of biologically salient stimuli. *Biological Psychiatry*, *66*(7), 649–655. doi:10.1016/j.biopsych.2009.05.014

van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2010). Enhanced resting-state connectivity of amygdala in the immediate aftermath of acute psychological stress. *NeuroImage, 53*, 348–354. http://doi.org/10.1016/j.neuroimage.2010.05.070

Von Känel, R., Baumert, J., Kolb, C., Cho, E-Y.N., & Ladwig, K-H. (2011). Chronic posttraumatic stress and its predictors in patients living with an implantable cardioverter defibrillator. *Journal of Affective Disorders, 131*(1-3), 344-352.

Wallraven, C., Breidt, M., Cunningham, DW., & Bülthoff, HH. (2008). Evaluating the perceptual realism of animated facial expressions. ACM Transactions on Applied Perception, 4(4), 23:1-23:20. doi:10.1145.1278760.1278764

Wartenberg, C., & Wiborg, P. (2003). Precision of exocentric distance judgments in desktop and cube presentation. *Presence: Teleoperators and Virtual Environments*, *12*(2), 196–206. doi:10.1162/105474603321640941

Wade, E., Dye, J., Mead, R., & Mataric, M. (2011). Assessing the quality and quantity of social interaction in a socially assistive robot-guided therapeutic setting. *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, 1-6.

Watson, D., Clark, L., Tellegen, A., & Sarason, Irwin G. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Waytz, A., & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. *Emotion, 14*, 434–444. http://doi.org/ 10.1037/a0036054

Weymar, M., Schwabe, L., Löw, A., & Hamm, A. O. (2012). Stress sensitizes the brain: Increased processing of unpleasant pictures after exposure to acute stress. *Journal of Cognitive Neuroscience, 24*, 1511–1518. http://doi.org/10.1162/jocn_a_00174

Wilkes, D. M., Franklin, S., Erdemir, E., Gordon, S., Strain, S., Miller, K., & Kawamura, K. (2010). Heterogeneous artificial agents for triage nurse assistance. *2010 10th IEEE-RAS International Conference on Humanoid Robots*, 130–137. doi:10.1109/ICHR.2010.5686839

Wirth, M. M., Scherer, S. M., Hoks, R. M., & Abercrombie, H. C. (2011). The effect of cortisol on emotional responses depends on order of cortisol and placebo administration in a within-subject design. *Psychoneuroendocrinology, 36*, 945–954. http://doi.org/10.1016/j.psyneuen.2010.11.010

Yi, C., Murry, M., & Gentzler, A. (2016). Perception of emotional expressions in adults. *Journal of Individual Differences*, *37*(1), 16-23.

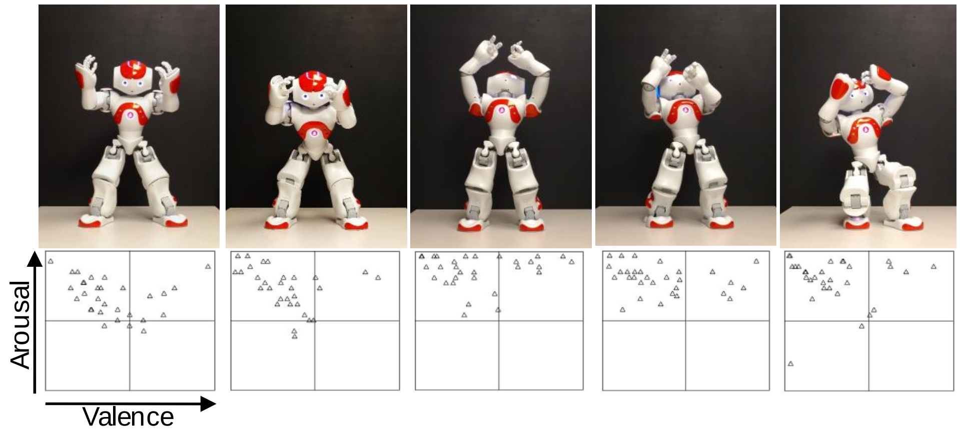
Young, M.E., & Brunet, P.M. (2011). It's about time: The influence of sociability – but not shyness – on the categorization of facial expressions of emotion. *Journal of Individual Differences*, *32*, 65-73.

Zanatto, D., Patacchiola, M., Goslin, J., & Cangelosi, A. (2016) Priming anthropomorphism: Can credibility of humanlike robots be transferred to non-humanlike robots? *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, 117-123.

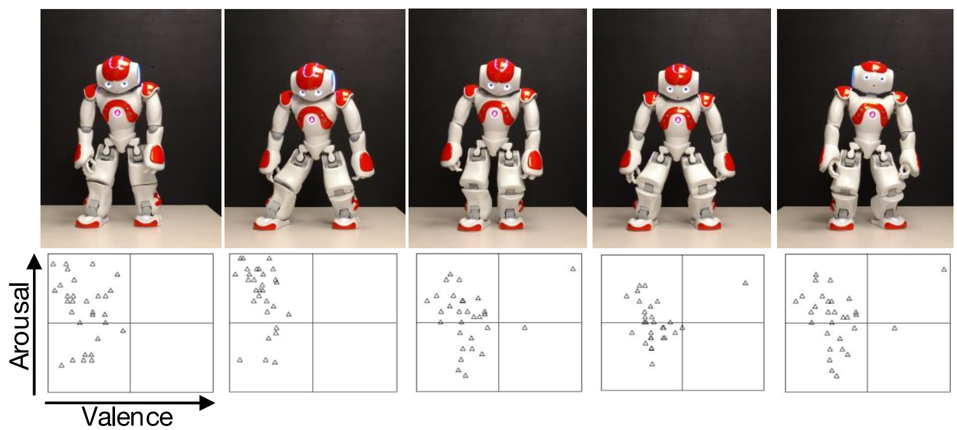
Appendices

Appendix A: Selected Poses for Study 2 and 3

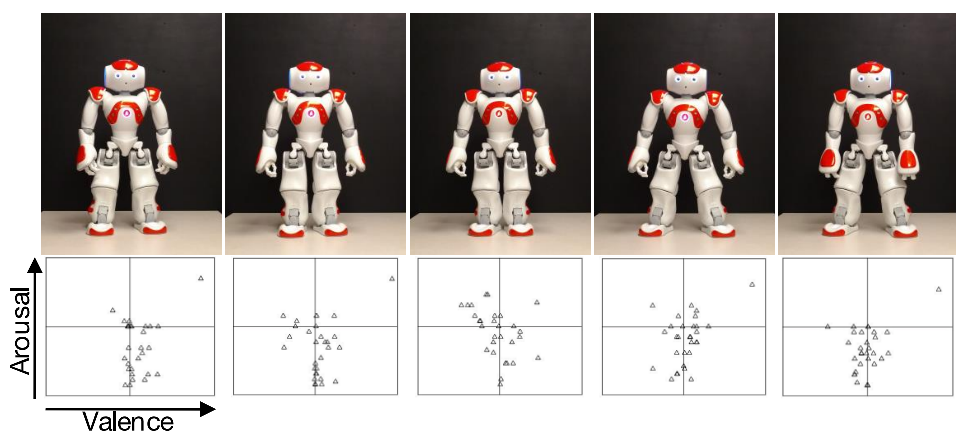
Negative Valence-High Arousal



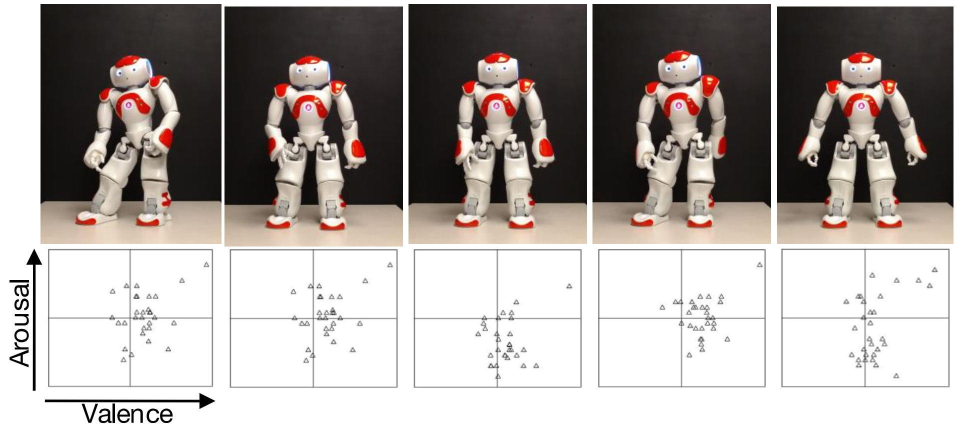
Negative Valence-Low Arousal



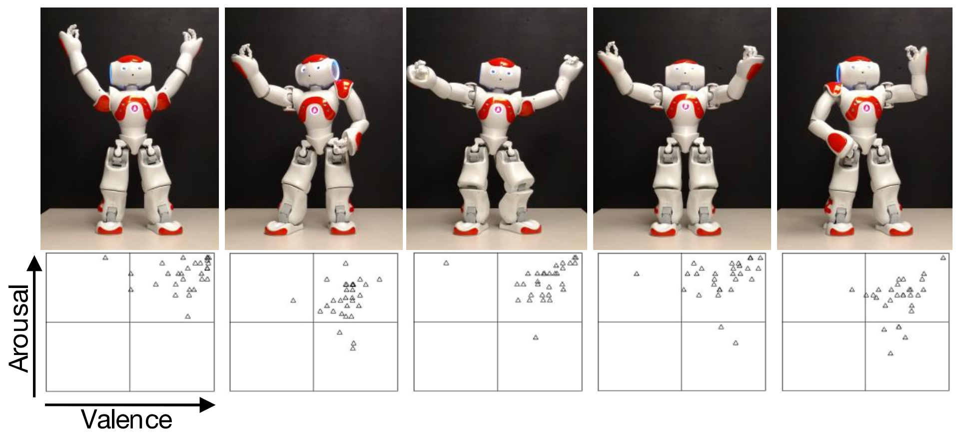
Neutral



Positive Valence-Low Arousal



Positive Valence-High Arousal



Appendix B: Pose Ratings from Study 1 (Selected and Non-Selected)

Negative Valence-High Arousal (NH) Robot Pose Ratings

Study 1 Pose #	Study 2&3 Pose #	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
41		5.27	5.50	5.27	5.25	2.51	1.93	2.08	0.19	0.32	0.20
42		4.36	4.00	6.48	7.00	4.04	3.59	4.33	0.57	0.53	0.49
43	1	4.06	4.00	6.44	6.50	4.56	3.84	4.05	0.79	0.69	0.68
44	2	3.51	3.00	6.98	7.25	3.13	4.18	3.73	0.65	0.76	0.72
45	3	3.96	3.00	8.09	8.50	6.72	6.55	7.76	0.36	0.75	0.62
46	4	3.74	3.00	7.56	8.00	5.43	5.69	7.52	0.71	0.71	0.65
47	5	3.20	2.00	7.49	8.00	8.19	3.34	6.63	0.36	0.34	0.59
48		4.31	4.00	6.92	7.00	6.52	7.02	3.73	0.66	0.57	0.57
49		4.99	5.00	5.83	6.00	1.77	3.75	3.37	0.22	0.86	0.37
50		4.73	5.00	6.35	6.50	2.90	3.70	4.70	0.70	0.70	0.53

Negative Valence-Low Arousal (NL) Robot Pose Ratings

Study 1 Pose #	Study 2&3 Pose #	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
11		3.27	3.00	5.15	5.25	2.27	1.98	2.41	0.58	0.77	0.64
12	6	2.57	2.25	5.81	6.50	0.96	3.10	1.52	0.26	0.35	0.45
13		2.87	3.00	5.52	6.00	3.07	1.45	2.20	0.33	0.51	0.46
14	7	2.33	2.00	6.68	7.50	0.96	1.89	0.66	-0.16	0.00	0.38
15	8	3.39	3.00	5.30	5.50	2.69	2.67	2.99	0.66	0.72	0.58
16		3.40	3.00	5.59	6.00	2.38	3.48	2.26	0.35	0.70	0.41
17	9	3.52	3.00	5.05	5.25	1.84	1.62	1.87	0.50	0.57	0.68
18		5.66	6.00	6.67	7.00	3.83	3.25	3.90	0.22	0.78	0.17
19	10	3.39	3.00	5.30	5.50	2.69	2.67	2.99	0.66	0.72	0.58
20		3.44	3.00	5.99	6.00	3.01	3.00	2.32	0.57	0.41	0.31

Neutral (NE) Robot Pose Ratings

Study 1 Pose #	Study 2&3 Pose #	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
31	11	5.37	5.00	3.80	4.00	0.87	0.87	1.07	0.56	0.76	0.50
32	12	5.15	5.00	3.66	4.00	1.41	1.73	1.70	0.52	0.44	0.58
33	13	4.81	5.00	5.28	4.50	1.83	2.12	1.95	0.06	0.40	0.00
34		5.37	5.00	4.13	4.75	1.57	1.63	1.43	0.72	0.68	0.44
36		4.87	5.00	4.07	4.50	0.69	1.86	2.08	0.37	0.51	0.17
37		5.18	5.00	4.01	4.25	2.21	1.61	1.58	0.47	0.69	0.57
38		5.22	5.00	3.83	4.25	0.86	1.02	1.36	0.20	0.58	-0.01
39	14	5.13	5.00	4.19	4.50	1.56	1.96	1.62	0.35	0.35	0.24
40	15	4.95	5.00	3.57	3.13	0.79	1.58	1.74	0.49	0.45	0.50

Positive Valence-Low Arousal (PL) Robot Pose Ratings

Study 1 Pose #	Study 2&3 Pose #	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
21	16	5.71	6.00	5.14	5.75	2.04	1.61	1.66	0.44	0.63	0.44
22	17	5.63	6.00	4.34	4.25	1.11	1.06	1.89	0.10	0.38	0.11
23		4.68	4.00	5.06	5.50	2.24	2.45	3.53	0.37	0.70	0.28
24	18	5.42	5.00	3.55	3.00	1.50	1.70	0.90	0.32	0.67	0.36
25		5.28	5.00	4.44	4.50	1.31	1.80	1.22	0.16	0.24	-0.08
26		5.51	5.75	7.46	8.00	6.60	6.94	6.03	0.30	0.28	0.29
27	19	6.04	6.00	5.06	5.00	1.09	1.64	2.03	0.30	0.35	0.28
28		5.17	5.00	4.24	4.00	0.94	1.77	1.78	0.28	0.51	0.15
29		4.93	5.00	4.52	5.00	1.75	1.57	2.25	0.61	0.67	0.42
30		5.16	5.00	5.19	5.00	1.60	2.21	3.11	0.63	0.73	0.54
35	20	5.53	5.00	4.44	5.00	1.84	1.15	1.67	0.74	0.69	0.69

Positive Valence-High Arousal (PH) Robot Pose Ratings

Study 1 Pose #	Study 2&3 Pose #	Valence Mean	Valence Median	Arousal Mean	Arousal Median	Block 1 Valence Variance	Block 2 Valence Variance	Block 3 Valence Variance	Blocks 1:2 Correlation	Blocks 2:3 Correlation	Blocks 1:3 Correlation
1	21	7.82	8.00	7.99	8.00	2.38	2.89	3.79	0.10	0.43	0.74
2		6.25	7.00	7.29	7.50	4.29	4.22	4.16	0.50	0.77	0.50
3		5.97	6.25	5.94	6.00	2.03	2.04	3.34	0.46	0.35	0.30
4		6.18	7.00	7.36	7.50	3.15	4.13	6.39	0.38	0.62	0.40
5	22	6.65	7.00	6.38	7.00	1.36	1.29	1.96	0.11	0.29	-0.02
6	23	7.38	7.50	7.53	8.00	1.36	2.00	3.73	0.71	0.64	0.50
7	24	7.18	8.00	7.73	8.00	4.97	3.45	3.05	0.32	0.39	0.17
8	25	6.75	7.00	6.47	7.00	1.32	1.29	1.22	0.55	0.61	0.69
9		5.64	6.00	5.31	6.00	2.12	1.90	3.36	0.34	0.62	0.36
10		5.24	5.00	5.36	5.75	1.70	1.54	2.95	0.40	0.38	0.45

Appendix C: Mean Ratings for Emotion Categories by Study

Negative Valence-High Arousal (NH) Robot Poses

		Valence				Arousal			
Study	Condition	Present 1	Present 2	Present 3	Overall Mean	Present 1	Present 2	Present 3	Overall Mean
1	Physical Low	3.82 (1.66)	3.56 (1.76)	3.72 (1.76)	3.69 (1.71)	7.20 (0.94)	7.39 (1.01)	7.38 (0.99)	7.32 (0.97)
2	Physical Low	4.37 (0.91)	4.05 (1.01)	4.19 (1.16)	4.21 (1.03)	6.84 (0.95)	6.98 (0.82)	6.83 (0.71)	6.90 (0.83)
2	Physical High	4.38 (1.10)	4.13 (1.19)	4.13 (1.16)	4.19 (1.14)	6.54 (0.98)	6.59 (1.08)	6.68 (1.09)	6.60 (1.04)
3	Physical Low	5.11 (1.45)	5.04 (1.37)	---	5.07 (1.35)	7.51 (0.71)	7.46 (0.67)	---	7.49 (0.64)
3	Physical High	4.39 (1.45)	4.30 (1.53)	---	4.35 (1.42)	7.14 (0.85)	7.00 (0.94)	---	7.07 (0.82)
3	Virtual Low	4.58 (1.42)	4.44 (0.99)	---	4.51 (1.06)	7.43 (0.63)	7.23 (0.69)	---	7.33 (0.56)
3	Virtual High	4.07 (1.27)	4.19 (1.40)	---	4.13 (1.26)	6.86 (0.90)	6.92 (1.05)	---	6.89 (0.92)

Negative Valence-Low Arousal (NL) Robot Poses

		Valence				Arousal			
Study	Condition	Present1	Present 2	Present 3	Overall Mean	Present 1	Present 2	Present 3	Overall Mean
1	Physical Low	2.92 (0.93)	3.14 (1.24)	3.05 (1.01)	3.04 (1.06)	5.71 (1.71)	5.53 (1.67)	5.72 (1.74)	5.65 (1.69)
2	Physical Low	3.45 (0.90)	3.65 (0.88)	3.85 (0.84)	3.65 (0.88)	3.75 (1.34)	3.85 (1.31)	3.95 (1.27)	3.86 (1.30)
2	Physical High	3.17 (0.67)	3.45 (0.83)	3.79 (0.61)	3.48 (0.74)	3.76 (1.34)	3.89 (1.50)	3.89 (1.18)	3.84 (1.33)
3	Physical Low	2.76 (0.64)	2.95 (0.76)	---	2.85 (0.65)	3.68 (1.20)	3.93 (1.37)	---	3.80 (1.21)
3	Physical High	2.91 (0.99)	2.89 (1.01)	---	2.90 (0.93)	3.53 (1.51)	3.42 (1.40)	---	3.48 (1.40)
3	Virtual Low	2.60 (0.78)	2.79 (0.68)	---	2.70 (0.67)	3.98 (1.77)	4.36 (1.78)	---	4.17 (1.72)
3	Virtual High	2.48 (0.72)	2.67 (0.82)	---	2.58 (0.72)	3.63 (1.75)	3.81 (1.57)	---	3.72 (1.63)

Neutral (NE) Robot Poses

		Valence				Arousal			
Study	Condition	Present 1	Present 2	Present 3	Overall Mean	Present 1	Present 2	Present 3	Overall Mean
1	Physical Low	5.04 (0.73)	5.12 (1.05)	5.05 (0.99)	5.07 (0.92)	3.85 (1.34)	4.16 (1.42)	3.90 (1.54)	3.97 (1.43)
2	Physical Low	4.97 (0.88)	5.20 (0.85)	5.31 (0.67)	5.14 (.81)	4.41 (1.46)	4.64 (1.30)	4.50 (1.48)	4.53 (1.40)
2	Physical High	5.13 (0.52)	5.19 (0.53)	5.09 (0.57)	5.16 (0.54)	4.27 (1.30)	4.39 (1.23)	4.39 (1.39)	4.35 (1.29)
3	Physical Low	4.96 (0.81)	5.01 (0.75)	– – –	4.98 (0.71)	3.88 (1.34)	4.30 (1.34)	– – –	4.09 (1.30)
3	Physical High	4.83 (0.89)	4.72 (0.74)	– – –	4.77 (0.79)	3.84 (1.50)	3.67 (1.55)	– – –	3.76 (1.49)
3	Virtual Low	4.52 (0.68)	4.73 (0.57)	– – –	4.62 (0.58)	3.57 (1.23)	3.86 (1.40)	– – –	3.71 (1.22)
3	Virtual High	4.50 (0.89)	4.76 (0.68)	– – –	4.63 (0.67)	3.13 (1.48)	3.31 (1.41)	– – –	3.22 (1.38)

Positive Valence-Low Arousal (PL) Robot Poses

		Valence				Arousal			
Study	Condition	Present 1	Present 2	Present 3	Overall Mean	Present 1	Present 2	Present 3	Overall Mean
1	Physical Low	5.68 (0.75)	5.71 (0.93)	5.61 (0.97)	5.67 (0.88)	4.47 (1.33)	4.52 (1.23)	4.53 (1.36)	4.51 (1.29)
2	Physical Low	5.45 (0.87)	5.59 (0.81)	5.75 (0.77)	5.60 (0.82)	4.57 (1.28)	4.93 (1.23)	5.03 (1.25)	4.85 (1.25)
2	Physical High	5.53 (0.49)	5.79 (0.82)	5.51 (0.82)	5.61 (0.73)	4.59 (1.07)	4.74 (1.36)	4.78 (1.36)	4.7 (1.26)
3	Physical Low	5.26 (1.05)	5.35 (0.91)	– – –	5.30 (0.94)	4.40 (1.29)	4.68 (1.30)	– – –	4.54 (1.21)
3	Physical High	5.37 (0.83)	5.32 (0.82)	– – –	5.34 (0.77)	4.45 (1.35)	4.43 (1.46)	– – –	4.44 (1.34)
3	Virtual Low	4.76 (0.80)	4.79 (0.61)	– – –	4.77 (0.62)	3.97 (1.25)	4.33 (1.24)	– – –	4.15 (1.11)
3	Virtual High	4.85 (0.88)	5.16 (1.02)	– – –	5.00 (0.89)	3.46 (1.30)	3.73 (1.25)	– – –	3.59 (1.20)

Positive Valence-High Arousal (PH) Robot Poses

		Valence				Arousal			
Study	Condition	Present 1	Present 2	Present 3	Overall Mean	Present 1	Present 2	Present 3	Overall Mean
1	Physical Low	7.23 (1.00)	7.08 (0.99)	7.17 (1.13)	7.16 (1.03)	7.24 (0.90)	7.11 (1.11)	7.29 (1.10)	7.22 (1.03)
2	Physical Low	7.01 (0.78)	7.09 (0.79)	7.10 (0.91)	7.07 (0.80)	6.24 (1.37)	6.70 (1.13)	6.73 (1.42)	6.57 (1.32)
2	Physical High	7.10 (0.78)	7.08 (0.86)	6.91 (0.90)	7.03 (0.84)	6.35 (1.07)	6.44 (1.30)	6.60 (1.20)	6.46 (1.18)
3	Physical Low	7.18 (0.91)	7.00 (1.34)	– – –	7.09 (1.05)	7.26 (0.94)	7.28 (0.99)	– – –	7.27 (0.93)
3	Physical High	7.02 (0.78)	6.57 (1.06)	– – –	6.79 (0.85)	6.94 (1.12)	6.92 (1.10)	– – –	6.93 (1.05)
3	Virtual Low	6.67 (0.79)	6.61 (0.87)	– – –	6.64 (0.72)	6.78 (1.16)	6.60 (1.48)	– – –	6.69 (1.28)
3	Virtual High	6.46 (1.20)	6.41 (1.23)	– – –	6.43 (1.15)	5.98 (1.13)	6.00 (1.54)	– – –	5.99 (1.26)

Appendix D: Robot Attitudes Questionnaire

Please rate your reaction to the robot used in this experiment based on these scales:

1. Fake	1	2	3	4	5	Natural
2. Machinelike	1	2	3	4	5	Humanlike
3. Unconscious	1	2	3	4	5	Conscious
4. Artificial	1	2	3	4	5	Lifelike
5. Moving rigidly	1	2	3	4	5	Moving elegantly
6. Dislike	1	2	3	4	5	Like
7. Unfriendly	1	2	3	4	5	Friendly
8. Unkind	1	2	3	4	5	Kind
9. Unpleasant1	1	2	3	4	5	Pleasant
10. Awful	1	2	3	4	5	Nice
11. Incompetent	1	2	3	4	5	Competent
12. Ignorant	1	2	3	4	5	Knowledgeable
13. Irresponsible	1	2	3	4	5	Responsible
14. Unintelligent	1	2	3	4	5	Intelligent
15. Foolish	1	2	3	4	5	Sensible

(Godspeed Scales – Bartneck, Kulić, Croft, & Zoghbi, 2009)

Please mark the following questions with a score 1 to 5 (1: minimal intensity, 5 maximum intensity):

I felt familiar with the robot:

16. In the beginning of the experiment

1 2 3 4 5

17. After the first block of poses

1 2 3 4 5

18. After the second block of poses

1 2 3 4 5

19. After the task

1 2 3 4 5

(Baddoura, Venture, Matsukata, 2012)

Please mark on the following scale:

1 (strongly disagree) 2 (disagree) 3 (undecided) 4 (agree) 5 (strongly agree)

20. If I saw someone verbally harassing another human I would be upset.

21. If I saw someone verbally harassing this robot I would be upset.

22. If I saw someone causing physical harm to another human I would be upset.

23. If I saw someone causing physical harm to this robot I would be upset.

24. This robot has feelings.

In general:

25. I feel that in the future, robots will be commonplace in society.

26. I would be comfortable with a robot helping me in a hospital.

27. I would appreciate the companionship of a robot while I was staying in a hospital.

28. I would be comfortable receiving care from a robot nurse if I received the care more quickly than from a human nurse.

29. I would be comfortable with a robot nurse assessing the severity of my injury or illness.

30. I would be comfortable with a robot assisting my elderly relatives if they were sick or injured.

31. I feel that if I depend on robots too much, something bad might happen.

32. If robots developed into living beings something bad might happen.

33. I am concerned that robots would be a bad influence on children.

Familiarity with Robots and Virtual Reality:

34. I know a lot about robots.

35. I have interacted with physical robots.

36. If you have interacted with physical robots, approximately how many times?

____

37. I know a lot about virtual reality.

38. I have experienced virtual reality.

39. If you have experienced virtual reality, approximately how many times?

____

40. This virtual reality experience felt realistic. (VR ONLY)

41. The robot used in this experiment physically exists. (VR ONLY)

42. The robot used in this experiment had a gender. (Study 3 ONLY)

43. If you perceived the robot as having a gender, which gender was it? (Study 3 ONLY)

(1: No gender, 2: Female, 3: Male, 4: A gender not listed here)