ROBOTIC FACES: EXPLORING DYNAMICAL PATTERNS OF SOCIAL INTERACTION BETWEEN HUMANS AND ROBOTS

Casey Bennett

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements for the degree

Doctor of Philosophy

in the School of Informatics and Computing

Indiana University

May 2015

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____

Selma Sabanovic, Ph.D.

_____

Randall Beer, Ph.D.

_____

Kay Connelly, Ph.D.

_____

David Crandall, Ph.D.

April 13, 2015

Copyright © 2015

Casey Bennett

For Dhari, for always being the light in the darkness.

Casey Bennett

EXPLORING DYNAMICAL PATTERNS OF SOCIAL INTERACTION BETWEEN HUMANS AND ROBOTS

The purpose of this dissertation is two-fold: 1) to develop an empirically-based design for an interactive robotic face, and 2) to understand how dynamical aspects of social interaction may be leveraged to design better interactive technologies and/or further our understanding of social cognition.

Understanding the role that dynamics plays in social cognition is a challenging problem. This is particularly true in studying cognition via human-robot interaction, which entails both the natural social cognition of the human and the "artificial intelligence" of the robot. Clearly, humans who are interacting with other humans (or even other mammals such as dogs) are cognizant of the social nature of the interaction – their behavior in those cases differs from that when interacting with inanimate objects such as tools. Humans (and many other animals) have some awareness of "social", some sense of other agents. However, it is not clear how or why.

Social interaction patterns vary across culture, context, and individual characteristics of the human interactor. These factors are subsumed into the larger *interaction system*, influencing the unfolding of the system over time (i.e. the dynamics). The overarching question is whether we can figure out how to utilize factors that influence the dynamics of the social interaction in order to imbue our interactive technologies (robots, clinical AI, decision support systems, etc.) with some "awareness of social", and potentially create more natural interaction paradigms for those technologies.

In this work, we explore the above questions across a range of studies, including lab-based experiments, field observations, and placing autonomous, interactive robotic faces in public spaces. We also discuss future work, how this research relates to making sense of what a robot "sees", creating data-driven models of robot social behavior, and development of robotic face personalities.

Keywords: Affective Communication; Emotion; Facial Expressions; Human-Robot Interaction; Robot Design; Social Interaction

Table of Contents

Chapter 1

Introduction

1.1 Problem

At its core, the purpose of this dissertation is two-fold: 1) to develop an empirically-based design for an interactive robotic face, and 2) to understand how dynamical aspects of social interaction may be leveraged to design better interactive technologies and/or further our understanding of social cognition. In this chapter, we explain the importance of delving into these two challenges. The motivation for such work starts from a basic premise: that social interaction is a "system" that *subsumes* the individual interactors and components of any interaction. Even when we interact with technology, much of what shapes the interaction goes beyond the design of the technology itself. Studying such a system requires the ability to rigorously and consistently manipulate aspects of the system. Technologies such as interactive robots (e.g. human-robot interaction, HRI) can afford such abilities, but only if robotic technologies are designed in an empirical way, allowing for replicable experimentation, i.e. "robotic science."

1.2 Question

The primary driving question here is: what makes an interaction "social"? There are corollaries to this primary question: what is social cognition? Where does it come from? Why do we humans and other animals exhibit such a capacity? Why does interacting with other items, such as tools or technology, not exhibit "social" features? Social interaction is a dynamic process influenced by a number of factors, but what are the factors, and what role do they play? Can we imbue our technologies with features or dynamical properties that make them interact more socially, and/or that encourage people to ascribe more social characteristics to them?

It is outside the scope of the current work to answer all those questions. Indeed, it may take lifetimes of work to ever do so, if we even can do so. Rather, our focus here is on beginning to drive

towards potential lines of evidence that may shed light on them. *The roots of sociality*. We do so from the perspective of human-robot interaction (HRI), using robots as tools to study social cognition, and the dynamics thereof, affords particular benefits. For instance, one advantage of using a robot as one of the interactors (averse to two humans) is that it allows us to "get inside the mind" of one of the interactors and purposely manipulate the interaction in a consistent manner across human subjects.

Understanding the role that various factors – such as environmental context or culture – play in social cognition is a challenging problem. This is particularly true in studying cognition via HRI, which entails both the natural social cognition of the human and the "artificial intelligence" of the robot. Clearly, humans interacting with humans have some sort of awareness of the social nature of the interaction – their behavior in those cases differs from that when interacting with inanimate objects such as tools. Humans have some "awareness of social", some sense of other agents (Froese & Di Paolo, 2011). However, it is not clear how or why, short of positing a special "module" of social cognition in the mammalian brain. Moreover, the synchronization that occurs between human interactors, whether the product of coupled oscillators in some dynamical system or otherwise, presents challenges. Constructing an emergent adaptivity into a robot in order to enable it to step into such a dynamical system (as one of the interactors) demands robot behavior that is emergent itself, i.e. "designed for emergence." However, as noted elsewhere, we have no idea how to systematically do so (Pfeiffer & Bongard, 2007). Hemmed in, we still are, by our Von Neumann computing paradigm.

A number of papers exist that have explored the dynamics of *interaction patterns* in human-robot interaction through a variety of temporal models – oscillating dynamical systems, Markov decision processes, etc. (e.g. Michalowski et al., 2007; Kahn et al., 2008; Kahn et al., 2010b). The work proposed here builds on this, exploring how such interaction patterns vary across culture, context, and individual characteristics of the human interactor. These factors are subsumed into the larger *interaction system*, influencing the unfolding of the system over time (i.e. the dynamics). Such influences should be detectable in the way people respond to the robot, and shape common patterns in the interaction data. In other words, they should be inherent in the sociality of the interaction. The overarching question is

whether we can figure out how to utilize dynamical aspects of the social interaction in order to imbue our interactive technologies (robots, clinical AI, decision support systems, etc.) with some "awareness of social", and potentially create more natural interaction paradigms for those technologies.

In plain language, interaction is a system, and if we want to design interactive technologies, we are really designing the system, not the technology itself:

> *"We must go beyond the view that defines interaction as simply the spatio-temporal coincidence of two agents that influence each other. We must move towards an understanding of how their history of coordination demarcates the interaction as an identifiable pattern with its own internal structure, and its own role to play in the process of understanding each other and the world." (De Jaegher & Di Paolo, 2007, pp. 492)*

1.3 Minimalist Robotics and "Robotic Science" as an Approach for Studying Social Interaction

A principle goal in this work is taking an empirical approach to designing robots, with a particular focus on robots for social interaction. The goal is to empirically develop socially-interactive robots, *from the ground up*, as well as to develop inexpensive and replicable robotic faces for experimental purposes.

In keeping with the empirical approach, we adopt a minimalist approach to robotic face design, grounded in over a half-century of psychological and computer science research on emotions and facial expressions (Bennett & Šabanović, 2014). The entire premise of that work (Ekman, 2009; Nelson & Russell, 2013; Pantic, 2009; Cohn, 2010) is that people are only attending to a small number of critical moving points/lines to detect emotion in faces. A minimalist approach to robot design understands that many "design features" may be superfluous for tasks such as basic social interaction; indeed, they may even be problematic in the sense of conflating factors for the research questions we are trying to answer.

Moreover, a minimalist approach is an explicit move towards developing *robotic science*, rather than approaching robots as simply an engineering or aesthetic endeavor. From that perspective, the costs of robots that "look cool" are, in fact, prohibitive to the replicability necessary for good science. One of the major challenges for robotic science is that many people are doing research with one-off, $100k robots, which makes it difficult to near-impossible for anyone to replicate. Our minimalist approach here

is the polar opposite: all the design schematics for the robotic face we use, including both hardware assembly instructions and software programming code, are available online. All the materials we use are easily accessible online, for less than $200 per robot. It takes advantage of newly emerging technologies, such as 3D printing for rapid prototyping. A minimalist approach encourages such replicability and accessibility. Replicability lies at the heart of the scientific process

The minimalist approach affords particular advantages to our empirical approach. The aim was to start as simple as possible, and build up from there in a grounded manner. Each step, as can be seen in subsequent chapters, builds on the previous one, moving from simple designs and simple questions to more complex 3D printed designs and deeper questions. An empirical approach demands such a systematic method – we cannot simply leap ahead without making a large number of assumptions. Rather, the goal is to minimize such assumptions, providing evidence for each hypothesis along the way.

To summarize, from a scientific standpoint, a minimalist approach to robotic design enhances replicability of the results via providing a robotic platform that can be exactly replicated elsewhere with minimal cost/effort (construction manual and programming code links are provided in Chapter 2). Replicability lies at the heart of the scientific process. Currently, many robot designs fail to meet such criteria, which makes it more difficult to build a sound body of scientific evidence for robotic design, human-robot interaction, and the like. Moreover, stripping the robotic face down from any conflating or superfluous factors (e.g. randomly adding ears, feathers, or other aesthetic features) that could affect human perceptions of the robot is key to analyzing the factors we are actually trying to study. In short, the work described here is focused on the *science* of designing a robotic face and its interactive behavior. A true *robotic science* demands such a particular methodological approach.

This goes hand-in-hand with an empirical approach to studying social interaction. Since our goal is to meticulously manipulate certain factors of the interaction, using the robotic face as one of the interactors, taking an empirical approach to robotic design enables us to more effectively create such experimental manipulation. Indeed, the relationship between social interaction experimentation and robotic design becomes an iterative process.

1.4 Social Construction: Social Interaction as a System

As noted above, we take as a basic premise here that social interaction is a "system" that *subsumes* the individual interactors and components of any interaction. What makes an interaction "social" is entailed in the *construction of social*: somewhere, emergent in our cognitive system, are a set of processes that construct the social from a number of perceptual cues and precepts. Such cues may arise from aspects outside the interactors themselves (e.g. context), playing an integral role in shaping it, as well as shaping the perceptions of the interactors. Moreover, the internal cognitive processes based on these perceptual cues are influenced by broader societal conditioning that may exist outside the scope of the interaction itself, e.g. culture (e.g. Hall, 1977; Shore, 1996; Nisbett, 2001, 2003). In short, *social construction* is a process that operates beyond the scope of the interactors themselves, and at multiple timescales beyond the time of the interaction itself.

As such, a natural avenue for exploring social interaction and social cognition is by attempting to manipulate those factors experimentally. Such manipulation holds potential to reveal the way those factors impact the "social" aspect of interaction, and more broadly how sociotechnical concepts like social construction relate to intrapersonal concepts like social cognition. Indeed, this strikes at the core of the theory of the reflexive nature between society at large and social cognition of the individual (Froese & Di Paolo, 2011). The *social interaction system* is a combination of internal and external in that sense, without a clear delineation between the two, and what is "real" in terms of the experience of each interactor is subject to debate (without resorting to debates over "qualia", phenomenology, and the like). But setting aside the question of "what is real", we can still posit one thing: *It is an experience*. And that experience – that system – involves the projection of the internal realities of the interactors, a perceptual blending of reality and illusion. Somewhere therein, lies the truth.

Accepting that the "experience" of social interaction is some perceptual combination of reality and illusion, that the social interaction system is comprised of many factors beyond the interactors (or even the interaction) itself, frees us from a number of constraints when it comes to designed interactive

technology. This may afford particular opportunities to imbue our technologies with some "awareness of social" based on those factors, and use them to create more naturalistic interaction between humans and technology. What the human interactor experiences may not be as based on the "reality" as a designer might assume. Indeed, we may be able to take advantage of such illusion by manipulating perceptual cues, without necessarily altering reality

From a technology design standpoint, it is critical to understand what role external factors such as culture and context may play in the dynamical process of perception formation during interaction (Šabanović, Bennett, & Lee, 2014). However, given that culture is dynamic and constantly in flux itself, it may not make sense to design robots/technology in toto for specific cultures, but rather to design robots/technology that are sensitive and adaptive to particular cultural factors and their temporal nature. In that vein, certain interpretations of the data here (e.g. Section 4.4.2) are driven by this philosophy, the dynamical viewpoint, and integrated into a much larger body of work around situated, embodied, dynamical (SED) approaches to cognitive systems ((Barsalou et al., 2006; Beer, 2000; Clark, 2013).

1.5 Chapter Layout

This work takes an explicit, iterative empirical approach. Each chapter builds on the previous one, moving from simple designs and simple questions to more complex 3D printed designs and deeper questions. This empirical approach lends itself naturally to a story about creating a robotic face, which, as we will see, goes far beyond the robot itself.

We start by laying out the background design philosophy of the robotic face platform utilized here (Chapter 2) at a high level, including details of its construction and the companion digital avatar used in several studies.

Chapter 3 systematically evaluates the facial expressions produced by the robotic face, and the effects of certain factors (e.g. added neck motion, degree of expression) on human perceptions thereof. It provides the basic validation for the facial expressions.

Chapter 4 explores the effect that environmental context has on human perception of robotic facial expressions, and whether such effects vary across culture. It also explores whether inducing such context effects might enable culture-neutral models of robots and affective interaction

Chapter 5 puts a twist on the previous chapter's question - what would happen if context congruency was varied – if the emotion expressed by the context was sometimes congruent, sometimes incongruent, with the robotic expressions? Do such effects vary across culture?

Of course, examining social interaction and human perceptions of the robotic face in a lab setting is useful in many ways, allowing us to conduct a number of controlled experiments to explore the effects of different factors. The question of course exists as to how people would interact with such a robotic face in naturalistic "in-the-wild" settings. In Chapter 6, we do just this, placing an autonomous, interactive robotic face in a public art museum for nearly a month. Can we identify common interaction "schemas" emergent in the data?

Chapter 7 provides a direct comparison of lab-setting interaction patterns and naturalistic-setting interaction patterns. The question is how such interaction patterns from a naturalistic setting compare to those from the lab. Are they the same, or are they different? And what implications might that hold for the way we study social interaction via robots and HRI?

Finally, Chapter 8 discusses future work, detailing a number of ongoing projects based on the research up to this point. This includes moving from analyzing/identifying behavioral patterns of interaction to developing models to guide future robot social behavior, largely based on the temporal dynamics of such interaction. We also explore various machine learning approaches for making sense of what the robot "sees" in its sparse visual data (averse to computationally-intensive algorithms for face detection and the like). Lastly, we discuss using evidence from human psychological research, such as that of Borderline Personality Disorder, as clues towards the development of robotic face personalities.

Chapter 2

The Robotic Face Platform

This chapter provides a general overview of the robotic face platform used throughout the rest of the chapters and studies in this work. The philosophy, capabilities, and technical details behind the development of an empirically-grounded robotic face are detailed. Some details specific to particular chapters/studies are related in those subsequent chapters – this chapter provides a general overview.

Abstract. None

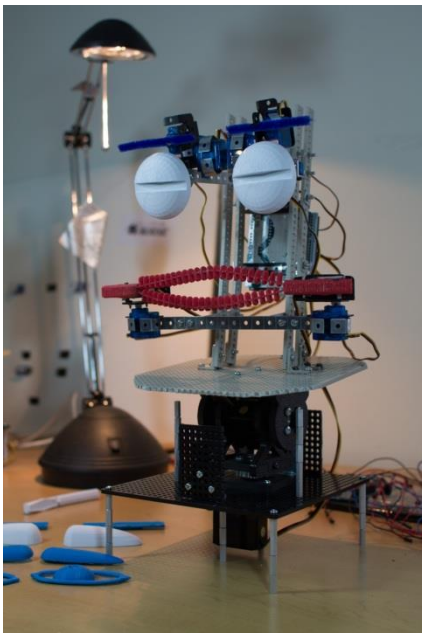
2.1 Robot Design Overview

The research detailed throughout this manuscript makes use of an interactive robotic face (MiRAE – Minimalist Robot for Affective Expression) capable of basic social behavior (Figure 2.1). The robotic face can detect faces and motion, respond to people, make facial expressions, and so forth. It has a basic visual attention system, mechanisms for altering the rhythm of its behaviors, and the ability to track environmental stimuli both relative to its sensory (retinotopic) and motor (spatiotopic) coordinates. The advantage of using a robot as one of the interactors (averse to two humans) is that it allows us to "get inside the mind" of one of the interactors and purposely manipulate the interaction in a consistent manner across human subjects. Studies with the same robot have validated human recognition capability of its facial expressions across multiple cultures and contexts, and explored free-form social interactions with people in a public museum exhibit (Bennett & Šabanović 2014; Bennett et al. 2014; Bennett & Šabanović 2015), which are detailed in the forthcoming chapters.

We provide a general overview of the robotic platform here, and the philosophy behind it, with details specific to different experiments provided in each relevant chapter.

MiRAE was designed to be an empirically grounded robotic face. Throughout subsequent chapters, we build from simple avatar faces, to embodied robotic faces, to context-enhanced facial expressions, and eventually to their applications to human-robot social interaction, and the dynamics thereof. In keeping with the empirical approach, we adopt a *minimalist* approach to robotic face design, grounded in over a half-century of psychological and computer science research on emotions and facial expressions (Bennett & Šabanović, 2014). The entire premise of that work (Ekman, 2009; Nelson & Russell, 2013; Pantic, 2009; Cohn, 2010) is that people are only attending to a small number of critical moving points/lines to detect emotion in faces. This is the basis for the Facial Action Coding System (FACS), which dominates the emotional facial expression literature and on which many robotic faces – including androids – are based (see Section 3.1.1).

Figure 2.1: MiRAE Robotic Face

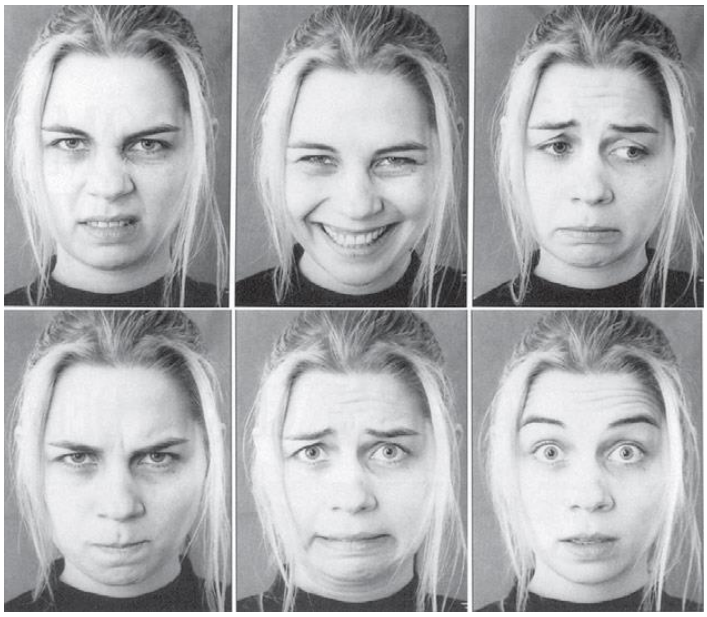


At least within the specific task context of emotional facial expression recognition, there is evidence that many realistic aspects of the face are not necessary, and may indeed even be conflating factors (e.g. by suggesting cultural affiliation, ingroup/outgroup effects). One of our studies here (Bennett & Šabanović, 2014, see Chapter 3) validated that principle in this exact robotic face, providing empirical evidence that simple moving lines work just as well for emotional expressions as more complex facial features (e.g. Kismet [Breazeal, 2003]). Other robotic research, such as Okada's Muu and Kozima's Keepon (Matsumoto et al., 2006; Kozima et al., 2009), further support such minimalism for affective interaction (not to mention Mori's work on the "Uncanny Valley" [Mori, 1970]).

According to the FACS and affective interaction theory, there is a set of six basic emotions - Happy, Sad, Angry, Fear/Worry, Surprise, and Disgust – which are rooted in evolution and displayed using similar features across human cultures (Figure 2.2). These features are referred to as Action Units (AUs, 44 in total), which capture all possible movements of the muscles of the human face. Activation

values for these AUs can be calculated and used to accurately identify the emotion expressed in a given human face, regardless of the idiosyncrasies of the individual face (Ekman & Friesen, 2003; Pantic & Bartlett, 2007). Given that the ability of people to recognize these facial expressions appears to be instinctive, it can be reasoned that humans may use these same AU features to recognize emotions in facial expressions of other people (Ekman, 2009; Calder & Young, 2005). There is evidence, however, that the display and reading of facial expressions may be variable across cultures (Shore, 1996; Yuki, Maddux, & Masuda, 2007; Jack et al., 2009), posing further questions for investigation.

Figure 2.2: Human facial expressions of six basic Ekman emotions

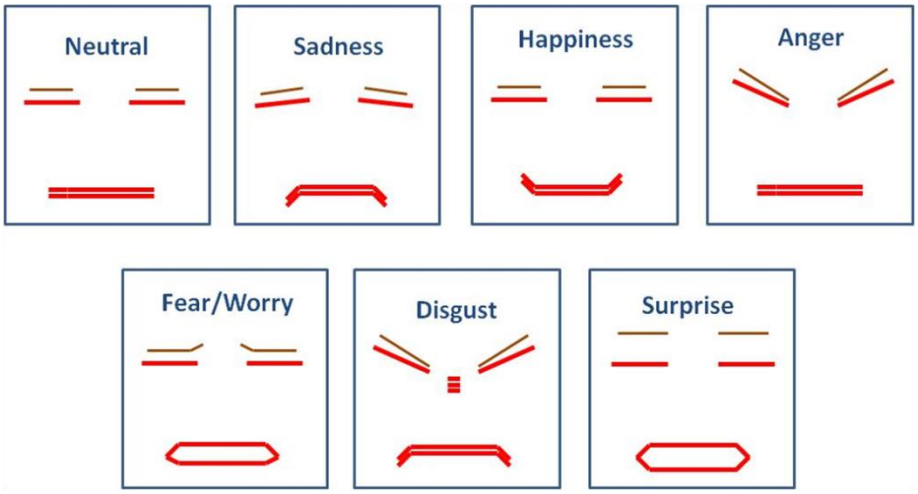


In order (left-to-right, top-to-bottom) – Disgust, Happiness, Sadness, Anger, Fear and Surprise (Cohn, 2010)

Feature selection techniques from machine learning have revealed that a small number of moving points/lines (i.e. a subset of 8-10 AUs) can be used to capture the vast majority of information in human facial expressions (~95%), such as those shown in Figure 2.2 above. This insight has been leveraged over the past several years in the computer vision community to develop automated techniques for classifying human facial expressions via computers (Pantic, 2009; Breazeal, 2003; Anderson & McOwen, 2006). These were translated into the schematic representation shown in Figure 2.3, comprising two principle

linear feature sets: upper (eye/brow) and lower (mouth). Similar sparse cues have also been indicated to play a role in human perception of emotion (Aronoff, Woike, & Hyman, 1992).

Figure 2.3: Schematic Facial Expressions



These simple schematic representations were used as the basis for the embodied robotic face design, as well as the digital avatar version (used in Chapter 3 and 4).

As described in Chapter 3, MiRAE has been shown capable (Bennett & Šabanović 2014, see Chapter 3) of producing higher, or at least comparable, identification accuracy rates (with Westerners) for all expressions as a number of other robotic faces, including Kismet (Breazeal, 2003), Eddie (Sosnowski et al. 2006), Feelix (Canamero & Fredslund, 2001), BERT (Bazo et al., 2010), and the android Geminoid-F (Becker-Asano & Ishiguro, 2011, values from Table 5 therein), as shown in Table 2.1 below (Bennett & Šabanović, 2014). This indicates that a minimalist robotic face such as MiRAE can provide a reliable, replicable, low-cost platform for investigating questions of affective social interaction and facial expression such as those addressed here.

Table 2.1: Robot Face Comparison

Expression	MiRAE (n=30)	Eddie (n=24)	Kismet (n=17)	Feelix (n=86)	BERT (n=10)	Geminoid (n=71)
Happy	97%	58%	82%	60%	99%	88%
Sad	100%	58%	82%	70%	100%	80%
Anger	87%	54%	76%	40%	64%	58%
Fear	43%	42%	47%	16%	44%	9%
Surprise	97%	75%	82%	37%	93%	55%
Disgust	-	58%	71%	-	18%	-
Average[a]	85%	57%	74%	45%	80%	58%

Facial expression identification average accuracy for the six Ekman emotions is shown for several robotic faces (including the one used here, MiRAE). The number of subjects (*n*) is shown for each study as well. Appropriate citations for each are provided in in text. [a]Averages do not include Disgust, since not all studies included it.

While more complex robots designed to capture facial aspects of nonverbal communication such as Kismet (Breazeal, 2003) or Eddie (Sosnowski et al. 2006) already exist, we explore whether simpler facial representations focused on two linear features (upper and lower) and their critical points may be able to convey most of the same information. Other aspects of the face could perhaps be omitted or left as purely aesthetic (and/or economic) choices. This minimalist approach could immensely reduce the complexity of constructing affective robots, or other artificial entities such as digital avatars, allowing for greater flexibility in robot design by freeing up constraints associated with mimicking non-critical aspects of human anatomy, as well as reducing costs. It also holds potential synergy for rapid prototyping in conjunction with new 3D printing techniques (see Chapter 3). Furthermore, such an approach raises interesting cognitive research questions about people's ability to make inferences using incomplete information during social interaction.

2.2 Embodied Face Construction

MiRAE was constructed using inexpensive, easily accessible components in keeping with a minimalist design approach. Principally, physical construction was accomplished using universal metal joints, plastic arms, and universal plates from Tamiya (http://www.tamiyausa.com); 10 sub-micro Hitec servos (http://www.hitecrcd.com); and various servo brackets. Arduino Uno v1.1 microcontrollers (http://www.arduino.cc) were used to create and control functionality of the robotic face. Combinations of servo motors were used to create the needed motion and degrees-of-freedom (DOF): 1 DOF for each eye, 2 DOF for each eyebrow, 2 DOF for the mouth corners/lips, and 2 DOF for the neck. Combined actuation of these simple DOFs could simulate complex motion, such as the parting of lips and baring of teeth. Facial features such as eyes, eyebrows, and the mouth were simulated using colored pipe cleaners affixed using gauge wire for some experiments (e.g. Chapter 3) or created via 3D-printing for other experiments (e.g. Chapter 6). It is also equipped with an onboard camera during social interaction experiments, mounted behind the mouth, for computer vision purposes. More explicit instructions are available the author's main website (http://www.caseybennett.com/Research.html) and at the author's lab website (http://r-house.soic.indiana.edu/mirae/MiRAE_Construction_Manual.pdf). All the programming code, including the C++ libraries and Python (see below), is also available on those websites. In addition, schematics for completely 3D printing the robot-face and head (as well as modifying it for other purposes) will be available from those websites.

In total (including the neck mechanism described in Section 3.2.1.3), the overall cost for the robotic face is approximately $150-175 USD. Total construction time averages roughly 6 hours. 3D printing can be utilized to create more realistic facial features like eyes, depending on the aims of the experiment (e.g. studying minimalist design principles vs. studying social interaction).

MiRAE's programming code is written as a C++/Arduino library, and easily allows facial expressions to be made with varying degrees of motion for each individual facial component (as a variable passed into the function calls). Visual functions are handled via OpenCV, typically on an offboard computer, and communicated serially to the onboard Arduino on the robot-face. This is handled

as a separate set of Python libraries (which also include various cognitive functions to make sense of the visual sensory data, see below). In total, this comprises about 3000 lines of code, including visual, cognitive, and motor functions. In a loose sense (in mammalian brain terms), the offboard computer then can be thought of as the cortex, with the Arduino as the cerebellum. The programming libraries, along with a construction manual for MiRAE, are available from the author's personal website (http://www.caseybennett.com/Research.html) and the lab website (http://r-house.soic.indiana.edu), in order to facilitate experimental replication.

The motor functions (Arduino code) were designed as a three-tiered system. The main program can call functions that specified facial expressions, passing in the *direction* (used to make or undo an expression) and *degree* (continuous value used to determine the strength or degree of the expression – i.e. smaller vs. larger). The facial expression functions in turn call lower functions that move specific facial components given a direction and degree – in essence these facial component functions roughly relate to specific AUs in the FACS. This approach provides several benefits – e.g. it allows for easy extensibility to include new expressions, AUs, or types of facial motion. It also permits a direct linkage between the programming code used to control the robot face and underlying theory about human emotion and facial expression. Also of note, motion (in both the embodied and digital avatar versions) was implemented as *gradated motion*, so that facial expressions occurred over a matter of a few hundred milliseconds (as they would in a real human face), rather than instantaneously. The platform also allows for nuanced control of the expressions and their level of intensity (i.e. degree) in experimental situations.

In terms of the cognitive and visual functions (the Python code), the programming is implemented in a modular fashion (object-oriented), including a visual attention system, affect system, and sensory-motor mapping system. This includes components for controlling visual attention, calculating/regulating its current emotional state, tracking visual stimuli, detecting motion (via optical flow and gray-scale intensity delta), estimating synchronous behavior of detected stimuli, mapping between retinotopic (sensory) and spatiotopic (motor) coordinates, translating detected sensory information into motor movements, and so forth. All of these components serve necessary functions for

the work described in the following chapters. For instance, mapping between retinotopic and spatiotopic coordinates is critical for maintaining stable perceptions of the world, e.g if the robot moves, the retinotopic positions of detected stimuli may shift, but spatiotopic coordinates stay the same (if one turns their head to the left, an object originally centered in the visual field will now appear shifted to the right). In other words, sensory stimuli are tracked relative to motor coordinates (spatiotopic), i.e. sort of like muscle memory. Another example is the visual attention system, which allows the robot to shift its attention across multiple stimuli, based on the behavior of the stimuli and attentional decay. This enables the robot to fluidly interact with multiple stimuli when present. The various systems are summarized in Table 2.2.

Table 2.2: Robot Face Systems

System	Function Name	Description
Communication	serial_setup	Sets of serial communication to Arduino
	comm_arduino	Sends commands to Arduino
	move_arduino	Provides movement information to Arduino
	get_arduino_init_vals	Retrieves initial state values from Arduino
Movement	calc_motion	Calculates motion
	calc_pred_motion	Calculates motion based on prediction of future stimulus location
	calc_saccade_motion	Calculates saccade motion
	calc_motion_corr	Calculates correlations between self-motion and stimulus motion
	robot_move	Moves robot
	move_error	Recovery for any movement errors (i.e. exception handling)
Cognitive	calc_sense	Calculates sensory information
	sensory_mapping	Maps retinotopic sensory data to spatiotopic coordinates
	rev_sensory_mapping	Reverses spatiotopic information back into retinotopic coordinates
	calc_sense_move	Calculates motion of detected stimuli
	predict_face_move	Predicts next location of detected faces
	calc_found	Calculates whether any stimuli detected
	calc_novel	Calculates novelty of stimuli
	regulate	Homeostatic regulatory mechanism of internal robot affect
	prop_attention	Propogates attentional stimuli from previous timestep
	calc_attention	Calculates attentional stimuli
	attention_decay	Attentional decay
	calc_affect	Calculates robot's affective state
	make_expression	Determines facial expression behavior
	clear_exp	Clears facial expressions
Visual	get_frame	Grabs a snapshot image from video stream
	detect_and_draw_faces	Detects relevant stimuli in environment - faces, people, eyes
	idle_detect	Periodically check for new stimulus while robot in idle mode
	slice_face	Slices out face image from larger scene for further processing
	delta_tracker	Detects low-level motion in visual field (e.g. optical flow, intensity gradient deltas)

There are a few additional functions not included in Table 2.2, related to motor control on the Arduino side, as well controlling the degree/strength of made facial expressions.

2.3 Digital Avatar

A digital avatar version was implemented for some of the experiments in Chapter 3 and 4, in order to compare the minimal features in an embodied vs. digital form. The digital avatar version was designed to look virtually identical to the schematic representations (Figure 2.3), and thus by extension as similar to the embodied version as possible. It was implemented using Python 2.7 (www.python.org) and the TkInter toolkit package (http://wiki.python.org/moin/TkInter). Programming was implemented using the same approach as for the embodied face (e.g. three-tiered design, gradated motion) as described above (Section 2. 2).

Chapter 3

Deriving Minimal Features for Human-Like Facial Expressions in Robotic Faces

This chapter explores the facial expression of emotion through a series of studies at the intersection of three fields – computer vision, psychology, and social robotics, focusing on the initial development of an empirically-grounded robotic face.

Abstract. This study (Bennett & Šabanović, 2014) explores deriving minimal features for a robotic face to convey information (via facial expressions) that people can perceive and understand. Recent research in computer vision has shown that a small number of moving points/lines can be used to capture the majority of information (~95%) in human facial expressions. Here, we apply such findings to a minimalist robot face design, which was run through a series of experiments with human subjects (n=75) exploring the effect of various factors, including added neck motion and degree of expression. Facial expression identification rates were similar to more complex robots. In addition, added neck motion significantly improved facial expression identification rates to 100% for all expressions (except Fear). The Negative Attitudes towards Robots (NARS) and Godspeed scales were also collected to examine user perceptions, e.g. perceived animacy and intelligence. The project aims to answer a number of fundamental questions about robotic face design, as well as to develop inexpensive and replicable robotic faces for experimental purposes.

3.1 Introduction

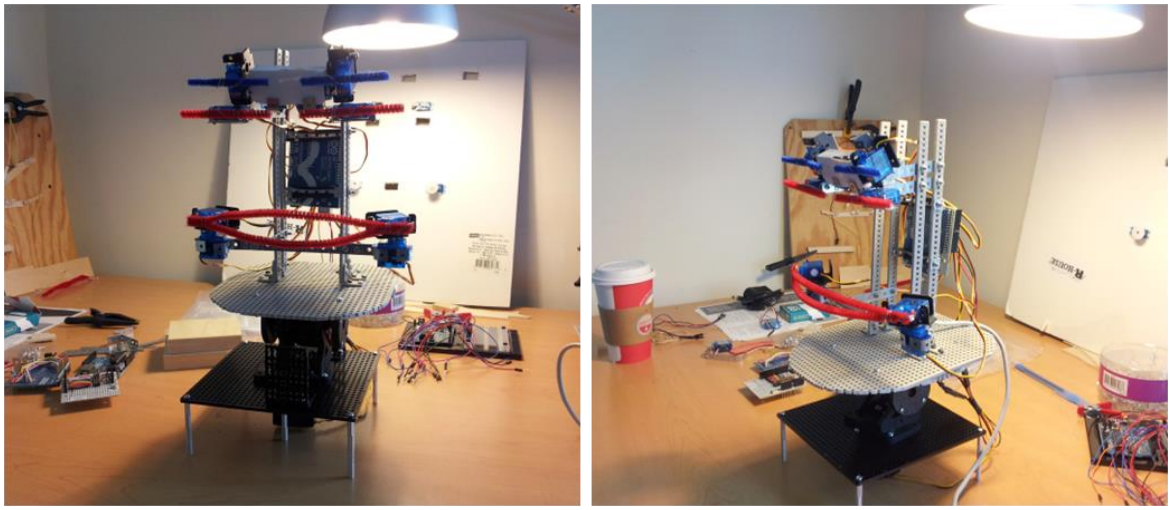
3.1.1 Background/Motivation

This chapter explores the facial expression of emotion through a series of studies at the intersection of three fields – computer vision, psychology, and social robotics, focusing on the initial development of an empirically-grounded robotic face. Extensive psychological research has shown the universality of certain human facial expressions (Ekman & Friesen, 2003), while recent computer vision research suggests that a small number of moving points/lines can be used to capture the majority of information in human facial expressions. This latter insight has been leveraged to develop automated techniques allowing computers to classify human facial expressions with high degrees of accuracy (~95%) for the six basic Ekman emotions: Happy, Sad, Angry, Fear/Worry, Surprise, and Disgust (see Section 3.2.1.1) (Cohn, 2010; Pantic, 2009). In combination, these previous studies suggest that humans may rely on sparse but specific cues to recognize the emotions of others. Inspired by these approaches, we present four experimental studies that seek to "flip" this finding in order to answer questions about human perception and robot design. Can a small number of moving lines in the face of a robot be used to communicate robotic facial expressions to humans in an understandable way? What factors may affect such perception?

This work has implications for the development of interactive robots – such as those used for companionship, collaboration, and therapeutic or assistive purposes – that need not only detect human facial expressions but also express them. While more complex robots designed to capture facial aspects of nonverbal communication such as Kismet (Breazeal, 2003) or Eddie (Sosnowski et al., 2006) already exist, we explore whether simpler facial representations focused on two linear features (upper and lower) and their critical points may be able to convey most of the same information. Other aspects of the face could perhaps be omitted or left as purely aesthetic (and/or economic) choices. This minimalist approach could immensely reduce the complexity of constructing affective robots, or other artificial entities such as digital avatars, allowing for greater flexibility in robot design by freeing up constraints associated with mimicking non-critical aspects of human anatomy, as well as reducing costs. It also holds potential

synergy for rapid prototyping in conjunction with new 3D printing techniques (see Sections 3.2.1.2 and 3.4.3). Furthermore, such an approach raises interesting cognitive research questions about people's ability to make inferences using incomplete information during social interaction. This research direction contributes to the existing agenda of studying the minimal set of cues that evoke social interpretations and responses from human interaction partners (e.g. Okada's Muu [Matsumoto, Fujii,, & Okada, 2006] and Kozima's Keepon [Kozima, Michalowski, & Nakagawa, 2009]).

Figure 3.1: MiRAE – Initial Prototype



Here we describe the development and results from initial research with such a minimalist robotic face – Minimalist Robot for Affective Expressions (MiRAE) – a robot platform we developed capable of performing an array of facial expressions and neck motions (Figure 3.1). MiRAE was designed to use easily accessible components (e.g. Arduino microcontrollers; see Section 3.2.1.2) and requires less than a day of construction time (~6 hours). The project aims to answer a number of fundamental questions about robotic face design, as well as to develop inexpensive and replicable robotic faces for experimental purposes. Our approach also addresses challenges with previous research projects in this area, such as the inclusion of unnecessary confounding variables (e.g. adding ears) or use of custom-made components that limit experimental replicability in the design of robotic faces. The broader goal of this approach is to

create a well-documented research platform that can serve as both a material and empirical contribution to the science of human-robot interaction (HRI), robot design, and cognitive research.

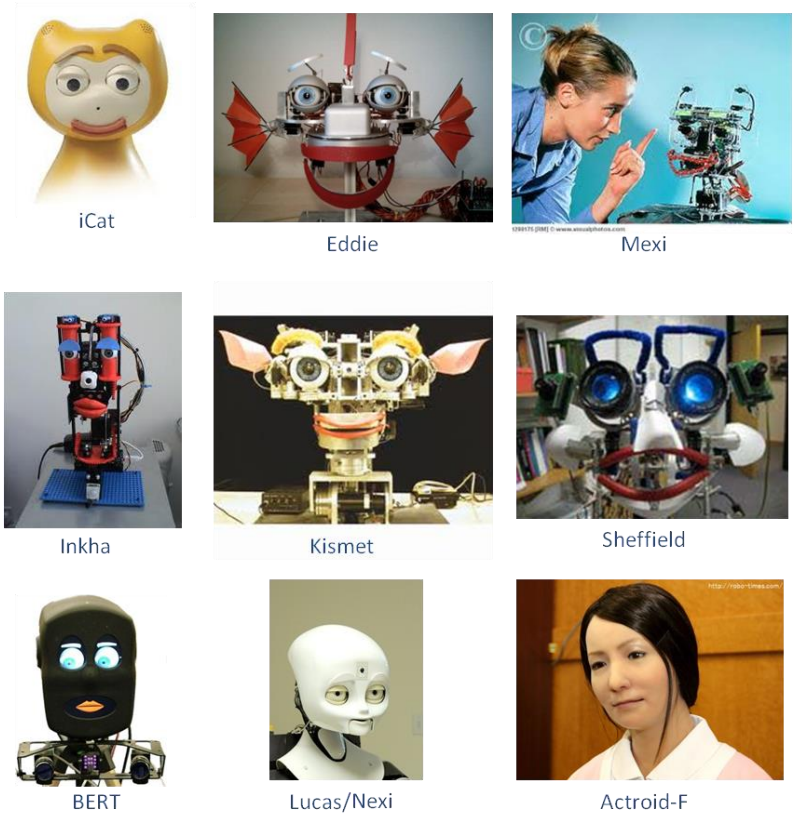
3.1.2 Robot Face Overview

There have been numerous attempts at designing robotic faces over the last thirty years. Many early attempts (pre-2000) are thoroughly reviewed by Fong et al. (2003). We describe various research efforts since 2000 below. These include a wide range of designs, from humanoid faces to animalistic faces to pure iconic/abstract faces. It has been argued elsewhere that iconic/abstract faces are easier to identify with for a broader range of people (Blow et al., 2006), but that they may not elicit the same visceral response as more humanoid robots (Chaminade et al., 2010). However, challenges also exist with humanoid robotic faces that seek near-replication of human facial features and affect. Their close resemblance to human faces engenders certain expectations in human observers that – when such robotic faces fail to achieve complete human-like behavior – triggers a strong negative response, a.k.a. the infamous "uncanny valley" (Mori, 1970; MacDorman et al., 2009). More broadly, there have been a few recent attempts to identify critical dimensions for robotic face design, though these relied largely on meta-reviews of existing robots rather than empirical research (Blow et al., 2006; Di Salvo et al., 2012), in contrast to the work described here.

Several advances have occurred in robotic face design in the last dozen or so years (Figure 3.2) both in terms of physical and computational design, largely based on the basic Ekman emotions and the Facial Action Coding System (see Section 3.2.1.1). A number of these newer robotic faces also underwent some degree of rigorous experimental testing. One example is Kismet, a fully embodied robotic face developed by Breazeal at MIT (2003). Kismet featured not only a sophisticated design capable of an array of facial motions, but also a complex artificial emotion system (see Section 3.1.3) that enabled seemingly naturalistic interaction with its environment (including humans) based on its emotional response to environmental stimuli. Eddie, developed by Sosnowski et al. (2006), is also capable of an array of facial motions similar to Kismet, and has been additionally evaluated using mechanisms to mimic

facial expressions of human observers (Mayer et al., 2010). BERT2 is a hybrid humanoid face mixing embodied and digital aspects (Bazo et al., 2010). Feelix, developed by Canamero and Fredslund (2001), was a robotic face designed from Lego Mindstorms™. Both BERT2 and Feelix implemented similar, though less complex, artificial emotion mechanisms like Kismet.

Figure 3.2: Examples of Robotic Faces



See text (Section 1.2) for appropriate citations.

In discussing the results of our experimentation with MiRAE, we focus on the four robotic faces mentioned above because: 1) they are primarily humanoid, and 2) they underwent some degree of rigorous experimental evaluation similar to that described for MiRAE here (see Section 3.2.2). However, many other robotic faces have been designed during the same time frame. These include the elephant-like Probo (Saldien et al., 2010), Kaspar (Blow et al., 2006), the retro-projected faces of Delaunay et al.

(2009), Sparky (Scheef et al., 2002), the androids Actroid-F (Yoshikawa et al. 2011) and Geminoid-F (Becker-Asano & Ishiguro, 2011), iCat (Van Breemen, Yan, & Meerbeek, 2005), ROMAN (Berns & Hirth, 2006), the teddy-bear-like EmotiRob (Saint-Aimé, Le Pévédic, & Duhaut, 2009), the Sheffield robot (Zhang & Sharkey, 2011), Mexi (Esau et al., 2003), and Lucas/Nexi (http://robotic.media.mit.edu/projects/robots/mds/overview/overview.html). This list highlights the range of robotic faces being developed and researched in recent years; it is, however, by no means exhaustive.

Additionally to understand robotic face research in the context of human-robot interaction, it is important to be cognizant of the distinction between the *capabilities* of a given robotic face to make certain facial expressions (e.g. the six basic Ekman emotions) and *applications* of such capabilities to actual interaction. In our view, facial expression capabilities (#1) and their use in human-robot interaction (#2) represent two distinct, though closely related, research questions. In the first question, we are interested in understanding the principles required for robotic faces to create facial expressions that people can perceive/understand, including identifying the minimal features and understanding the effects of facial components, design aesthetics, degree of motion, etc. For the second question, we are interested in the application of facial expression capabilities to simulate/study specific social interaction scenarios and/or behavior. The focus of this latter research agenda is more broadly on the interaction itself, in which facial expressions represent only one component subsumed in the broader system. The second question also often comprises the use of computational models of artificial emotion (e.g. Kismet). Moreover, not all the aforementioned robotic face studies address the second question. We detail robot/agent emotions below (Section 3.1.3). In this chapter, we focus on the first question.

3.1.3 Robot/Agent Emotions

Emotions, as well as non-verbal communication of such emotions, serve a critical role in biological organisms (Gadanho & Hallam, 2001; Ekman, 2009). Emotions can form part of the basis for:

1) **Attentional Control** – what features are important to pay attention to in the environment (should I look at that leopard or the rock?)

2) **Reflexive Behavioral Tendencies** – reflex behaviors in emergency situations (it's a leopard, don't think, run away!)
3) **Social Interaction/Communication** – critical adaptive behavior in social species (there's a leopard behind you, hence the fear in my face)

Emotions may also play a role in decision-making, memory, somato-sensory responses, and other cognitive processes (Gadanho & Hallam, 2001; Bechara, Damasio, & Damasio, 2000; Dolan, 2002; Breazeal, 2009). There is strong evidence for the adaptive role that emotions may have played in the course of evolution, both in humans and other animals (Gadanho & Hallam, 2001; Ekman, 2009).

Artificial emotions (and/or more broadly affective computing) refer to the ability of technology (computers, robots, artificial agents, etc.) to both recognize and express emotions, typically through the use of computational models (Robinson & El Kaliouby. 2009). Artificial emotions have been proposed as a potential "cognitive control architecture" in multi-agent systems (Canamero, 1997; Gadanho & Hallam, 2001). For instance, Gadanho and Hallam (2001) used artificial emotions as a "filter" between perceptions and actions in order to synthesize appropriate behaviors from noisy perceptual information. Numerous computational models for artificial emotion have also been implemented in robotic face platforms, most notably in Kismet (Breazeal, 2001), but also in others like Probo (Saldien et al., 2010) and the Roboceptionist (Kirby, Forlizzi, & Simmons, 2010). Implementations vary conceptually across robotic platforms, but generally utilize some mathematical formulation to convert perceptions into emotions, facial expressions, and/or behaviors. For instance, the architecture deployed in Kismet utilized four "cognitive" stages – perception, cognitive appraisal, emotional activation, and behavioral (i.e. facial expression/posture) activation – which capture Russell's affect space (see Section 3.2.1.1) as three mathematical values (arousal, valence, and stance) that can be calculated and communicated across the cognitive process (Breazeal, 2003). The end result is a set of numerical values that trigger an emotional response (and related behavior/facial expression) when appropriate (i.e. when some threshold is exceeded).

These computational models, both in robots and multi-agent systems, also enable the use of emotions to address the problem of action selection/switching, which is the challenge in an agent or organism of determining when to continue a current behavior or switch to a new one (Bryson & Tanguy, 2010). Computational models allow the conceptualization of artificial emotions as *trajectories* that bias behavioral tendencies, with thresholds representing the equivalent of attractor basins from a dynamical systems theory perspective. Various time scales of operation for these biases can also be conceived of as constructs reflecting the differences between short-term emotions and long-term moods/drives (Gadanho & Hallam, 2001; Ekman & Friesen, 2003; Bryson & Tanguy, 2010).

In short, emotions play a critical role in biological organisms, and artificial emotions hold promise to play a similarly critical role in artificial entities. They can potentially create naturalistic social behavior between robots and humans through relatively simplistic mechanisms (along with addressing other problems, e.g. cognitive control, action switching, etc.). However, in order to achieve such social interaction, it is necessary to address the research problem of robotic face design and facial expressions in a systematic way, so as to understand the fundamental features needed to convey information (including artificial emotions) to humans in a way they can perceive and understand. This could be accelerated via inexpensive and replicable robotic platforms and the application of rigorous experimental evaluation (see Section 3.1.1).

3.1.4 Potential Applications of Minimalist Robotic Facial Expressions

As noted in Section 3.1.1, this work has implications for the development of socially interactive robots – such as those used for therapeutic or assistive purposes – that need not only detect human facial expressions but also express them. The role of *social intelligence* (including things like artificial emotion and facial expressions) has elsewhere been argued to be a critical component for development of socially interactive robots and artificial intelligence in general (Dautenhaun, 2007). Although many of the aforementioned robots (Section 3.1.2) have been primarily focused on understanding facial-expression-

based human-robot interaction in lab settings, the overarching goal is to apply the findings of such research to robots interacting with people in real-world settings and/or for practical purposes.

Examples of applications of robots capable of emotional display and used for practical purposes include the Rubi education robot (Movellan et al., 2005), service robots (Kwon et al., 2007), the nurse-bot PEARL (Pollack et al., 2002), museum tour-guide robots (Faber et al., 2009), patient care robots (Allison, Nejat, & Kao, 2009), and socially assistive robotics (SAR) for autism therapy (Scasselati, Admoni, & Mataric, 2012). Emotional expression in these types of robots can help users understand the robot's intentions and state, such as if it is over-stimulated or interested in an object (Breazeal, 2003). Emotional cues can also be used to manage the behavior of human interaction partners to fit the robot's needs. For example a museum guide robot's angry expression can cue people blocking its path to move out its way so that it can continue guiding them (Thrun et al., 1999).

However, many of these examples currently utilize only limited facial cues and/or expressions, and the design of the emotional expression capabilities is not based on rigorous empirical testing of the underlying principles. The approach utilized here can contribute to the future design of socially interactive robots by providing a minimum set of necessary components that such robots must include and on which they can build further. For example, in the case of robots used for autism treatment, the minimal set of cues could be used as a baseline from which individuals can be taught to interpret more complex emotional expressions (Ogino, Watanabe, & Asada, 2008). Furthermore, a simple minimalist robotic platform can serve as the basis for studying human cognition, including social cognition, as has been argued previously – this, however, necessitates certain capabilities (see Section 3.4.2)

3.2 Methods

3.2.1 Robot Face Design

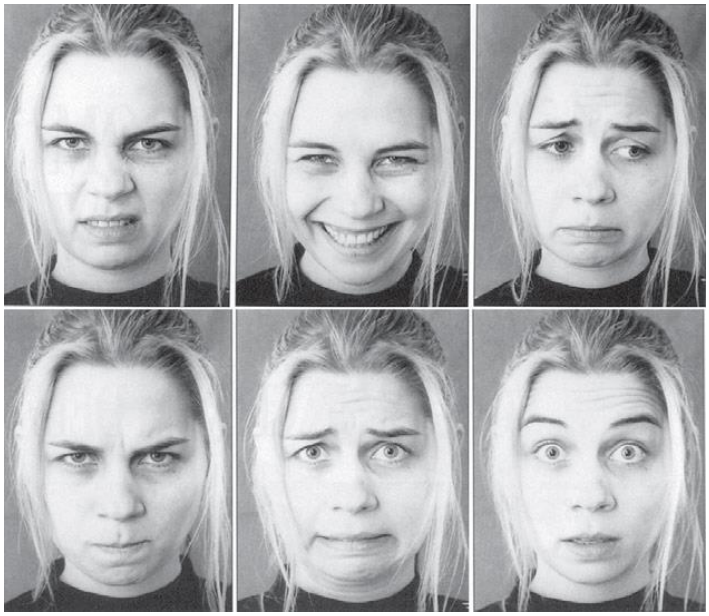
3.2.1.1 Design Overview

The robotic face design utilized here was inspired by recent research in computer vision on human facial expressions and based on Ekman's theories of emotion and the Facial Action Coding

System (FACS) (Ekman & Friesen, 2003). According to this theory, there is a set of six basic emotions - Happy, Sad, Angry, Fear/Worry, Surprise, and Disgust – which are rooted in evolution and displayed using similar features across human cultures (Figure 3.3). These features are referred to as *Action Units* (AUs, 44 in total), which capture all possible movement of the muscles of the human face. Activation values for these AUs can be calculated and used to accurately identify the emotion expressed in a given human face, regardless of the idiosyncrasies of the individual face (Ekman & Friesen, 2003; Pantic & Bartlett, 2007). Given that the ability of people to recognize these facial expressions appears to be instinctive, it can be reasoned that humans may use these same AU features to recognize emotions in facial expressions of other people (Calder & Young, 2005; Ekman, 2009). There is evidence, however, that the display and reading of facial expressions may be variable across cultures (Shore, 1996; Yuki, Maddux, & Masuda, 2007; Jack et al., 2009), posing further questions for investigation (see Section 3.4.3).

Of note, there is debate as to how to conceptualize these emotions, primarily between Ekman's categorical view (Ekman & Friesen, 2003) and Russell's three-dimensional *affect space* view (a.k.a. the circumplex model) (Russell & Fernández-Dols, 1997). In short, Russell's model utilizes three continuous-valued dimensions (arousal, valence, and stance) and treats all "emotions" as manifestations in the resulting 3D *affect space*. In other words, what we classify as emotions (and/or emotional facial expressions) are in actuality ill-defined points in continuous space, rather than distinct categories (Fugate, 2013). This view contrasts with the basic Ekman categorical emotions (described above). However, this debate has been repeatedly detailed in the literature (Cohn, 2010; Breazeal, 2003, Sosnowski et al., 2006; Bazo et al., 2010; Saldien et al., 2010), and primarily centers on what emotions constitute (not whether they exist). As such, we will not address it here.

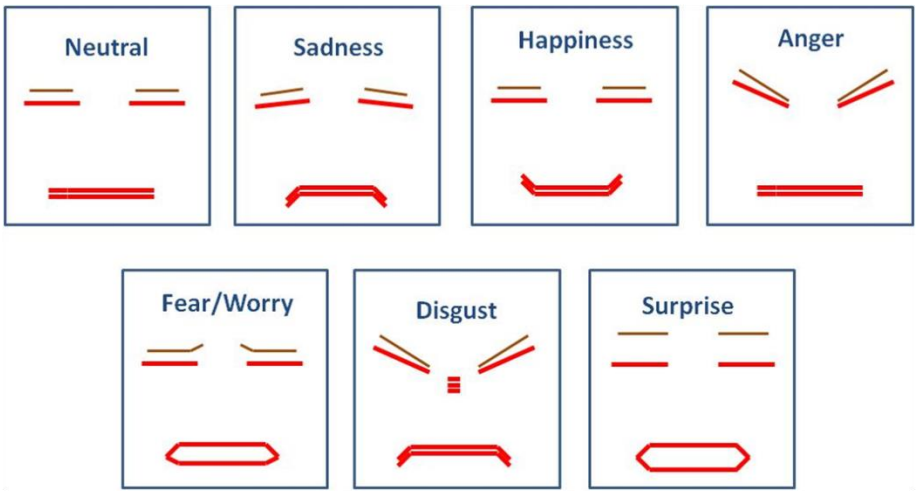
Figure 3.3: Human facial expressions of six basic Ekman emotions



In order (left-to-right, top-to-bottom) – Disgust, Happiness, Sadness, Anger, Fear and Surprise (Cohn 2010).

Feature selection techniques from machine learning have revealed that a small number of moving points/lines (i.e. a subset of 8-10 AUs) can be used to capture the vast majority of information in human facial expressions (~95%). This insight had been leveraged over the past several years in the computer vision community to develop automated techniques for classifying human facial expressions via computers (Pantic, 2009; Cohn, 2010; Anderson & McOwen, 2006). These were translated into the schematic representation shown in Figure 3.4, comprising two principle linear feature sets: upper (eye/brow) and lower (mouth). Similar sparse cues have also been indicated to play a role in human perception of emotion (Aronoff, Woike, & Hyman, 1992).

Figure 3.4: Schematic Facial Expressions



These simple schematic representations were used as the basis for the embodied robotic face design (Section 3.2.1.2) as well as the digital avatar version (Section 3.2.1.4) below. The goal of the experiments was to start with a simple, minimalist robotic face with as little complexity as possible, perform thorough scientific experimentation of its facial expression capabilities with human users, and then build from that. For instance, if one degree-of-freedom (DOF) for eye motion turned out to be insufficient for a given task, then additional DOFs could be added. If simple lines proved insufficient, then more robust shapes could be evaluated. In short, we wanted to minimize our *a priori* assumptions about what was and was not important. We also wanted the embodied robotic face to be as similar to the digital avatar version for experimental purposes (Experiment #1, see Section 3.2.2).

Although many other researchers have utilized similar approaches in the design of robotic faces (Breazeal, 2003; Sosnowski et al., 2006; Saint-Aime et al., 2007), our approach differs in its strict adherence to the minimal features (AUs) without addition of extraneous (and/or potentially confounding) attributes, e.g. ears or other aesthetic properties, as well as in the iterative process of designing and evaluating these features.

3.2.1.2 Embodied Face Design

The embodied face (MiRAE) was constructed using inexpensive, easily accessible components in keeping with a minimalist design approach (see Figure 3.1 above). Principally, physical construction was accomplished using universal metal joints, plastic arms, and universal plates from Tamiya (http://www.tamiyausa.com); 10 sub-micro Hitec servos (http://www.hitecrcd.com); and various servo brackets. Arduino Uno v1.1 microcontrollers (http://www.arduino.cc) were used to create and control functionality of the robotic face. Combinations of servo motors were used to create the needed motion and degrees-of-freedom (DOF): 1 DOF for each eye, 2 DOF for each eyebrow, 2 DOF for the mouth corners/lips, and 2 DOF for the neck. Combined actuation of these simple DOFs could simulate complex motion, such as the parting of lips and bearing of teeth (see Figure 3.6 below). Facial features such as eyes, eyebrows, and the mouth were simulated using colored pipe cleaners affixed using gauge wire. More explicit instructions are available the author's website (http://www.caseybennett.com/Research.html) and at the author's lab website (http://r-house.soic.indiana.edu/mirae/MiRAE_Construction_Manual.pdf). All the programming code, including the C++ libraries (see below), is also available on those websites. In addition, schematics for completely 3D printing the robot-face and head (as well as modifying it for other purposes) will be available from those websites.

In total (including the neck mechanism described below, Section 3.2.1.3), the overall cost for the robotic face is approximately $150-175 USD. Total construction time averages roughly 6 hours. The potential exists to enhance the current bare-bones design – such as through the use of 3D printing to create more realistic facial features like eyes. However, the goal here was to minimize aesthetic properties so as to focus on the effect of the features themselves, as well as provide a direct comparison to the digital avatar (see Section 3.2.1.4).

The programming code to control the robotic face was written in the Arduino language, which is based on C++, as a C++ library (available online, see above). Some extensions to the basic Arduino language were written as structures to handle multi-variable function returns. The code was designed as a

three-tiered system. The main program could call functions that specified facial expressions, passing in the *direction* (used to make or undo an expression) and *degree* (continuous value used to determine the strength or degree of the expression – i.e. smaller vs. larger). The facial expression functions would in turn call lower functions that moved specific facial components given a direction and degree – in essence these facial component functions roughly relate to specific AUs in the Facial Action Coding System (FACS). This approach provides several benefits – e.g. it allows for easy extensibility to include new expressions, AUs, or types of facial motion. It also permits a direct linkage between the programming code used to control the robot face and underlying theory about human emotion and facial expression. Also of note, motion (in both the embodied and digital avatar versions) was implemented as *gradated motion*, so that facial expressions occurred over a matter of a few hundred milliseconds (as they would in a real human face), rather than instantaneously. The platform also allows for nuanced control of the expressions and their level of intensity (i.e. degree) in experimental situations.

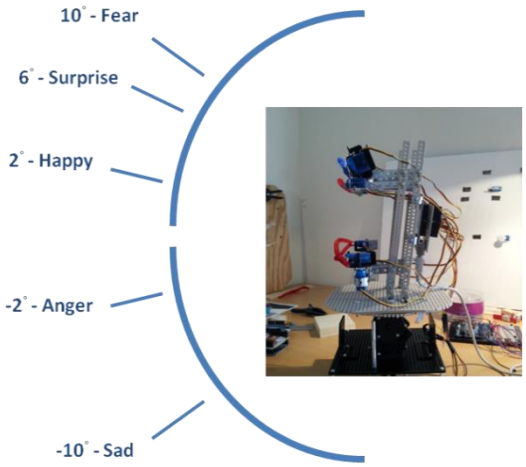
A comparison of the DOF of MiRAE and several other robotic faces designed for human-like facial expressions is provide in Table 3.3 in the Results section (Section 3.3.1).

3.2.1.3 Added Neck Motion/Neck Posture

An additional question of interest was the effect of added neck motion (or neck posture) on human facial expression identification (FEI) in robotic faces. Previous work has studied neck motion/posture in both embodied and digital robots as it relates to general human-robot interaction and communication (Breazeal, 2003; Gratch et al. 2002, Sidner et al., 2006), but not specifically related to facial expression identification. In order to address this, a neck mechanism was constructed using a ServoCity SPT200 Heavy-Duty Pan & Tilt System and two Hitec high-torque HS-485HB servos. The "face" of MiRAE was mounted on top of this mechanism, allowing for both vertical (tilt, i.e. up/down) and horizontal (pan, i.e. left/right) rotational motion, similar to a human neck. More explicit details are provided online (http://r-house.soic.indiana.edu/mirae/MiRAE_Construction_Manual.pdf).

The final aspect was determination of how much motion should be applied for the each of the Ekman emotions. Surprisingly, literature on neck motion/posture as it relates to facial expressions is limited, even in humans (De Gelder, 2009). As such, we ran preliminary trials (using lab personnel only) *before* the actual experiments described below to arrive at reasonable estimates. We found that a relatively small amount of vertical (i.e. rotational tilt) neck motion created a rather large effect. The utilized values for each emotion (with negative values indicating down and positive indicating up, see Figure 3.5) were as follows: Happy (2°), Sad (-10°), Anger (-2°), Fear (10°), and Surprise (6°). Disgust neck motion is still yet to-be-determined (see Section 3.4.2). A video of MiRAE making these facial expressions, plus the neck motion, is available online (http://r-house.soic.indiana.edu/mirae/MiRAE_neck_video.mpg).

Figure 3.5: Added Neck Motion



Note: Angle representations in figure are not to scale.

3.2.1.4 Digital Avatar

A digital avatar version was implemented for the first experiment (Section 3.2.2) so as to compare the minimal features in an embodied vs. digital form. The digital avatar version was designed to

look virtually identical to the schematic representations (Figure 3.4), and thus by extension as similar to the embodied version as possible. It was implemented using Python 2.7 (www.python.org) and the TkInter toolkit package (http://wiki.python.org/moin/TkInter). Programming was implemented using the same approach as for the embodied face (e.g. three-tiered design, gradated motion) as described above (Section 3.2.1.2).

3.2.2 Experimental Design

We conducted a series of four experiments to evaluate human abilities to perceive and understand robotic non-verbal affective cues while varying factors related to robot and study design. The four experiments included evaluations of:

1) Embodied robotic face vs. digital avatar version
2) Effect of additional neck motion (vs. no neck motion)
3) Effect of "priming" subjects using human facial expressions (vs. no priming)
4) Effect of the degree of expressions (smaller vs. larger)

We recruited 75 unique subjects across all experiments (total n=75), 30 for the first experiment and 15 apiece for the other three. Subjects were randomly assigned to experiments, and each subject participated in only one experiment. Importantly, we were concerned about potential effects of repeatedly showing human subjects another entity making facial expressions, the so-called *priming* effect from psychology (we test this in Experiment #3). All subjects were college undergraduates in the United States (i.e. generally 18-23 years old) from various disciplines (e.g. computer science, psychology) and of varying gender (approximately 54.7% female). All the experiments were performed during the same 3 month time period (October 2012 thru January 2013).

In all experiments, subjects observed the robotic face (and/or digital avatar, if applicable) making a randomized pre-set series of facial expressions (the six Ekman emotions, less Disgust; see below) and responded to a three-item Facial Expression Identification (FEI) instrument for each expression. On the

FEI, subjects were asked to first identify the expression (Question #1) and to rate the strength of expression (Question #2). The FEI used a similar 7-option forced-choice design for Question #1 as was used in studies with Kismet, Eddie, etc. for comparability purposes (FEI available online in English and Japanese: http://r-house.soic.indiana.edu/mirae/FEI_Instrument.docx) (Breazeal, 2003; Sosnowksi et al., 2006). The FEI also asked subjects an additional question (Question #3) for each expression, allowing (but not requiring) them to select one or more "other expressions" they thought the robot might be displaying, if desired. For instance, if the subject selected Surprise as the most likely emotion for a given expression on Question #1, but also thought the expression might have been Fear, they could circle Fear on Question #3. Question #3 was included on the FEI since there has been some criticism of the forced-choice emotion labeling task, going all the way back to Ekman even (Russell, 1994). In subsequent sections, we refer to *main accuracy* based on the single answer from Question #1, and *other accuracy* when including answers from both Question #1 and #3. *Strength ratings* are based on Question #2. Finally, subjects were administered the Negative Attitudes toward Robots Scale (NARS, prior to each experiment) (Nomura & Kanda, 2003) and Godspeed instruments (after each experiment) (Bartneck et al., 2009).

During each experiment, the robotic face (and/or digital avatar, if applicable) made each expression for several seconds, then returned to a neutral face. A pause of 15 seconds was provided between expressions to allow participant to fill out the FEI. Participants simply watched the robot, i.e. there was no interaction. The robot (nor avatar) did not speak or make affective sounds. The robot is capable of actual "interaction", e.g. it can see that people are there and react to them, track them, etc. However, this was not done in any of the experiments reported here.

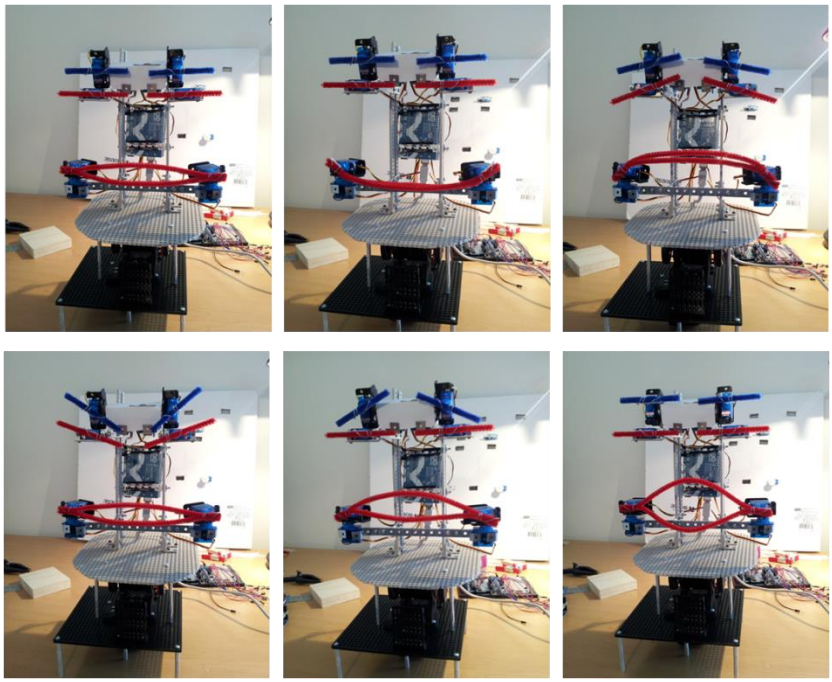
For Experiment #1, 30 subjects were randomized into two groups: one group seeing the embodied robot version first and the other group seeing the digital avatar version first. Both groups saw both versions (embodied and digital), only the order differed. This was done to rule out any potential effects due to the ordering. For subsequent experiments, the results of this first experiment were considered the "baseline" (i.e. the control).

For Experiment #2, 15 subjects observed the embodied robotic face making facial expressions with the addition of the added neck motion described in Section 3.2.1.3.

For Experiment #3, 15 subjects observed the embodied robotic face making facial expressions *after* being shown a priming stimulus of a human making the six Ekman emotions with each expression labeled (Figure 3.3 above). Subjects observed a printed copy of the priming stimulus for approximately 30-45 seconds total (i.e. about 5-7 seconds per expression), guided by the experimenter. Nothing was pointed out to subjects regarding details about the facial expressions or specific facial cues.

For Experiment #4, 15 subjects observed the embodied robotic face making facial expressions with roughly half the degree (50%, in numerical terms) of that seen in the first experiment– i.e. subjects in this experiment observed the "smaller" degree (50%), while subjects in the first experiment observed the "larger" degree (100%). Explicit numerical values for the degrees of motion of individual components are provided in the detailed instructions on the authors' website (see Section 3.2.1.2).

Figure 3.6: MiRAE Display of Emotions



Expression at apex of motion, without neck motion. In order (left-to-right, top-to-bottom) – Neutral, Happiness, Sadness, Anger, Fear and Surprise.

Of note, even though MiRAE has the ability to express Disgust using a similar approach as Kismet or Eddie (i.e. "Lip Twist"), we did not address Disgust in this current research, for reasons discussed in Section 3.4.2. Figure 3.6 shows MiRAE's basic display at the apex of the five expressions (without added neck motion) used in all experiments, plus the neutral starting expression.

3.2.3 Analysis Plan

The analysis plan was conceived *a priori*, and served as the basis for the experimental design detailed above (Section 3.2.2). Due to time/costs constraints (i.e. limited sample size), we chose to specifically test only four specific hypotheses, rather than all possible hypotheses. Each experiment above was designed to test one of these hypotheses. The study had two main parts: 1) A paired within-subjects design for Experiment 1 only, comparing the accuracy/strength-ratings of the embodied robotic and digital faces, and 2) a between-subjects design re-using the results from Experiment 1 (for the embodied robot) as the baseline (i.e. the control) and comparing each Experiment #2-4 to it to test for effects (listed in Section 3.2.2).

We thus do not test for all group-comparisons directly, e.g. added neck motion (Experiment #2) vs. reduced degree of expression. Rather, we compare each treatment condition in Experiments 2-4 with the "baseline" (i.e. Experiment #1). Again, this was a consequence of time/costs constraints - the approach was decided upon *a priori*, in order to allow us to take the best advantage of the sample size and power we had available.

Each of the four hypotheses was thus a two-group comparison: either embodied vs. digital (for Experiment #1), or control vs. effect (for Experiments #2-4)). For Experiment 1, this entailed paired *t*-test comparison, and for all other experiments it entailed independent-samples *t*-tests. We also tested for ordering effects (the effect of showing either the embodied or digital face first) in Experiment 1 using an

independent-samples *t*-test. Significance was measured at the $p<.05$ level (two-tailed). Effect sizes are reported using Cohen's *d*.

Given sufficient time and money, a larger study with a full fixed-effects ANOVA testing more hypotheses/group-comparisons would be of great interest. However, such an experimental design would require testing at least twice as many hypotheses and a much larger sample size to achieve sufficient statistical power – even our current reduced-hypothesis analysis has only modest statistical power (approximately 0.5, as calculated post-hoc). In practice, this sort of approach is still uncharted territory in many ways as far as robotic face experiments go; as such, this study represents a good preliminary step in that direction.

3.3 Results

3.3.1 Embodied vs. Digital Results

Table 3.1 shows the identification results between the embodied robotic face (MiRAE) and the digital avatar version (Experiment #1), including the accuracy of the main identified emotion and the accuracy when including the "other" identified emotions (see Section 3.2.2). In general, the results are comparable, though the digital avatar version was slightly higher for most expressions (avg. main accuracy, digital vs. embodied: 88% vs. 84%). The difference was not significant for main accuracy (paired *t*-test: $t(29)=1.44$, $p=.161$), though it was when including other accuracy ($t(29)=2.11$, $p=.043$, effect size=.49). This difference made some intuitive sense, in that it was easier to maintain better fidelity to the FACS in the digital version. However, the perceived strength ratings were on average slightly lower for the digital avatar (but not significant). Also of note, the perceived strength of expression significantly correlated with the identification accuracy ($r^2=.896$ for the embodied version). A confusion matrix of the results is provided in supplementary Table s1.

Table 3.1: Main Results of Expression Recognition

	Expression	Main Accuracy	Other Accuracy	Strength Rating
Embodied	Happy	96.7%	96.7%	7.31
	Sad	100.0%	100.0%	8.30
	Anger	86.7%	93.3%	7.25
	Fear	43.3%	63.3%	6.25
	Surprise	96.7%	100.0%	7.96
Digital	Happy	100.0%	100.0%	6.93
	Sad	100.0%	100.0%	8.09
	Anger	100.0%	100.0%	7.98
	Fear	53.3%	66.7%	6.38
	Surprise	86.7%	100.0%	7.22

Facial expression identification results from Experiment #1 for the Embodied Robot (top) and Digital Avatar (bottom) are shown. For each expression, main accuracy, other accuracy, and average strength ratings from the FEI instrument are provided (definitions for each are provided in Section 3.2.2).

Table 3.2 shows the comparison between MiRAE and several other recent robotic faces: Kismet (Breazeal, 2003), Eddie (Sosnowski et al. 2006), Feelix (Canamero & Fredslund, 2001), BERT (Bazo et al., 2010), and the android Geminoid-F (Becker-Asano & Ishiguro, 2011, values from Table 5 therein, Americans only). Generally, MiRAE produced higher, or at least comparable, identification accuracy rates for all expressions, despite its minimalist design and ease/brevity of construction. Across all faces, similar patterns can be observed (e.g. relative dip in Fear identification). Note that many robotic faces from the last decade are not included because similar rigorous experimental evaluation was never performed/reported.

Table 3.2: Robot Face Comparison

Expression	MiRAE (n=30)	Eddie (n=24)	Kismet (n=17)	Feelix (n=86)	BERT (n=10)	Geminoid (n=71)
Happy	97%	58%	82%	60%	99%	88%
Sad	100%	58%	82%	70%	100%	80%
Anger	87%	54%	76%	40%	64%	58%
Fear	43%	42%	47%	16%	44%	9%
Surprise	97%	75%	82%	37%	93%	55%
Disgust	-	58%	71%	-	18%	-
Average[a]	85%	57%	74%	45%	80%	58%

Facial expression identification average accuracy for the six Ekman emotions is shown for several robotic faces (including the own used here, MiRAE). The number of subjects (*n*) is shown for each study as well. Appropriate citations for each are provided in text. [a]Averages do not include Disgust, since not all studies included it.

Table 3.3 provides the degrees-of-freedom (DOF) for different facial features across the same robotic faces in Table 3.2. In particular, the results suggests that moveable facial features such as eyelids, ears, and animal-like crowns, which were used in some of the previous studies, appear to be dispensable for creating recognizable robotic facial expressions for affect (the Ekman emotions). Comparable recognition rates were obtained in the current study using only movement of the eyes, eyebrows, and mouth. It should also be noted that several of the robots (e.g. Feelix, Geminoid) took advantage of facial symmetry to reduce the DOF listed in Table 3.3, e.g. using only a single motor to rotate both eyebrows rather than two separate motors, while several (Kismet, Eddie, MiRAE) did not. In the current study, only symmetrical expressions were used, which indicates the DOF of those latter robots could be potentially reduced (by roughly half), at least for the Ekman emotions. Further research would be needed to empirically establish the value, if any, of asymmetrical expression capabilities.

Table 3.3: Robot Face DOF Comparison

Feature	MiRAE	EDDIE	Kismet	Feelix	Bert	Geminoid
Eyes	2	3	3	-	2	2
Eyelids	-	4	2	-	1	1
Pupil Size	-	-	-	-	2	-
Brows	4	4	4	1	4	2
Ears	-	6	4	-	-	-
Crown	-	1	-	-	-	-
Mouth	4	5	5	3	4	2
Neck	2	-	3	-	-	4
Total	12	23	21	4	13	11

Degrees-of-Freedom (DOF) are shown for each robotic face. This can be compared facial expression recognition accuracy in Table 3.2. Dashes indicate that there was no DOF for the given feature for a particular robotic face. Appropriate citations for each are provided in text

Godspeed and NARS values were also analyzed for differences between the embodied robotic face and the digital avatar version. Table 3.4 shows the values across the five domains of the Godspeed scale. The numbers were comparable across all domains, with only animacy and likeability showing statistically higher values for the embodied version (at the $p<.05$ level, paired t-test: $t(29)=2.18$, $p=.037$ and $t(29)=2.36$, $p=.025$, respectively). Analysis of the NARS revealed no significant relationships with either identification accuracy or Godspeed ratings (data not shown).

Table 3.4: Embodied vs. Digital – Godspeed Ratings

Category	Embodied	Digital	Sig.	Effect Size
Anthropomorphism	2.26	2.08	0.228	
Animacy	2.44	2.14	0.037*	0.40
Likeability	3.58	3.26	0.025*	0.52
Perceived Intelligence	2.86	2.85	0.940	
Perceived Safety	3.83	3.91	0.500	

Ratings for each of the five Godspeed-scale domains are provided for the Embodied Robot and the Digital Avatar on the left. Significance values (p-values) testing for significant differences between the Embodied Robot and the Digital Avatar in each

domain – based on paired *t*-tests – are provided on the right. Significance values that are statistically significant are starred and effect sizes are provided.

Also of note, there was no significant difference due to ordering (whether subjects saw the digital or embodied version first) based on overall main identification accuracy (independent samples *t*-test: $t(28)=1.57$, $p=.127$), other accuracy ($t(28)=.866$, $p=.384$) or strength ratings ($t(28)=1.35$, $p=.181$).

3.3.2 Added Neck Motion Results

Table 3.5 shows the results from the added neck motion (or neck posture, Experiment #2). Identification accuracy values were pushed to 100% for all expressions, except for Fear (n=15). In short, identification accuracy was generally higher for the embodied robotic face with neck motion (Table 3.5) than without it (Table 3.1), avg. main accuracy: 87% vs. 84%. Effects were similar to Experiment 1, in that main accuracy differences were not significant, but differences were statistically significant when including other accuracy (independent samples *t*-test: $t(43)=2.50$, $p=.016$, effect size=.72). This also effectively eliminated any difference in identification accuracy between embodied and digital versions as seen in Experiment 1. Strength ratings were significantly increased for all expressions (independent samples *t*-test: $t(43)=2.11$, $p=.042$, effect size=.62), except surprise which remained roughly stable.

Table 3.5: Added Neck Motion

	Expression	Main Accuracy	Other Accuracy	Strength Rating
Embodied	Happy	100.0%	100.0%	9.07
	Sad	100.0%	100.0%	8.80
	Anger	100.0%	100.0%	8.00
	Fear	40.0%	86.7%	8.60
	Surprise	100.0%	100.0%	7.80

Even for Fear, the "other" accuracy did significantly increase, indicating that subjects generally recognized that the expression could be Fear, even if they were unsure. Interestingly, when Fear was misidentified in Experiment #1 (Section 3.3.1), it was most often misidentified as Sad (approximately 83% of misidentifications). In contrast, for Experiment #2, it was most often misidentified as Surprise (approximately 90% of misidentifications). This may suggest confusion due to the neck motion used for Fear and Surprise during Experiment #2, and indicate that alternative neck/body postures for Fear should be explored.

Godspeed and NARS ratings were also collected for the robotic face with added neck motion. However, the data were not significantly different from those reported for the embodied robotic face without neck motion (Table 3.4) and are omitted for brevity.

3.3.3 Primer Effect Results

A question of interest was the effect of "priming" subjects by showing them images of a human making the same facial expressions prior to observing the robotic face. We hypothesized this may increase the identification accuracy by alerting an individual's cognitive processes to prepare for specific facial cues (either consciously or unconsciously). Importantly, we were concerned about potential effects of repeatedly showing human subjects another entity making facial expressions, the so-called *priming* effect from psychology. The priming effect is notorious in many psychology experiments (Hermans, De Houwer, & Eelen, 1994; Henson et al., 2003; Pierno et al., 2008), but potentially an unacknowledged source of bias in robotic facial expression experiments.

Priming showed mixed results in effects on perceptions of robotic facial expressions. Comparing Table 3.6 to the embodied results in Table 3.1, average main accuracy increased to 89% vs. 84%. This followed a similar pattern to that seen with the digital avatar (Experiment #1) and/or added neck motion (Experiment #2). However, this was not statistically significant for either main or other accuracy. Strength ratings, on the other hand, were significantly increased (with the exception of Fear; independent samples *t*-test: $t(43)=2.10$, $p=.042$, effect size=.59). In short, priming mainly affected people's perception

of the intensity of the expression, without significantly altering their interpretation of which emotion was being communicated.

Table 3.6: Primer Results

	Expression	Main Accuracy	Other Accuracy	Strength Rating
Embodied	Happy	100.0%	100.0%	9.07
	Sad	100.0%	100.0%	8.80
	Anger	100.0%	100.0%	8.00
	Fear	40.0%	86.7%	8.60
	Surprise	100.0%	100.0%	7.80

In this case, the primer was shown immediately before the robotic face experiment (short-term). It is not clear if this effect is long-term as well. It does raise some possible concerns about repeatedly showing human subjects another entity making facial expressions (either human or robotic). In this study, *we took precautions so as to only use each subject once for a single experiment*, as noted in Section 3.2.2. However, in other studies where subjects may have possibly been re-used across experiments, reported results could potentially be erroneous due to such an effect. This warrants caution for experimental design in future studies of human-robot interaction and robotic facial expression.

Godspeed and NARS ratings were also collected for the robotic face with primer effects. However, the data were not significantly different from those reported for the embodied robotic face without the primer (Table 3.4) and are omitted for brevity.

3.3.4 Varying Degrees of Expression Results

Table 3.7 shows the results when the robotic face made expressions using one-half the degree of motion as in Experiment #1 (i.e. 50% less motion). In short, there were no statistically significant differences for main accuracy, other accuracy, strength ratings, Godspeed, or NARS (e.g. compare Table

3.7 vs. the embodied results in Table 3.1). Interestingly, human subjects were generally still able to identify the emotion being expressed with similar accuracy as with the larger degree of motion. Fear and Happy were the only expressions that exhibited any notable decline. In contrast to our hypothesis, there was *not* a consistent reduction in strength rating across emotions when a smaller degree of motion was used – some expressions actually increased slightly.

Table 3.7: Smaller Degree of Expression Results

	Expression	Main Accuracy	Other Accuracy	Strength Rating
Embodied	Happy	80.0%	80.0%	6.50
	Sad	93.3%	93.3%	7.42
	Anger	93.3%	93.3%	7.78
	Fear	20.0%	46.7%	6.50
	Surprise	100.0%	100.0%	8.20

An open question is where the lower bound lies for motion in robotic facial expressions that people could still perceive and understand. Our hypothesis was that there would be a continuously decreasing identification accuracy rate and strength rating as degree of motion was reduced. However, at least in a comparison of two specific degrees of motion in a minimalist robot, this phenomenon was not consistently observed across expressions.

3.4. Discussion

3.4.1 General Discussion

This study explored deriving minimal features for a robotic face to convey information (via facial expressions) that people can perceive/understand. The robotic face (MiRAE) was run through a series of experiments with human subjects (n=75) exploring the effect of various factors, including added neck motion and degree of expression. Facial expression identification rates were similar to more complex robots. Results suggest that movement of certain facial features (e.g. eyelids, ears, animal-like crowns) is

not requisite for creating recognizable facial expressions of affect (Ekman emotions) – movement of the eyes, eyebrows, and mouth alone is sufficient. In addition, added neck motion improved facial expression identification rates to 100% for all expressions (except Fear), as well as significantly increasing perceived strength of expression.

The embodied robotic face was also compared with a digital avatar version. Facial expression identification accuracy was higher for the digital avatar, which was attributed to the fact that it was easier to maintain fidelity with the FACS in the digital version. However, adding neck motion to the embodied robotic face eliminated this difference. On the other hand, perceived strength of expression ratings were slightly lower for the digital version, and Godspeed ratings revealed significantly higher perceptions of animacy and likeability for the embodied robotic face versus the digital avatar, which may make the former more appropriate for socially-interactive and assistive purposes.

Additional findings included that perceived strength of expression correlated strongly with identification accuracy rates. There was also an apparent effect on perceived strength due to "priming" subjects using human facial expression images as stimuli, which may suggest caution in experimental design of robotic face studies. Alternatively, it might also suggest that people could be quickly and explicitly trained to recognize robotic expressions with high accuracy, possibly as an alternative to including additional components, added motion, and/or more sophisticated features. Lastly, we found that human subjects were still able to identify robotic facial expressions even when half the degree of motion was used. An open question exists as to where the lower bound of motion lies for robotic expressions that people can perceive and understand.

3.4.2 Implications/Limitations

As mentioned in Section 3.1.1, this study has a number of implications for robotic face design. The results shown here hold promise to immensely reduce the complexity of constructing affective robots, allowing for greater flexibility in robotic design for social interaction. It may also free up constraints associated with mimicking non-critical aspects of human anatomy. More broadly, the

minimalist approach could be applied to many aspects of robotic form – e.g. previous applications to the study of affective, attentional and rhythmic cues in Keepon (Matsumoto, Fujii, & Okada, 2006) and relational interaction in Muu (Kozima, Michalowski, & Nakagawa, 2009) – as well as exploration of embodied cognition of artificial emotions.

The development of inexpensive robotic platforms utilizing widely available materials also holds promise to enhance replicable and methodologically-rigorous experimentation in human-robot interaction – and thus the advancement of "robotic science" – in contrast to robotics as purely an endeavor in engineering. Moreover, such an approach can enhance our understanding of human cognition. MacDorman and Ishiguro suggest that robots can act as unprecedented research tools for the study of social cognition by providing controllable stimuli in experimental and field studies; their behaviors can be carefully controlled, finely tuned and varied, and repeated exactly and indefinitely, which is challenging even for well-trained human confederates (Ishiguro, 2005; MacDorman & Ishiguro, 2006). Their proposed "android science," however, relies on very complex and expensive platforms that will be difficult to make widely available to the research community. Others have suggested that robots can be used to validate specific models of human embodied cognition, which can be implemented on robotic platforms and tested to see whether the expected human-like behavior is displayed (Scassellati, 2000; Barsalou, Breazeal, & Smith, 2007). Our suggested minimalist platform makes both of these approaches to the scientific study of human-human and human-robot interaction feasible. We focus particularly on answering cognitive science questions around people's abilities to make inferences using incomplete information during social interaction. Using robots to study human cognition, including social cognition, in lieu of human confederates necessitates robots capable of making signaling cues based on the way humans do. Having a simple minimalist platform containing such capabilities with which to study the various aspects of human affective perception and interaction – but without the cost and complexity – would be a boon to not only robotics, but also psychology and cognitive science.

There are many factors associated with facial expressions, social communication, and the recognition thereof that are not addressed in this study. Some of these are mentioned in Section 3.4.3. In

general, it should be noted that there is much work left to do to evaluate the importance of many aspects of such communication, both in human-to-human and human-robot interaction. This study only represents a small slice of the numerous questions that could be asked.

Additionally, the study did not consider the emotion of Disgust. MiRAE is capable of making this facial expression using the same approach as Kismet (Breazeal, 2003) and Eddie (Sosnowski et al., 2006). This approach entails the use of what we refer to as a "Lip Twist" expression, which involves the twisting of the lips so that one mouth corner is raise and the other lowered while simultaneously cocking the eyebrows in some fashion. However, this "Lip Twist" expression is not based on the FACS or pre-existing literature (Section 3.2.1.1) – it is in essence a substitute expression contrived to compensate for the difficulty in making nose-wrinkling motions. In the FACS, the Disgust expression is primarily indicated by a "Nose Wrinkle" expression, which involves wrinkling the upper bridge of the nose along with some movement of the eyes/brows and mouth. As such, in keeping with a strict adherence to the FACS and the existing body of literature, we chose not to address Disgust at this time. The goal of the study was to evaluate a robot making facial expressions using the same minimal features a human would – how one could do so for the nose wrinkling motion in Disgust is still an open question. To our knowledge, no robotic face has yet convincingly implemented such a capability. The closest examples are the android-type faces with skin-like coverings, e.g. the Actroid-F (Yoshikawa et al., 2011) or ROMAN (Berns & Hirth, 2006), but it is still unclear how effective these are in empirical terms (e.g. a study of the Geminoid-F did not include Disgust [Becker-Asano, 2011]) or how one might mimic such an effect without full-blown android faces. In short, further work is needed.

3.4.3 Future Work

Future work plans to extend upon the research described here. Another study (Bennett & Šabanović 2015, described in Chapter 4) involved a series of cross-cultural experiments between Japan and the U.S. to explore cultural variability in robotic facial cues during non-verbal communication. Evidence suggests that people from different cultural backgrounds may focus on different facial features

more than others (e.g. East Asians focus more on the eyes) (Shore, 1996; Yuki, Maddux, & Masuda, 2007; Jack et al., 2009; Trovato et al., 2013). There is also research that points to the variable significance of context as compared to the individual characteristics of the social actor (Nisbett, 2003). These experiments aim to answer such questions in the scope of human-robot interaction and understand the implications for robotic face design.

Additionally, a number of other factors remain to be evaluated. Simple questions like the shape and color of key facial components, like the eyes or lips (or whether their shape/color matter at all), are still open questions. The typical assumption is that the design of biological organisms has some explicit purpose, but in reality some aspects of the design of biological organisms, including humans, may simply be vestiges of the evolutionary process, i.e. less than optimal and/or unnecessary for the purposes of artificial agents. Systematic evaluation of additional aesthetic features – ears, hair, skin – can also potentially enhance our understanding. Advances in 3D printing present new possibilities for rapid prototyping and experimentation with such components. However, a challenge still exists as to what exactly these parts should look like, which parts you actually need, and how these parts should move in a communicative robotic face. Beyond the face itself, further investigation of the myriad of effects related to body posture and gesture is also warranted. Even more broadly, several other potential variables have thus far been the subject of only limited research – e.g. saliency of context (Zhang & Sharkey, 2011), interplay of human mental models with robot shape/form (Powers & Kiesler, 2006), and the effects of gaze tracking on human-robot interaction (Sidner & Lee, 2007). In summary, these various factors are representative of the numerous opportunities and need for additional future research. We address some of these factors (e.g. culture, context) in subsequent chapters.

Chapter 4

The Effects of Culture and Context on Perceptions of Robotic Facial Expressions

This chapter builds on the previous chapter's focus on empirically-grounded design of a minimalist robot face, exploring the effect that environmental context has on human perception of robotic facial expressions, and whether such effects vary across culture. It also explores whether inducing such context effects might enable culture-neutral models of robots and affective interaction.

Abstract. We report two experimental studies (Bennett & Šabanović, 2015) of human perceptions of robotic facial expressions while systematically varying context effects and the cultural background of subjects (n=93). Except for Fear, East Asian and Western subjects were not significantly different in recognition rates, and, while Westerners were better at judging affect from mouth movement alone, East Asians were not any better at judging affect based on eye/brow movement alone. Moreover, context effects appeared capable of over-riding such cultural differences, most notably for Fear. The results seem to run counter to previous theories of cultural differences in facial expression based on emoticons and eye fixation patterns. We connect this to broader research in cognitive science – suggesting the findings support a dynamical systems view of social cognition as an emergent phenomenon. The results here suggest that, if we can induce appropriate context effects, it may be possible to create culture-neutral models of robots and affective interaction.

4.1 Introduction

4.1.1 Background

While the previous chapter focused on the design of an empirically-grounded robot face, there are many factors *external* to the robot itself that may affect human-robot interaction and human perceptions thereof. Environmental context (e.g. lighting, music) can affect such perceptions in fundamental ways – the possibility exists that such contextual effects might vary across culture (Picard, 1997). This chapter explores such context effects and their variability across cultures.

Scientific inquiry stretching back over a century has contributed to an ongoing debate about the nature and classification of human emotions and their related facial expressions (e.g. Ekman, 2009; Nelson & Russell, 2013; Breazeal, 2003; Sosnowski et al., 2006; Pantic, 2009; Cohn, 2010). Even Charles Darwin played a role (Darwin, 1872). The main points of contention can be summarized as such: Does a basic set of universal human emotions (and their related facial expressions) exist across culture, gender, context, etc.? Moreover, are there universal facial cues associated with these expressions that we can distill out from the broader array of complex and/or idiosyncratic facial movements?

Research by Ekman and colleagues going back to the 1960's suggested that there was indeed such a basic set of universal human emotions and/or facial expressions (Ekman & Friesen, 2003; Ekman, 2009). This eventually led to the development of the Facial Action Coding System (FACS), which could be used to identify facial expressions via specific facial cues. These facial cues are referred to as *action units*, and intended to encode the movement of specific facial muscles. However, that research on the universality of emotions/expressions was challenged on multiple grounds based on the work of Russell (Russell & Fernandez-Dolz, 1997), Matsumoto (Matsumoto, 1992), and others from the 1980's onwards, using studies done with human images and confederates. Recent work over the last few years using digital avatars has further challenged the universality of basic "Ekman emotions" on the basis of variations due to culture, context, and age (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). However, that work, based heavily on visual fixation patterns, has been disputed by more recent research (see Section 4.1.2). In spite of these scientific controversies, sophisticated automated

facial expression recognition technology has been developed over the last decade such that computers can, at least for posed Western expressions, achieve roughly 95% accuracy for identifying human facial expressions (Pantic, 2009; Cohn, 2010). Furthermore, most robotic faces with affective expression capabilities built over the last decade continue to be based on the basic Ekman emotions and their associated facial expressions (Breazeal, 2003; Sosnowski et al., 2006; Canamero & Fredslund, 2001; Bazo et al., 2010; Saldien et al., 2010; Becker-Asano & Ishiguro, 2011; Bennett & Šabanović, 2013; Bennett & Šabanović, 2014). In short, the literature is full of conflicting evidence on the subject, suggesting a need for novel lines of evidence.

This chapter is aimed at that need, contributing to the debate over human perception of affective facial expressions and to the application of such research in robot design through two experimental studies in which participants interacted with a previously validated minimalist face robot (MiRAE). MiRAE was designed with the aim of utilizing the minimal facial cues necessary to convey facial expressions in ways humans can perceive/understand (Bennett & Šabanović, 2014). The first study investigated the effect of cultural differences in perceptions of robotic facial expressions, using three human-subject groups: Japanese (living in Japan), native East Asians (living in the United States), and Westerners (i.e. Americans). A second study evaluated the effects of context on those perceptions. Both experiments seek to understand how situational factors (e.g. context, culture) affect people's perceptions of affective facial expressions. These were part of a broader series of seven experiments involving nearly two-hundred-twenty human subjects, interacting in-person with the robot (Bennett & Šabanović, 2014, Bennett et al., 2014), the other experiments being detailed in Chapters 3 and 5. A novel contribution of this work is the simultaneous manipulation of participant culture and context together that allows us to analyze the effects of and interactions between both of these two factors on people's perceptions of a robot's affective expression.

This research is developed through reference to cognitive science and psychological theories, which suggest that our perceptions and modes of interaction are contextually dependent and dynamically constructed and biased by cues in our environment – culture, context, interaction partners, etc. (see

Related Work and Discussion sections below). Emotions perceived in others' faces – including robots – may be an internal construct in the mind of the perceiver, based on a number of perceptual and cognitive processes.

4.1.2 Related Work

Even if facial expressions of emotions are variable in humans, it is not precisely clear as to how or why. While certain aspects of emotional and cognitive development may be universal, researchers have shown that the specific ways in which people engage in affective interaction can vary across culturally-situated norms and context scenarios. For instance, Nisbett et al. (2001) suggested that different "cognitive styles" in Western and East Asian cultures define aspects of the environment that are worthy of attention (e.g. characteristics of the environment or of the individual) and acceptable communication patterns (e.g. implicit vs. explicit). Such cognitive differences between Western and East Asian subjects may indicate that the two groups vary in regard to their attention to the context of interaction as indicative of its affective valence. Similarly, Shore (1996) argued that "social-orientational models" in particular "provide a degree of standardization in emotional response within a community," and designate appropriate roles/behaviors within interaction as well as culturally normative rules for displaying, perceiving, and experiencing affect (pp.62-63). Ekman, Friesen, and Izard themselves suggested a similar "Deception hypothesis" in the 1970's to explain culturally-based affective expression encoding rules (Ekman, 1971). More recently, Elfenbein (2013) has proposed a "Dialect hypothesis" for affective communication, which posits isomorphisms between affective expressions and linguistic distributions/development.

In addition to such research with humans, researchers in recent years have also used digital avatars as stimuli for testing people's perceptions of emotion. However, the evidence derived from these studies is subject to debate. For example, much of this recent work on cross-cultural differences is rooted in what we refer to as the "Emoticon hypothesis". In short, this posits that since emoticons are different for specific features (e.g. eyes, mouth) between Western and Eastern/Asian styles, displays of emotions

by humans between those groups must therefore be different across those features as well (for example, East Asians focus more on the eyes, and Westerners more on the mouth) (Yuki et al., 2007). Several papers in the last few years have studied visual fixation patterns as the basis for these putative cultural differences in facial expressions, arguing that the patterns support the Emoticon hypothesis (Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). However, more recent papers have provided evidence countering the use of visual fixation patterns, noting that people are engaged in a range of information-gathering activities for a variety of purposes (not simply judging affect) when looking at other faces, including determining culture, gender, confidence, sexual attraction, social referencing, etc. (Arizpe et al., 2012; Blais et al., 2012; Peterson & Eckstein, 2012). Furthermore, interpretation of results from studies using digital avatars is complicated by their common use of cartoon-like facial representations that are sometimes difficult to clearly relate back to the FACS and/or facial displays directly based on them.

Other recent empirical work has provided evidence for significant, culturally-variable effects due to context on facial expression recognition, using both digital avatars and human faces (Righart & de Gelder, 2008; Barrett et al., 2011; Lee et al. 2012). This literature has focused on the variable importance of context between cultures (particularly Asian and Western cultures) as an explanation for such cross-cultural differences, related to the arguments of Shore (1996) and Nisbett et al. (2001) above. Researchers have also challenged the sole focus on facial expressions, suggesting body posture/gesture plays a significant role as well (Kleinsmith et al., 2006; de Gelder, 2009). A further complicating matter is the possible effect of variations in language and cultural connotations of emotion-label words (Perlovsky, 2009; Ruttkay, 2009). Lindquist & Gendron (2013) even suggest a dynamical systems perspective of emotion perception and word-label grounding to explain such variation.

To summarize, the debate over the universality of human emotions and facial expressions, as well as their mechanisms of display/interpretation, is complex and rife with conflicting evidence. As noted in Section 4.1.1, our research contributes to the production of new modes of evidence, via human-robot interaction, for systematically approaching this debate.

4.1.3 The Role of Human-Robot Interaction

There are multiple motivations for utilizing robotic faces to study the question of human display and perception of emotional expressions, both academic and pragmatic. On the one hand, robotic faces provide a three-dimensional, embodied platform that can be used as a controllable/consistent/modifiable surrogate for human images or confederates when investigating questions of human cognition and perception (Adams et al., 2000; Scasselati, 2006; Kozima et al., 2009). On the other hand, if we endeavor to add faces and facial expressions to robots in order to enhance human-robot interaction and communication, then understanding how to do so effectively is of immense importance. This is doubly true if robots are also meant to interpret human facial movements. If indeed factors like culture and context matter to human perception and performance of affective facial expressions, then future human-robot interaction design requires an empirically-based understanding of how and why.

Despite the aforementioned work using human images (Matsumoto, 1992; Russell & Fernandez-Dolz, 1997; etc.) and digital avatars (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; etc.) to investigate human facial expressions, as well as numerous papers evaluating the ability of robotic faces to display the basic Ekman emotions, limited research has been performed evaluating the purported *universality* of the Ekman-based facial expressions and facial cues using robotic faces. Becker-Asano and Ishiguro (2011) evaluated the android Geminoid-F robot across three cultural groups (Americans, Europeans, and Asians), showing clear differences across them. However, the study utilized only still, posed images of the robot distributed over the Internet, and even the Western subjects struggled to identify many of the expressions (e.g. Anger, Surprise) with high accuracy. Elsewhere, Zhang and Sharkey (2011) have evaluated the effects of context on robotic facial expression identification by humans, and Embgen et al. (2012) have conducted robotic studies on emotional body language in lieu of facial expressions.

More broadly, a number of researchers have investigated cross-cultural differences in perceptions of robots, though not necessarily for the specific purpose of affective communication (Bartneck & Okada, 2001; Bartneck et al., 2007). While many such studies agree that cultural factors influence how people

perceive and behave toward robots, there is a surprising lack of agreement on the nature of these differences. A popular view among scholars is that Japanese (and possibly other Asian) subjects are more positive towards robots in general and identify them as more lifelike and animate (e.g. Geraci, 2006; Kaplan, 2004). Bartneck et al. (2007) suggest the opposite – that US participants have the most positive attitudes toward robots, particularly in terms of their willingness to interact with them on a daily basis. MacDorman et al. (2009) find more similarities than differences in how pleasant or threatening US and Japanese participants deem robots to be. Lee and Šabanović's (2014) survey study of perceptions of robots among participants in the US, South Korea, and Turkey show that, while differences among these populations exist, they are not directly correlated with broad cultural factors such as animistic or Christian beliefs, or with media portrayals of robots. These divergent results suggest that more situated contextual factors beyond broadly defined national cultures may be responsible for differential perceptions and attitudes toward interactive robotic technologies, particularly variables related to the social context of the interaction.

The studies reported here explored the effect of cultural background and environmental context on people's perceptions of affective expressions of a robotic face. We used the same robot in studies performed face-to-face with participants in the USA and in Japan, so that all subjects were able to directly interact with the robotic face, rather than only watch pre-captured images or videos of the robot in action, which is known to have drawbacks (Krumhuber et al., 2013). The studies also involved using different "cultural variants" of facial expressions to test previous research findings (see Section 4.3.2), as well as experiments simultaneously varying both culture and context.

4.2 Methods

4.2.1 General Overview/Subjects

Two experiments are reported in this chapter. They are part of a broader series of seven experiments investigating the minimal features needed for a robotic face to communicate facial affect in a way humans could perceive and understand. The other five experiments have been previously reported

(Bennett & Šabanović, 2014, Bennett et al., 2014), and are detailed in Chapters 3 and 5 of this work. In total, 216 human subjects participated in all the experiments, of which 93 participated in the two reported here.

Three groups of subjects were utilized: Japanese (living in Japan), native East Asians (living in the United States), and Westerners (i.e. Americans). We use the term "Westerners" here to be consistent with Jack et al. (2009) and others. The Japanese were college students recruited in Japan from a university in Yokohama. The East Asians were a mixture of Japanese, South Korean, and Chinese college students, who had lived in the United States on average for 10 months (and generally no longer than one year) and had passed an English proficiency entrance exam (TOEFL). The Westerners were all American-born college students, primarily Caucasian. The age range across all groups was approximately 18-23 (i.e. college-aged). The gender mix was roughly 50-50, with the percent male being 53.2% (Westerners), 56% (Japanese), and 46.9% (East Asians living in the U.S.). The breakdown by experiment was: 57.7% (Experiment #1A), 47.9% (Experiment #1B), and 50% (Experiment #2). Most participants came from either the computer science or psychology programs.

For the two experiments reported here, the first (#1a and #1b) examined the effects of culture, and the second (#2) examined the effects of context on the participants' perceptions of affective expressions performed by the robotic face. Experiment #1a used a sample of Japanese subjects only (n=15). Experiment #1b used subjects from all three groups (n=48, 16 per group). Experiment #2 used samples of East Asian and Western subjects only (n=30, 15 per group). Subjects were not re-used across experiments, due to potential priming effects from repeatedly showing them facial expressions described in Section 3.3.3 (Bennett & Šabanović, 2014). Sample sizes were determined from previously observed effect sizes described in Chapter 3 (Bennett & Šabanović, 2014) with consideration for time/costs constraints.

The experiments were performed in-person through face-to-face interaction between the robot and participants at universities in the United States and in Japan. All experiments were performed in a

conference room against a neutral off-white background wall. For experiments involving the digital avatar and context videos, these were shown using a laptop in the same room setup.

4.2.2 Robotic Face

The platform used here (MiRAE) is a minimalist robotic face that is capable of displaying a variety of facial expressions (Bennett & Šabanović, 2014). In a previous study (detailed in Chapter 3), MiRAE was shown capable of producing higher, or at least comparable, identification accuracy rates (with Westerners) for all expressions as a number of other robotic faces, including Kismet (Breazeal, 2003), Eddie (Sosnowski et al. 2006), Feelix (Canamero & Fredslund, 2001), BERT (Bazo et al., 2010), and the android Geminoid-F (Becker-Asano & Ishiguro, 2011, values from Table 5 therein), as shown in Table 3.2 in Chapter 3. This indicates that a minimalist robotic face such as MiRAE can provide a reliable, replicable, low-cost platform for investigating questions of affect and facial expression such as those addressed here.

The minimalist approach for the robotic face used here is grounded in over a half-century of psychological and computer science research on emotions and facial expressions (Bennett & Šabanović, 2014). The entire premise of that work (Ekman, 2009; Nelson & Russell, 2013; Pantic, 2009; Cohn, 2010) is that people are only attending to a small number of critical moving points/lines to detect emotion in faces. This is the basis for the FACS, which dominates the emotional facial expression literature and on which many robotic faces – including androids – are based (see Section 4.1.1). At least within the specific task context of emotional facial expression recognition, there is evidence that many realistic aspects of the face are not necessary, and may indeed even be conflating factors (e.g. by suggesting cultural affiliation, ingroup/outgroup effects). Our previous study (Bennett & Šabanović, 2014, see Chapter 3) validated that principle in this exact robotic face, providing empirical evidence that simple moving lines work just as well for emotional expressions as more complex facial features (e.g. Kismet, see above). Other robotic research, such as Okada's Muu and Kozima's Keepon (Matsumoto et al., 2006;

Kozima et al., 2009), further support such minimalism for affective interaction (not to mention Mori's work on the "Uncanny Valley" [Mori, 1970]).

Examples of MiRAE displaying various facial expressions can be seen in Figure 3.6 in Chapter 3. The dimensions of the robotic face are similar to an actual human face, approximately 8 inches tall by 6.5 inches wide. MiRAE also has the ability to move its neck with two degrees-of-freedom (pan and tilt), though this ability was not used in the experiments described here.

MiRAE's programming code is written as a C++/Arduino library, and easily allows facial expressions to be made with varying degrees of motion for each individual facial component (as a variable passed into the function calls). These programming libraries, along with a construction manual for MiRAE, are available from the lab website (http://r-house.soic.indiana.edu) and the first author's personal website (http://www.caseybennett.com/Research.html), in order to facilitate experimental replication.

4.2.3 Experimental Design

For the two experiments reported here, the first (#1a and #1b) examined the effects of culture, and the second (#2) examined the effects of context across culture. Experiment-specific details are provided in Section 4.3. Here we describe the protocol and instruments used across all experiments in general.

First, we should be clear that all the experiments described here, as well the companion studies in Chapters 3 and 5 (from which some of the comparison data is derived) (Bennett & Šabanović, 2014; Bennett et al., 2014) are actually the same experiment – in terms of protocol, instruments used, and the robotic face – except for whatever independent variable was being manipulated (e.g. neck motion or added context stimuli). The only exception to this were some minor differences in the physical setup in Experiment 2 here due to the addition of the context stimuli. The robotic face was physically transported to and from Asia from the United States, so that all subjects could interact with the exact same artifact.

In all experiments, subjects observed the robotic face (and/or digital avatar, if applicable) making a randomized pre-set series of facial expressions (the six Ekman emotions, less Disgust). During each

experiment, the robotic face (and/or digital avatar, if applicable) made each expression for several seconds, then returned to a neutral face. A pause of 15 seconds was provided between expressions to allow participant to fill out the FEI instrument (see next paragraph). Participants simply watched the robot, i.e. there was no interactive behavior used in these experiments. The robot (nor avatar) did not speak or make affective sounds. There were no repetitions within subjects, nor did subjects participate in multiple conditions/experiments (to avoid any "priming effect", see Section 4.2.1). Subjects were randomly assigned to conditions/experiments. Finally, for terminological clarity, we will use the term "eye/brow movement" to refer to the simultaneous movement of both eyes and eyebrows henceforth.

For all experiments, the same Facial Expression Identification (FEI) instrument was used as in previous studies (Bennett & Šabanović, 2014), detailed in Section 3.2.2. The FEI contains three questions. First, subjects were asked to identify the expression (Question #1) and to rate the strength of expression (Question #2). The FEI used a similar 7-option forced-choice design for Question #1 as was used in studies with Kismet, Eddie, etc. for comparability purposes (Breazeal, 2003; Sosnowski et al. 2006), although there are some issues with the forced-choice design (Nelson & Russell, 2013; Barrett et al., 2011; Fugate, 2013). The FEI also asked subjects an additional question (Question #3) for each expression, allowing (but not requiring) them to select one or more "other expressions" they thought the robot might be displaying beyond the primary one in Question #1, if desired (see [Bennett & Šabanović, 2014] or Chapter 3 for a complete description). This is the basis for the *main accuracy* (Question #1) and *other accuracy* (Question #3) in subsequent tables. The FEI is available online (in both English and Japanese) at the lab website http://r-house.soic.indiana.edu (English version: http://r-house.soic.indiana.edu/mirae/FEI_Instrument.docx).

Additionally like the previous studies, both the Godspeed (Bartneck et al., 2009) and Negative Attitudes towards Robots (NARS: Nomura & Kanda, 2003) scales were collected to evaluate user perceptions. The NARS is a commonly used metric in human-robot interaction (HRI) research, developed to measure people's attitudes towards robots *in general* and consisting of three subscales: situation of interaction, social influence of robots, and emotion in robots during interaction (Nomura et

al., 2006). The NARS has often been used prior to a human-robot interaction to evaluate whether and how pre-existing attitudes affect people's behavior towards robots, as well as before and after interaction to see if the interaction itself has changed people's general attitudes toward robots. Our use of the NARS in this study was in the former sense. The Godspeed Scale was designed to gauge people's perceptions of *specific* robots and consists of five subscales: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. It is generally used in participant evaluations of robots they interact with or see; in our case we used it to measure people's perceptions of MiRAE following their interaction with it. Psychometric analyses of the NARS (Nomura et al., 2006) and Godspeed (Bartneck et al., 2009) have been previously provided, with Cronbach Alpha values consistently above 0.7. The NARS was collected prior to the interaction, the Godspeed after the interaction. Note that no significant differences in the overall NARS were found, so it will not be discussed further in this chapter. For brevity, the Godspeed will only be discussed for Experiment 1a here.

All Westerner subjects and East Asian (living in the US) subjects were administered all forms, including the FEI instrument, in English. The East Asian subjects were all US university students who had passed an English proficiency entrance exam (TOEFL) prior to admission. The Japanese (living in Japan) subjects were administered the forms translated into Japanese. The 7 emotion-label options on the FEI instrument were translated into Japanese as: 怒り (*ikari* - Anger), 幸せ (*shiawase* - Happy), 悲しい (*kanashii* - Sad), 恐怖 (*kyofu* - Fear), 驚き (*odoroki* - Surprise), 嫌悪感 (*keno-kan* - Disgust), 退屈 (*taikutsu* - Bored).

4.2.4 Analysis

The analysis of data varied by experiment in accordance with the number of groups and conditions in each experiment. This included *t*-tests for Experiment #1a and ANOVAs for Experiments #1b and #2. Effect sizes are reported using Pearson's *r*. Specifics for each experiment are provided in the relevant subsections of Section 4.3.

Previous evaluation of statistical power suggested an *a priori* power estimate somewhere in the range 0.6 (Bennett & Šabanović, 2014), which is capable of detecting modest effects (but not smaller ones). However, since we had no basis for projecting effect sizes for most of the hypotheses reported here, it is only an estimate. Post-hoc calculations of statistical power were thus also performed, which may be informative for future experiments. For Experiment #1a, observed power was 0.64. For Experiment #1b, power was 0.98 for expression variant, and 0.32 for culture (not surprising given the small differences across cultural groups). For Experiment #2, power was 0.67 for context effects, and 0.8 for culture. Given sufficient time and money, replicating the results here with a larger study would be of great interest.

4.3 Experiments

4.3.1 Experiment 1a

4.3.1.1 Experiment 1a – Methods

Experiment #1a used a Japanese sample (n=15) to *replicate a previously reported study of the baseline facial expression identification results of the same robot face with Westerners* (see experiment 1 in Chapter 3 [Bennett & Šabanović, 2014]), in order to provide a baseline comparison and ground the results of Experiment #1b. The hypothesis, based on previous research (Section 4.1.2), was that there would be significant differences in recognition accuracy across cultures. This experiment also involved subjects observing a digital avatar designed to appear nearly identical to the embodied robotic face, as shown in Figure 3.4 in Chapter 3 (see [Bennett & Šabanović, 2014] or Chapter 3 for complete digital avatar description). The aim was *not* to build a state-of-the-art digital face, but to build a minimalistic avatar whose appearance and motion closely resembled the embodied robot face, for comparison purposes. The order in which subjects saw the robot and avatar (robot-1st, avatar-2nd or avatar-1st, robot-2nd) was randomized. The digital avatar was *only* used in this experiment (#1a). The experiment was exactly identical to the previous study, except for the use of Japanese subjects rather than Westerners.

For both Experiment #1a and #1b (see below), subjects observed the robotic face (and digital avatar in #1a) making a randomized pre-set series of facial expressions (the six Ekman emotions, less Disgust). As detailed in the previous chapter (Section 3.4.2), Disgust is problematic since most studies on robotic facial expressions don't actually use the Ekman "Nose Wrinkle" Disgust expression based on the FACS (e.g. Kismet [Breazeal, 2003] and Eddie [Sosnowski et al., 2006]) but rather use a contrived "Lip Twist" expression as a substitute, or do not use Disgust at all (e.g. Geminoid [Becker-Asano & Ishiguro, 2011]). To our knowledge, no robotic face has yet convincingly implemented an empirically-validated, FACS-based Disgust expression capability. In short, further work is needed.

4.3.1.2 Experiment 1a – Analysis

For Experiment #1a, we used *t*-tests (independent samples, two-tailed, equal variances not assumed) to test for differences between the original Western participants in the previously reported study described in Chapter 3 (Bennett & Šabanović, 2014) and the Japanese subjects evaluated in this study.

4.3.1.3 Experiment 1a - Results

Results, for both the embodied robotic face and the digital avatar, are shown in Table 4,1 for both the Japanese and Western subject samples (Western results reproduced from experiment 1 in Chapter 3 [Bennett & Šabanović, 2014]). A few things are notable. First, except for Fear, the identification accuracy is nearly identical for the Westerners and the Japanese, despite the fact that the facial expressions in this experiment were based on the Ekman FACS system that is purportedly biased towards Western displays of emotion (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). Fear is clearly different between the two groups (43% vs. 0%), and the Japanese clearly had trouble identifying it. However, it should be noted that – even among Westerners across an array of humanoid robotic faces (MiRAE, Kismet, Eddie, BERT, Feelix, Geminoid) – Fear is only identified on average 34% of the time (see Table 1 above) (Bennett & Šabanović, 2014). *T*-tests between the two groups for overall

accuracy were significantly different when including Fear ($t(43)=2.65$, p=.011, effect size=0.54), but not significant without it ($t(43)=0.53$, p=.601).

Table 4.1: Experiment 1a – Main Results

		Western			Japanese		
	Expression	Main Accuracy	Other Accuracy	Strength Rating	Main Accuracy	Other Accuracy	Strength Rating
Embodied	Happy	96.7%	96.7%	7.31	100.0%	100.0%	5.86
	Sad	100.0%	100.0%	8.30	86.7%	100.0%	7.67
	Anger	86.7%	93.3%	7.25	100.0%	100.0%	6.47
	Fear	43.3%	63.3%	6.25	0.0%	6.7%	N/A
	Surprise	96.7%	100.0%	7.96	93.3%	93.3%	5.93
Digital	Happy	100.0%	100.0%	6.93	100.0%	100.0%	4.67
	Sad	100.0%	100.0%	8.09	86.7%	93.3%	7.46
	Anger	100.0%	100.0%	7.98	100.0%	100.0%	8.07
	Fear	53.3%	66.7%	6.38	0.0%	20.0%	N/A
	Surprise	86.7%	100.0%	7.22	73.3%	100.0%	5.09

The identification results for the digital avatar followed similar patterns (significantly different with Fear, non-significant without). Strength ratings (not including Fear, since it was never identified by Japanese) were significantly different ($t(43)=2.86$, p=.008, effect size=0.41), with Westerners having higher average ratings (7.7 vs. 6.4).

Table 4.2: Experiment 1a – Godspeed

	Western	Japanese			
Category	Embodied	Embodied	*t*-value	Sign.	Effect Size
Anthropomorphism	2.26 (.84)	2.89 (.57)	2.97	0.005*	0.41
Animacy	2.44 (.81)	3.24 (.58)	3.78	0.001*	0.50
Likeability	3.58 (.62)	3.77 (.41)	1.24	0.221	
Perceived Intelligence	2.86 (.81)	3.15 (.47)	1.49	0.143	
Perceived Safety	3.83 (.69)	3.00 (.42)	4.99	0.000*	0.59

Mean Values for both Western and Japanese subjects are shown, with standard deviations in parentheses. T-test values are provided to the right, with statistically significant differences ($p < 0.05$) are starred with an asterisk. Effect sizes are provided for any significant differences.

Godspeed ratings were also evaluated between the two groups for the embodied robotic face. These can be seen in Table 4.2. Several categories were significantly different between Japanese and Westerners, with anthropomorphism and animacy being rated higher by the Japanese and perceived safety being rated higher by Westerners. It is not clear exactly why this is the case. The pattern was identical for the digital avatar (data not shown).

4.3.2 Experiment 1b

4.3.2.1 Experiment 1b – Methods

Experiment #1b *evaluated two cultural variants of the baseline robotic facial expressions – an "East Asian" variant and a "Western" variant" – based on the "Emoticon hypothesis" and previous research findings that posits that East Asians focus more on the eyes and Westerners more on the mouth in interpreting facial expressions* (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). The hypothesis was that East Asians would have higher recognition accuracy for the "East Asian" variant, and the Westerners would have higher recognition accuracy for the "Western variant." In short, this resulted in the eye/brow facial feature motion being effectively turned off for the Western expressions, and the mouth facial feature motion being effectively turned off for the East Asian expressions. The exception was Anger – where the only movement in the original was in the eyes and eyebrows – which was left the same between the two variants (since there was no mouth movement to manipulate). By "effectively", we mean that the motion was set to ~10% of the original motion, so as to still be perceptible but so small as to not indicate any particular expression. Previously in Chapter 3, we have shown that reducing the degree of motion by as much as 50% for the robot face holistically (i.e. all facial features simultaneously) had no effect on human perception of affective expression (Bennett & Šabanović, 2014). The 10% motion was in effect a small twitching motion, and was tested (for all facial features simultaneously) with several lab personnel prior to the experimental phase to verify that they conveyed no recognizable emotion/expression.

For Experiment #1b, three groups of participants were recruited each containing 16 individuals (in total, n=48) for each cultural group (see Section 4.2.1). Each group was randomly divided in half into two sub-groups (n=8), each of which saw only one of the variants. In other words, we had 6 sub-groups that varied by both the culture of the subjects and the facial expression variant observed.

1.1.1 4.3.2.2 Experiment 1b – Analysis

For Experiment #1b, we used a two-way, fixed-effects, between-subjects ANOVA to test for differences between the three cultural groups and the two cultural variants of facial expression. Post-hoc Bonferroni *t*-tests were used to determine the source of any differences.

4.3.2.3 Experiment 1b - Results

Overall results for Experiment #1b are shown in Table 4.3 below. Of note, we point out the similar identification patterns for Fear between the Japanese from Japan and the native East Asians living in the United States.

Table 4.3: Experiment 1b – Main Results

		Western			Japanese			East Asian		
Expression Variant	Expression	Main Accuracy	Other Accuracy	Strength Rating	Main Accuracy	Other Accuracy	Strength Rating	Main Accuracy	Other Accuracy	Strength Rating
Western	Happy	100.0%	100.0%	7.38	100.0%	100.0%	5.63	100.0%	100.0%	7.38
	Sad	100.0%	100.0%	7.25	75.0%	100.0%	7.00	75.0%	87.5%	8.33
	Anger	100.0%	100.0%	7.25	87.5%	87.5%	6.28	87.5%	87.5%	7.71
	Fear	37.5%	37.5%	6.67	0.0%	12.5%	N/A	0.0%	0.0%	N/A
	Surprise	100.0%	100.0%	7.38	50.0%	50.0%	4.25	100.0%	100.0%	7.60
East Asian	Happy	50.0%	50.0%	5.50	62.5%	87.5%	5.00	50.0%	50.0%	7.00
	Sad	62.5%	87.5%	7.80	87.5%	87.5%	5.57	100.0%	100.0%	7.50
	Anger	87.5%	87.5%	8.00	100.0%	100.0%	7.38	75.0%	75.0%	8.33
	Fear	12.5%	37.5%	6.00	0.0%	12.5%	N/A	0.0%	12.5%	N/A
	Surprise	62.5%	100.0%	7.00	37.5%	62.5%	5.33	37.5%	75.0%	8.33

The results from Table 4.3 are succinctly summarized in Table 4.4. In brief, all of the cultural groups struggled to identify the East Asian expression variants (eye/brow movement only), with accuracy averaging 53.3%. Identification of the Western expression variants did vary across groups, with the

Westerners having higher values, the Japanese lower values, and the East Asians living in the US somewhere in between. This pattern held even when Fear was removed, as well as Anger (which was unchanged between the variants, see Section 4.3.2.1). Strength ratings, however, were consistent across cultural groups for different expression variants.

Table 4.4: Experiment 1b – Summary

		Western	Japanese	East Asian	
Expression Variant					**Average**
Western	Main Accuracy	87.5% (10.4)	62.5% (16.6)	72.5% (14.9)	74.2%
	Strength	7.26 (.64)	6.01 (1.20)	7.58 (0.97)	7.01
East Asian	Main Accuracy	52.5% (18.3)	57.5% (16.6)	50.0% (20.9)	53.3%
	Strength	7.48 (1.64)	5.97 (1.52)	7.57 (1.55)	6.95

Mean values are provided for each cultural group/condition, with standard deviations in parentheses.

These patterns were investigated for statistical significance via a two-way ANOVA (described in Section 4.2.4). The results are shown in Table 4.5. Significant effects on accuracy were found for expression variant ($F(1,42)=17.43$, $p<.001$) but not for cultural background. The interaction effect was near significance ($F(2,42)=3.04$, $p=.058$), but not below the .05 threshold. It is possible that a larger sample size might return a significant result for the interaction effect, however. Strength ratings showed the opposite, significant variation due to cultural background, but not due to expression variant. Post-hoc test showed the significant strength differences were between the Japanese and both other groups, but not between the Westerners and East Asians living in the US.

Table 4.5: Experiment 1b – ANOVA

	Main Accuracy		Strength Rating	
	F	**Sign.**	**F**	**Sign.**
Culture	1.59	0.216	6.96	0.002*

Exp. Variant	17.4	0.000*	0.026	0.873
Culture * Exp. Variant	3.04	0.058	0.047	0.954

F-values attaining statistical significance ($p < 0.05$) are starred with an asterisk.

To summarize the first experiment (#1a and #1b), Westerners were better at identifying robotic facial expressions from mouth movement alone than Japanese subjects (East Asians living in the US fell in between). However, none of the subject groups were significantly better at identifying facial expressions from eye/brow movement alone. Moreover, when expressions were made normally with all facial features (eyes, brows, mouth), there were no significant differences between Westerners and Japanese, except for Fear.

4.3.3 Experiment 2

4.3.3.1 Experiment 2 – Methods

For **Experiment #2**, *we evaluated the effect of the broader interaction context on participants' perceptions of the face robot's expressions*. The hypothesis, based on previous research (Section 4.1.2), was that context would have a larger effect on recognition accuracy for East Asians than Westerners. Subjects watched a series of videos alongside the robot-face. The videos were taken from a previous psychological study (Gross & Levenson, 1995), which validated the clips' consistent ability to elicit certain emotional responses that tie to the Ekman emotions (Happy, Sad, Anger, etc.). The same video clips were obtained in digital format and cut to length using the FRAPS software (version 3.5, http://www.fraps.com/), for the same five affective expressions as in Experiments #1a and #1b. The clips used were generally a couple minutes long, from the following (see Table 1 in [Gross & Levenson, 1995] for specific scenes/times): *When Harry Met Sally* (Happy), *Bambi* (sad), *The Shining* (Fear), *Sea of Love* (Surprise), and *Cry Freedom* (Anger). The robot face was set to automatically trigger the facial expression ("react") to match the elicited emotion of each video, at an appropriate time-point (as judged by the researchers) in the latter half of each video. Subjects were then asked to identify the expression of

the robot between videos, as well as rate the strength of expression (see below). Aside from the inclusion of the video-watching, this experiment was identical to Experiments #1a and #1b in terms of protocol. As noted in Section 4.2.1, this experiment included two groups: Westerners and native East Asians living in the U.S. (n=30, 15 per group). Results were compared with non-context-exposed Western/Asian subjects from previous experiments (Western: n=30, Asian: n=15), with the experimental protocol being exactly the same except for the addition of context stimuli (i.e. the movie clips) during the interaction (Bennett & Šabanović, 2014; Bennett et al., 2014).

In terms of the experimental setup, the robot was placed so as to create a triadic interaction between robot, computer screen, and human subject (i.e. roughly a triangular type arrangement). Every subject was explicitly instructed prior to the experiment that the robot would "watch the video with them, and react to the video at some point, and that they should mark down the robot's reaction." A written briefing script was used by investigators to facilitate consistency.

4.3.3.2 Experiment 2 – Analysis

For Experiment #2, we used the same ANOVA approach as in Experiment #1b (Section 4.3.2.2). This included a two-way, fixed-effects, between-subjects ANOVA to evaluate differences between the two cultural groups used (Westerners and East Asians living in the U.S.) and the two context exposure conditions (context-exposed vs. non-context-exposed). Post-hoc Bonferroni *t*-tests were used to determine the source of any differences.

4.3.3.3 Experiment 2 - Results

The main results for Experiment #2 are shown in Table 4.6. The results show a significant increase in facial expression identification when context is supplied. This was primarily due to Fear identification, which increased from 43.3% to 100% in Westerners and from 0% to 80% in East Asians, as most of the other expressions were already in the 90-100% accuracy range without context. Of note,

there was also a notable drop in identification of Happy in East Asians, which we discuss below. The results from Table 4.6 are summarized in Table 4.7.

Table 4.6: Experiment 2 – Main Results

		Western			East Asian		
	Expression	**Main Accuracy**	**Other Accuracy**	**Strength Rating**	**Main Accuracy**	**Other Accuracy**	**Strength Rating**
Non-Context	Happy	96.7%	96.7%	7.31	100.0%	100.0%	5.86
	Sad	100.0%	100.0%	8.30	86.7%	100.0%	7.67
	Anger	86.7%	93.3%	7.25	100.0%	100.0%	6.47
	Fear	43.3%	63.3%	6.25	0.0%	6.7%	N/A
	Surprise	96.7%	100.0%	7.96	93.3%	93.3%	5.93
Context	Happy	93.3%	93.3%	5.53	60.0%	80.0%	5.67
	Sad	100.0%	100.0%	8.67	100.0%	100.0%	8.07
	Anger	93.3%	100.0%	7.50	93.3%	93.3%	7.50
	Fear	100.0%	100.0%	6.47	80.0%	100.0%	6.79
	Surprise	80.0%	100.0%	8.18	80.0%	100.0%	6.71

Table 4.7: Experiment 2 - Summary

		Western	East Asian	
				Average
Non-Context	Main Accuracy	84.0% (14.2)	74.7% (9.2)	80.9%
	Strength	7.65 (1.36)	6.72 (1.36)	7.19
Context	Main Accuracy	92.0% (12.6)	82.7% (16.6)	87.3%
	Strength	7.32 (1.54)	7.25 (1.27)	7.28

Mean values are provided for each cultural group/condition, with standard deviations in parentheses.

These patterns were investigated for statistical significance via a two-way ANOVA (see Section 4.2.4). The results are shown in Table 4.8. Significant effects on accuracy were found for both culture ($F(1,71)=8.02$, $p=.006$) and context ($F(1,71)=5.89$, $p=.018$). The interaction effect was not significant. There were no significant effects on strength ratings. In other words, both context and culture significantly affected facial expression perception, but context effects of similar size were present regardless of cultural background.

Table 4.8: Experiment 2 - ANOVA

	Main Accuracy		**Strength Rating**	
	F	**Sign.**	**F**	**Sign.**
Culture	8.02	0.006*	3.75	0.057
Context	5.89	0.018*	0.55	0.463
Culture * Context	0.00	1.000	3.05	0.085

F-values attaining statistical significance ($p < 0.05$) are starred with an asterisk.

There were some differences across cultures, notably in the identification of Happy. Many of the East Asians identified the expression as Disgust, despite the fact that the robot expression was unchanged from previous experiments. We attribute this to the context stimuli used for that emotion (the fake orgasm scene from the film *When Harry Met Sally*), which created some discomfort and/or embarrassment in several of the East Asian participants (a few of them reported this, unsolicited, to the researcher). We also qualitatively evaluated the patterns of emotions identified as "other expression" on the FEI (Question #3, see Section 4.2.3) for the Westerners, which asked what if any other emotions and expression might represent beyond the primary one (data not shown for brevity). Of note, there were much higher rates of responses of Disgust for the Anger expression (80% vs. 50%, context vs. non-context) as well as higher rates of Fear for the Surprise expression (80% vs. 30%). Taken into account

with the effects of context on Fear identification, these results are interesting, seeing as the robotic facial expressions themselves did not change at all.

One issue here is that many of the emotional facial expressions were already at or near 100% accuracy without context. However, a companion study to this described in Chapter 5 (Bennett et al., 2014) looked at both congruent vs. incongruent context, and showed significant differences across all emotions, except (curiously) surprise. When provided incongruent context, subjects had a higher mis-recognition rate for all emotional facial expressions, revealing differences across most of them. In short, the results from Experiment #2 presented here have been partially replicated, providing further evidence for the conclusions here.

4.4 Discussion

4.4.1 General Discussion

We conducted experiments on the effects of both culture and context on perceptions of robotic facial expressions during human-robot interaction. The first set of experiments looked at the effects of culture and hypothesized culturally-variant expressions, while a second looked at the interaction of culture and context. The results are summarized below (main findings underlined).

Previous research on cultural differences in facial expressions has suggested that East Asians focus on the eyes more when viewing facial expressions in others, largely based on the "Emoticon hypothesis" and evidence from visual fixation experiments (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). However, more recent research has disputed this evidence (see Introduction) (Arizpe et al., 2012; Blais et al., 2012; Peterson & Eckstein, 2012). Here we investigated this hypothesis using robotic facial expressions. Our findings indicate that the issue is more complicated than those previous hypotheses might suggest. In the first experiment (#1a), we found that, except for Fear, Westerners (living in the US) and Japanese (living in Japan) were not significantly different when facial expressions were made normally (i.e. all facial features utilized). A second experiment (#1b) studied two hypothesized culturally-variant facial expressions using only mouth movement (Western) and only

eye/brow movement (East Asian). We found that even though Westerners were relatively better at discerning facial expressions from mouth movement alone, Japanese were just as poor at identifying facial expressions from eye/brow movement alone, with East Asians living in the US falling somewhere in between.

These findings suggest that even if East Asians (such as Japanese) are looking at the eyes more when viewing other faces, it may be for reasons other than judging affect (as recently argued [Arizpe et al., 2012; Blais et al., 2012; Peterson & Eckstein, 2012], see Section 4.1.2). The results could also suggest that East Asians utilize more holistic facial feature information to judge affect in other faces. This conforms to existing research suggesting that East Asians have a more holistic cognitive style that encourages extracting meaning from relationships of multiple relevant points of attention, rather than from individual components of a scene (e.g. Nisbett et al., 2001). As for Fear, clearly current robotic facial expressions based on Ekman's FACS system appear to be ineffective for East Asians. However, we note that, even among Westerners, identification rates for Fear only average 34% across a range of humanoid robotic faces (see Section 4.3.1). Furthermore, Fear has been previously shown to elicit lower levels of rater agreement among research participants viewing human facial expressions, across multiple cultural groups (Biehl et al., 1997). Why this is the case remains uncertain. It is one of the most complex expressions to produce in terms of the number and control of muscles used. Its infrequency of use in daily life might also be a factor in the difficulty people have in identifying it.

The differences between Japanese and other subjects in terms of their ratings of the strength of the emotions portrayed by MiRAE can be compared to previously documented evaluations of human emotions, in which Japanese participants rated expressions as having a lower intensity than Americans (Biehl et al., 1997, pp.17). These differences in intensity might be related to the learned nature of display and decoding rules for emotional expression and to different socially normative acceptability of different expressions and levels of intensity of emotional expression in different cultures (e.g. Matsumoto, 1992). This would follow findings from previous work on identification of human emotions (e.g. Friesen, 1973),

in which Japanese subjects masked negative emotions with smiles. This idea merits further study in human-robot and human-computer interaction.

As for context effects (Experiment #2), both context and culture significantly affected facial expression perception, but context effects of similar size were present regardless of cultural background. In other words, context improved recognition accuracy across cultures, and to practically the same degree. In particular, Fear – a notoriously difficult emotion to convey via robotic facial expressions – increased to nearly 100% with added context, regardless of cultural background of the subjects. These findings concur with previously reported context effects in both humans/avatars (Righart & de Gelder, 2008; Barrett et al., 2011; Lee et al. 2012) as well as robots (Zhang & Sharkey, 2011). We were also able to replicate these effects in a companion study described in Chapter 5, in which we looked at the effects of both incongruent and congruent context on people's perceptions of a robots affective facial expressions which showed significant differences across all emotions, except for surprise (Bennett et al., 2014).

These findings are potentially useful for constructing robotic faces that may interact via facial expressions with different cultures, as well as for designing interactive robots or avatars that utilize facial expression identification across different cultures.

4.4.2 Implications

The results of these studies presented here have a number of potentially intriguing implications. The context effects seen in Experiment #2 seem to suggest that human subjects may be *projecting* their own internal emotions onto the facial expressions of others, including robots. Given that the context videos have been previously shown to reliably elicit certain emotions in human subjects, and the fact that the robotic facial expression stimuli were exactly the same across conditions, we arrive at such an interpretation. This concurs with other recent research findings into the role of emotion formation and cognition in human-human interaction, which may be informative for human-robot interaction.

There is evidence that such projection may in fact be a key part of such affective communication between humans. Lindquist and Gendron (2013) have proposed a "Construction hypothesis" of emotion,

which is essentially a dynamical systems view of emotion perception, where language, emotion labels, and/or other context may ground our perceptions of both emotion and facial expressions. As they noted, "this leaves open the possibility, as the data reviewed here suggest, that emotions seen on other people's face are constructed in the mind of the perceiver" (pp.70). Barrett et al. (2011) make a similar dynamical systems argument for the effects of context (including language). They also point out that context – from a human cognition standpoint – really relates to the way the brain makes predictions using visual (or other sensory) data. Recent studies provide further evidence for this explanation. Righart and de Gelder (2008) found context biases the pattern of error responses in facial expression identification of human faces. This is similar to our finding for "other expression" attribution patterns in Experiment #2 (see Section 4.3.3.3). Elsewhere, Lee et al. (2012) found evidence of inter-individual differences modulating the effects of context on facial expression identification.

More broadly, this relates to scholarship on the cultural aspects of social cognition and technology (e.g. Hall, 1977; Shore, 1996; Nisbett, 2001, 2003), which suggests that culturally appropriate social cues, including modes of communication, temporal interaction patterns, and expectations regarding affective display, are foundational to human sociality and that a breach of cultural norms can provide a significant barrier to successful interaction. The results here support this perspective. In previous work, Šabanović (2010, 2014) showed that various cultural models of affect, social cognition, and interaction with technology are embodied in social robot design in both explicit and implicit ways. Such culturally-situated design choices, however, generally reproduce stereotypical notions of cultural difference rather than developing technologies that can fit empirically based constructions of the cultural dynamics of social interaction. A more reflexive understanding of culture's role in social interaction suggests a dynamic model, in which cultural models are not simply copied, but are "repeatedly assembled": core cultural models dynamically change as they are adapted to fit contemporary circumstances (Caporael, 1997). In the development of affectively-expressive interactive technologies, this viewpoint supports the adoption of a dynamic and relational model of affect construction, which would address the situated nature of cultural expression within social interaction.

Such a dynamical systems view of emotion and affective interaction also feeds into concepts about embodied cognition and the development of robotic (and/or other artificially intelligent) interactive systems. If, as Barsalou et al. and others have suggested, higher cognition is primarily intended for the mediation of perception and action via dynamic mechanisms, then emotions are biasing factors that prime our anticipatory response systems for subsequent events (Barsalou et al., 2006; Beer, 2000). Indeed, affective communication, including facial expressions, could even be seen as a kind of *context* itself in that view. In a counter-intuitive sense, they are context created by social interaction for the explicit purpose of facilitating further social interaction. For instance, if the goal is to communicate information about food or dangers in the environment, then affective communication can provide enabling context that simplifies the need for interpretation and understanding of *future* sensory signals (including social ones) in terms of behavior/action-selection (Barsalou et al., 2006). This is an equivalent argument to Clark (2013) that we utilize social cues to "load the dice" in terms of minimizing costly prediction errors and facilitating our own cognition (see Section 3.2 therein). Or, in other words, self-structuring of sensory information into a rolling "cognitive niche" (Sterelny, 2007; Clark 2013). From another angle, this can be seen as a social-interaction-based form of cognitive scaffolding, in the vein of Gibson and visual scaffolding (Gibson, 1979). This also concurs with other recent suggestions of social cognition as an emergent phenomenon from social interaction itself (De Jaegher et al., 2010; Froese & Ziemke, 2009; Froese & Di Paolo, 2010; McGann et al. 2013). The socio-cultural and cognitive science literature both point in the same direction – that affective interaction is not necessarily about communicating some "information" about the current state of the world, but rather about biasing what we expect to experience next, both internally and externally.

Such evidence holds intriguing possibilities for robotics. If emotions perceived in others are indeed an internal construct in the mind of the perceiver based on a number of dynamic perceptual and cognitive processes, then the question exists of how we might take advantage of those processes to facilitate human-robot interaction. Facial expressions, or other direct forms of communication, may only be one piece of the puzzle. The results here suggest that, if we can induce appropriate context effects, it

may be possible to create *culture-neutral models* of robots and affective interaction. Inducement of such context effects, for instance, could stem from creation of environmental conditions that correspond with certain attractor basins in human cognition. Individual-specific models could potentially be learned via machine learning methods, allowing the robot to adapt to individual people. Such an approach may be an alternative and/or potentially more effective path than direct affect communication (e.g. trying to make culturally-specific expressions or cues for every single cultural group). This is a similar concept as approaches being explored for dynamic/adaptive production of synthetic emotions in robots and intelligent agents, although from the polar opposite direction (Picard, 1997; Canamero, 2005; Asada et al. 2009; Bosse et al. 2010).

4.4.3 Limitations

There are some limitations to this study. For example, there are confounding factors we cannot rule out cross-culturally, including the effects of language. Different emotion-label words may have different cultural connotations (a.k.a. linguistic relativity), which can affect response answers (Perlovsky, 2009; Ruttkay 2009; Davies et al. 1998). Such linguistic relativity might also tie into the aforementioned view of emotions and facial expressions from a dynamical systems and embodied cognition perspective. Additionally, there are issues with the forced-choice response design – although given how common that methodology is in this area (Nelson & Russell, 2013; Barrett et al., 2011; Fugate, 2013), it becomes difficult to directly compare results to other work if other designs are utilized. Moreover, from a dynamical systems perspective, categorization is a fundamental aspect of higher cognition, as categories relate to attractor basins for otherwise continuous-valued perceptions. In that sense, it is challenging to understand or study any aspect of human cognition without categorization.

Caution should also be taken in generalizing the results seen here. There may be, for instance, tasks other than affective facial interaction where these results do not apply. Those tasks may necessitate less minimalist face/facial components for a robot, or other non-facial (i.e. bodily) cues in order to communicate information.

Other limitations include the sample size – some statistical tests here, particularly several that were near the .05 threshold for significance, might attain significance if these experiments were replicated with larger sample size. There were also some issues with the film clips used (particularly happiness, as noted in Section 4.3.3.3), though they were chosen because they had been previously validated to elicit certain emotional responses in a published study (Gross & Levenson, 1995). Those issues may hint at the interplay of cultural norms and context-based emotional cues. From a broader perspective, this study also leaves a number of unanswered questions that deserve further study, e.g. more deeply investigating synergistic effects between culture and context. We discuss some of these in the next section.

4.4.4 Future Directions

This work suggests a number of future directions for research. For instance, the congruence between context and facial display of emotion may have a variable effect on emotion recognition cross-culturally (Boiger & Mesquita 2012). Such work is detailed in a companion study (Bennett et al., 2014), described in Chapter 5. Temporal dynamics in social cognition and interaction may also play a role. Modeling of those dynamics, in the spirit of Beer (1995), Auvray et al. (2009), and Ikegami & Suzuki (2008), may help elucidate fundamental building blocks of minimal cognition and social interaction. Moreover, exploring such interaction dynamics, both in laboratory and "robots in the wild" experiments, is warranted (Šabanović et al. 2006, MacDorman & Ishiguro 2006). A study exploring the latter involved placing an interactive version of MiRAE – which could respond to the presence of people in its vicinity – into a month-long public art display to explore more naturalistic, free-form social interaction. This is detailed in Chapters 6 and 7. Finally, design aspects that may affect the interaction and/or affective communication can be explored with 3D printing, allowing for rapid prototyping and testing of component design that vary in terms of shape, size, texture, range of motion, realism, etc. Understanding how certain design choices affect human-robot interaction, and their interplay with contextual factors, is fundamental. We are currently working on a project involving such 3D printed robotic face design.

Many other opportunities exist in terms of contextual effects, and how they are elicited, that may inform our understanding of social interaction and the artificial construction thereof.

Chapter 5

Context Congruency and Robotic Facial Expressions: Do Effects on Human Perceptions Vary across Culture?

This chapter builds on the previous chapter's exploration of the effects that environmental context has on human perception of robotic facial expressions across culture. The question here is what would happen if context congruency was varied – if the emotion expressed by the context was sometimes congruent, sometimes incongruent, with the robotic expressions?

Abstract. We performed an experimental study (Bennett et al. 2014) of the effects of context congruency on human perceptions (n=48) of robotic facial expressions across cultures (Western and East Asian individuals). We found that context congruency had a significant effect on human perceptions, and that this effect varied by the emotional valence of the context and facial expression. Moreover, these effects occurred regardless of the cultural background of the participants. In short, there were predictable patterns in the effects of congruent/incongruent environmental context on perceptions of robot affect across Western and East Asian individuals. We argue that these findings fit with a dynamical systems view of social cognition as an emergent phenomenon. Taking advantage of such context effects may ease the constraints for developing culturally-specific affective cues in human-robot interaction, opening the possibility to create culture-neutral models of robots and affective interaction.

5.1 Introduction

5.1.1 Background

A fundamental question for human-robot interaction (HRI) is whether – and to what degree – variables *external to the robot* affect the perceptions a human user has of what the robot is communicating. This includes affective interaction (Picard, 1997). For instance, environmental context (due to music, lighting, etc.) is known to elicit resonant emotions in people (Gross & Levenson, 1995). Certain colors of light elicit happiness, certain sounds evoke fear, certain scenes evoke surprise, and so forth. Moreover, the effects of such environmental context may vary depending on the characteristics of the person, e.g. their cultural background.

Many researchers have explored affective communication by robots, such as facial expressions (Breazeal, 2003; Sosnowski et al., 2006; Canamero & Fredslund, 2001; Bazo et al., 2010; Saldien et al., 2010; Becker-Asano & Ishiguro, 2011) and other less explicit emotional cues (Matsumoto & Okada, 2006; Kozima, Michalowski, & Nakagawa, 2009). As in interaction among humans, context effects can play a role in how people perceive such cues performed by a robot.

In the previous chapter (Chapter 4), we have empirically shown that context effects of similar size were present regardless of the participant's cultural background (Bennett & Šabanović, 2015). Providing context known to elicit matching emotions significantly improved human recognition of the robotic facial expressions over non-context experiments, even though the facial expressions were exactly the same in both conditions. The results suggested a form of *projection.* Emotions perceived in the faces of others – including robots – appeared to be an internal construct in the mind of the perceiver, based on a number of perceptual and cognitive processes (Bennett & Šabanović, 2015; Barrett, Mesquita, & Gendron, 2011). This was equally true across human subjects from Western and Asian cultural backgrounds.

In that previous study (Chapter 4), the provided context was always congruent with the robotic facial expression, i.e. the emotion elicited by the context was the same as the emotion communicated by the robot's facial expression. A separate, but related, question is what would happen if context congruency was varied – if the emotion expressed by the context was sometimes congruent, sometimes incongruent,

with the robotic expressions (Kret & de Gelder, 2010; Kret et al., 2013). In this study, we empirically explore *whether the effects of context congruency on human perceptions of robotic facial expressions vary across culture*.

5.1.2 Related Work

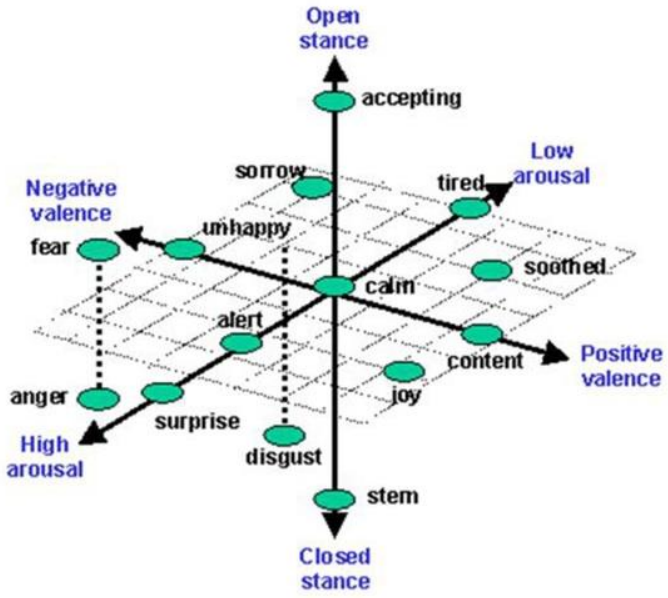
This work is informed by previous research on emotions, facial expressions, and robotic faces, which we review here along with scholarship on the interplay of both culture and context on affective interaction. Additional information can be found in Section 4.1.

5.1.2.1 Emotion, Facial Expressions, and Robotic Faces

In the section, we provide a brief overview (for brevity) of emotion and facial expressions, and their use in robotic faces. We have provided a more extensive overview of the scientific literature on robotic facial expressions and human emotion in Chapter 3, as well as previous papers (Bennett & Šabanović, 2014; Bennett & Šabanović, 2015; Bennett & Šabanović, 2013).

The scientific study of emotions in humans has a long and venerable history going back nearly two centuries (Darwin, 1872). Over the last half century, scholarly debate has focused on emotional facial expressions and how to classify them (Ekman, 2009; Nelson & Russell, 2013; Pantic, 2009; Cohn, 2010). A principle question is whether a basic set of universal human emotions (and their related facial expressions) exist across culture, gender, context, etc.? The study of facial expressions of emotion has evolved into two major camps during this time period: 1) Ekman et al., who argue for 6-7 "basic" *categorical* emotional expressions that are universal across cultures (Ekman, 2009), and 2) Russell et al., who argue that facial expressions are emergent states from a *continuous*, multi-dimensional space of affect (circumplex model), typically defined by three principle axes: valence, arousal, and stance (Figure 5.1) (Nelson & Russell, 2013; Schiano et al., 2000). Valence, which relates to the positivity/negativity of the emotion/expression, is of particular interest in the present study.

Figure 5.1: 3-Dimensional Affect Space (From [3])



Various robotic faces have been constructed over the last decade that integrate aspects of both Ekman and Russell's theoretical approaches (e.g. Breazeal, 2003; Sosnowski et al., 2006; Canamero & Fredslund, 2001; Bazo et al., 2010; Saldien et al., 2010; Becker-Asano & Ishiguro, 2011).

5.1.2.2 Culture and Affective Interaction

Numerous theories about the role of culture in affective interaction, including facial expressions, exist. One primary theory is the "Emoticon hypothesis", which posits cultural differences in facial expressions based on differences in emoticons between Western and East Asian cultures (e.g. East Asians focus more on the eyes, and Westerners more on the mouth) (Yuki, Maddux, & Masuda, 2007). A number of papers have studied visual fixation patterns as the basis for these putative differences in recent years, related to that hypothesis (Jack et al., 2009; Koda et al., 2010). However, more recent studies have provided evidence countering the use of such visual fixation patterns, noting that people are engaged in a range of information-gathering activities for a variety of purposes (not simply judging affect) when

looking at other faces (Arizpe et al., 2012; Blais et al., 2012; Peterson & Eckstein, 2012). Recent work in human-robot interaction has provided empirical evidence that also runs counter to this hypothesis (Bennett & Šabanović, 2015). In short, the empirical basis at this point for the Emoticon hypothesis is tenuous at best.

Broader socio-cultural research has examined the possibility of different "cognitive styles" in affective interaction across cultures that prescribe salient features of an individual's environment and appropriate modes of communication (Nisbett et al., 2001). Culturally variable "social-orientational models" may designate appropriate roles/behaviors within interaction as well as culturally-normative rules for displaying, perceiving, and experiencing affect (Shore, 1996). Along similar lines, Ekman, Friesen, and Izard themselves suggested a "Deception hypothesis" in the 1970's to explain culturally-based affective expression encoding rules (Ekman, 1971). More recently, Elfenbein has proposed a "Dialect hypothesis" for affective communication, which posits isomorphisms between affective expressions and linguistic distributions/development (Elfenbein, 2013).

5.1.2.3 Context Congruency and Culture

An ongoing debate in recent years is whether cultural differences influence the role context plays in affective interaction, including perceptions of facial expressions (Barrett, Mesquita, & Gendron, 2011; Righart & de Gelder, 2008; Lee, Choi, & Cho, 2012). Across cultures, context is considered important for discerning emotions, with evidence suggesting that without context cues emotion recognition decreases (Barrett, Mesquita, & Gendron, 2011; Barrett & Kensinger, 2010; Kitayama, Mesquita, & Karasawa, 2006). For example, Western participants (from the Netherlands) displayed faster reaction times to correctly identify emotions when a background image invoked an emotion congruent with displayed facial expressions. This trend varied by the valence of the expression (Righart & de Gelder, 2008). Recent work has shown the importance of context in perceptions of robotic facial expressions across cultures as well (Bennett & Šabanović, 2015; Zhang & Skarkey, 2011), also evidenced by the results in Chapter 4.

There is some research suggesting that people in Eastern Asian cultures pay greater attention to context than do Westerners. This has been shown on neutral tasks such as describing the contents of a fishbowl (Masuda & Nisbett, 2001) and on tasks of detecting emotion in faces (Ko et al., 2011; Masuda et al., 2008). In particular, the effects of context congruency varied across culture, having a greater effect on East Asians than Westerners.

However, these findings on cultural variability are subject to debate. They mainly involve looking at pictures of static faces and images on a computer screen, not direct interaction with a physically embodied human or robotic face. Thus, an open question is whether the effects of context congruency may vary across cultures in face-to-face interaction with a robot. We empirically explore that question here.

5.2 Methods

5.2.1 General Overview/Subjects

This chapter reports on a single experiment involving robotic facial expression recognition, in which we systematically varied the congruency of the environmental context in respect to the affective facial expressions made by the robot and the cultural background of human subjects.

Two groups of subjects participated in the study: native East Asians (living in the United States), and Westerners (i.e. Americans). We use the term "Westerners" here to be consistent with Jack et al. and others (Jack et al. 2009). The East Asians were a mixture of Japanese, South Korean, and Chinese college students, who had lived in the United States on average for 6 months (and generally no longer than one year) and had passed an English proficiency entrance exam (TOEFL). The Westerners were all American-born college students, primarily Caucasian. The gender mix was 58.3% females. Subjects were college age (18-25 years old). Results with subjects outside this age/gender composition, of course, may vary from those seen here. Most participants came from either the computer science or psychology programs.

A total of 48 subjects were recruited (n=48), 24 each for the two cultural groups. There were three experimental conditions (see Section 5.2.3), resulting in n=8 for each condition for each cultural group. Sample sizes were based on estimated effect sizes from previous studies (see Section 3.2.3) (Bennett & Šabanović, 2014; Bennett & Šabanović, 2015).

5.2.2 Robotic Face

The platform used here (MiRAE) is a minimalist robotic face that is capable of displaying a variety of facial expressions, described in Chapters 2 and 3.2.1 (Bennett & Šabanović, 2014). In previous studies (see previous chapters), MiRAE was shown capable of producing higher, or at least comparable, identification accuracy rates (with Westerners) for all expressions as a number of other robotic faces, including Kismet (Breazeal, 2003), Eddie (Sosnowski et al. 2006), Feelix (Canamero & Fredslund, 2001), BERT (Bazo et al., 2010), and the android Geminoid-F (Becker-Asano & Ishiguro, 2011, values from Table 5 therein), as shown in Table 3.2 in Chapter 3.

Examples of MiRAE displaying various facial expressions can be seen Figure 3.6 in Chapter 3. More details on MiRAE, its construction, and design philosophy can be found in Chapter 2 and Section 3.2.1.

5.2.3 Experimental Design

The experiment took place in the R-House HRI Lab at Indiana University, Bloomington. The experiment design was the same as in previously reported experiments (experiment #2 in Chapter 4, Bennett & Šabanović, 2015), except that the context congruency was varied in this case. After giving informed consent, subjects were asked to watch a series of videos alongside the robot-face. The videos were taken from a previously validated psychological study (Gross & Levenson, 1995), which verified the clips' ability to consistently elicit certain emotional responses that tie to the Ekman emotions (e.g. Happy, Sad, Anger). The same video clips were obtained in digital format and cut to length using the FRAPS software (version 3.5, http://www.fraps.com/), for the same five affective expressions as used in previous

experiments described in Chapters 3 and 4: Happy, Sad, Surprise, Fear, Anger ((Bennett & Šabanović, 2014; Bennett & Šabanović, 2015).

The clips used were generally a couple minutes long, excerpted from the following films (see Table 1 in [Gross & Levenson, 1995) for specific scenes/times): *When Harry Met Sally* (Happy), *Bambi* (Sad), *The Shining* (Fear), *Sea of Love* (Surprise), and *Cry Freedom* (Anger). The robot face was set to automatically trigger the facial expression ("react") to either match (*congruent*) or not match (*incongruent*) the elicited emotion of each video, depending on the experimental condition (see below). Expressions were triggered at an appropriate time-point (as judged by the researchers) in the latter half of each video. Subjects were then asked to identify the expression of the robot between videos, as well as to rate the strength of expression (see below). Results were compared with non-context-exposed subjects from previous studies described in chapters 3 and 4 (Bennett & Šabanović, 2014; Bennett & Šabanović, 2015).

The goal was to evaluate the effects of context congruency on human perceptions of robotic facial expressions. However, such effects may depend on the *degree* of incongruency, i.e. how similar the elicited emotion of the context is to the emotion of the facial expression. In this study, we define similarity based on emotional valence, which is a primary component of emotion classification systems (see Section 5.1.2.1). Previous studies have also suggested that the effects of context congruency may vary by valence (Righart & de Gelder, 2008). In order to account for similarity as a conflating factor, three experimental conditions were used, in which we "switched" certain expressions so that they were incongruent with the context (Table 5.1). Other expressions were left as congruent with the context. Each expression was shown only once for each subject, to avoid potential priming effects (see Section 3.3.3, and Bennett & Šabanović, 2014). For Condition 1, positive-valence emotional expressions (Happy, Surprise) were switched with each other. For Condition 2, negative-valence emotional expressions were switched (Sad, Fear, Anger). For Condition 3, we switched expressions *across* valence, so that positive-valence expressions were shown with negative-valence context, and vice versa (Fear was left congruent as a control).

Table 5.1: Experimental Conditions

	Expression Shown		
Context	**Positive Switch**	**Negative Switch**	**Cross Switch**
Happy	Surprise	-	Sad
Sad	-	Anger	Happy
Anger	-	Fear	Surprise
Fear	-	Sad	-
Surprise	Happy	-	Anger

** Entries with a dash were unchanged (i.e. context and facial expression were congruent)

For all experiments, the same Facial Expression Identification (FEI) instrument was used as in the previous studies described in Chapters 3 and 4 (Bennett & Šabanović, 2014; Bennett & Šabanović, 2015). The FEI contains three questions. First, subjects were asked to identify the expression (Question #1) and to rate the strength of expression (Question #2). The FEI used a similar 7-option forced-choice design for Question #1 as was used in studies with Kismet, Eddie, etc. for comparability purposes (Breazeal, 2003; Sosnowski et al., 2006), although there are some issues with the forced-choice design (see Barrett, Mesquita, & Gendron, 2011; Nelson & Russell, 2013; Fugate, 2013). The FEI also asked subjects an additional question (Question #3) for each expression, allowing (but not requiring) them to select one or more "other expressions" they thought the robot might be displaying beyond the primary one in Question #1, if desired (see Section 3.2.2 for a complete description). The FEI is available online at the author's personal website (http://www.caseybennett.com/research.html) or the lab website (http://r-house.soic.indiana.edu). Like previous studies described in Chapters 3 and 4 ((Bennett & Šabanović, 2014; Bennett & Šabanović, 2015), both the Godspeed and NARS scales were collected, but are not discussed here for brevity.

5.2.4 Analysis

The analysis of the data consisted of two separate parts in order to answer two primary questions: 1) whether the effects of context congruency on perceptions of robotic facial expressions varied by culture, and 2) whether such effects depended on the similarity of emotional valence of the facial expressions and the context.

For the first question, we used a two-way, fixed-effects, *within-subjects* ANOVA to test for differences between congruent and incongruent context across the two cultural groups. Repeated measures for each subject were the recognition accuracies of facial expressions for congruent context and for incongruent context.

For the second question, we used a two-way, fixed-effects, *between-subjects* ANOVA to test for differences in recognition accuracy between the three different conditions across the two cultural groups. The three conditions varied by which expressions were incongruent with the context, based on emotional valence (see Section 5.2.3). Post-hoc Bonferroni *t*-tests were used to determine the source of any differences across the conditions.

5.3 Results

This section is broken into two parts which each address one of the primary questions of the chapter (see Section 5.2.4).

5.3.1 Effects of Context Congruency across Cultures

One primary question was whether the effects of context congruency on perceptions of robotic facial expressions varied by culture. A summary of facial expression recognition accuracy rates by context congruency and culture is shown in Table 5.2. The "None" context values were taken from previously reported studies described in Chapters 3 and 4 (Bennett & Šabanović, 2014; Bennett & Šabanović, 2015).

Table 5.2: Recognition Accuracy by Context Congruency and Culture

Context	Western	East Asian
None[1]	84.0%	74.7%
Congruent	93.8%	87.5%
Incongruent	51.4%	45.8%

As Table 5.2 shows, congruent context produced facial expression recognition rates that were nearly twice that of incongruent context, regardless of culture. Moreover, incongruent context significantly reduced recognition rates over providing no context at all. Meanwhile, congruent context increased facial expression recognition rates by about 10-12% over no context, which replicates previous findings from Chapter 4 (Bennett & Šabanović, 2015). These patterns occurred regardless of the cultural background of the subjects.

The patterns were investigated for significance via a two-way, *within-subjects* ANOVA (see Section 5.2.4). The results are shown in Table 5.3. Significant effects on accuracy were found for context congruency ($p<.001$) but not for cultural background. The interaction effect was not significant.

Table 5.3: Context Congruency and Culture - ANOVA

	F	Sign.
Culture	0.639	0.428
Congruency	46.38	0.000*
Culture * Congruency	0.02	0.883

**Significant values ($p<.05$) marked with an asterisk

A breakdown of the recognition rates for each facial expression by cultural group is provided in Table 5.4. There are some interesting patterns of note, although caution is warranted when sub-dividing the experimental data to that degree. In general, recognition rates for all expressions were reduced when

context was incongruent, except for Surprise (curiously). Surprise recognition actually increased slightly. This was true regardless of whether Surprise was switched with expressions of similar (positive) or opposite (negative) valence. The reason for this is unclear. The result warrants further experimental investigation.

Table 5.4: Individual Facial Expressions and Context Congruency

Context	Expression	Western	East Asian
Congruent	Happy	100.0%	100.0%
	Sad	100.0%	87.5%
	Anger	100.0%	87.5%
	Fear	93.8%	87.5%
	Surprise	75.0%	75.0%
Incongruent	Happy	50.0%	43.8%
	Sad	37.5%	37.5%
	Anger	50.0%	25.0%
	Fear	0.0%	25.0%
	Surprise	93.8%	87.5%

In summary, context congruency had a significant influence in human perceptions of robotic facial expressions. Providing incongruent context was worse than providing no context at all. There were no significant differences in context effects due to culture.

5.3.2 Effects of Emotional Valence Similarity between Facial Expression and Context

A second primary question was whether the effects of context congruency depended on the similarity of emotional valence of the facial expressions and the context. In order to test this, we had three experimental conditions that varied by emotional valence similarity (see Section 5.2.3). A summary of facial expression recognition rates by condition and culture is shown in Table 5.5.

Table 5.5: Recognition Accuracy by Condition and Culture

Condition	Context	Western	East Asian
Positive Switch	Congruent	100.0%	87.5%
	Incongruent	75.0%	62.5%
Negative Switch	Congruent	87.5%	87.5%
	Incongruent	16.7%	37.5%
Cross Switch	Congruent	87.5%	87.5%
	Incongruent	65.6%	46.9%

As can be seen in the table, congruent context produced fairly stable recognition rates across conditions and culture. However, incongruent context produced recognition rates that varied significantly depending on the condition. The pattern (positive<cross<negative) was the same for both cultural groups, although the specific values differed. In short, perceptions of negative-valence emotional expressions (Sad, Fear, Anger) were more heavily affected by context congruency than their positive-valence counterparts. Switching across valence fell somewhere in between.

The patterns were investigated for significance via a two-way, between-subjects ANOVA (see Section 5.2.4). The results are shown in Table 5.6. Significant effects were found for condition ($p<.001$) – i.e. emotional valence similarity – but not for cultural background. The interaction effect was not significant. Post-hoc Bonferroni t-tests revealed the significant differences were between the negative switch condition (Condition 2) and the other two.

Table 5.6: Condition by Culture - ANOVA

	Main Accuracy	
	F	**Sign.**
Culture	0.503	0.482
Condition	9.577	0.000*
Culture * Condition	1.94	0.157

**Significant values ($p<.05$) marked with an asterisk

In summary, the similarity between emotional valence of the facial expressions and the context had a significant effect on in human perceptions of robotic facial expressions. In particular, the effects were significantly larger for facial expressions of negative-valence emotions. These results suggest that there may be predictable patterns in the effects of congruent/incongruent environmental context on perceptions of robot affect, regardless of culture.

5.4 Discussion

We performed an empirical study (n=48) investigating the effects of context congruency on perceptions of robotic facial expressions across cultures. There were two key findings. First, context congruency had a significant effect on human perceptions of robotic facial expressions. This effect occurred regardless of culture, and was even of similar size. Providing incongruent context was worse than providing no context at all. Second, the similarity of emotional valence between the context and the facial expressions played a significant role, whereas negative-valence emotions were more affected by context congruency. Again, this effect occurred regardless of culture.

The results suggest that there may be predictable patterns in the effects of environmental context on perceptions of robot affect, regardless of culture. These patterns are shaped by the congruence/incongruence of the context, as well as its emotional valence. As has been suggested previously, these patterns fit the notion that emotions which humans perceive in others' faces – including robots – may be an internal construct in the mind of the perceiver, based on a number of perceptual and cognitive processes (Bennett & Šabanović, 2015; Barrett, Mesquita, & Gendron, 2011, Lindquist & Gendron, 2013). In other words, humans appear to be *projecting* their own internal emotions. More broadly, this fits into recent cognitive science research on social cognition as an emergent phenomenon (Barsalou, Breazeal, & Smith, 2007; De Jaegher & Di Paolo, 2007, De Jaegher, Di Paolo, & Gallagher, 2010). We discuss this idea in more detail in Section 4.4.2.

From a robot design standpoint, understanding these sorts of phenomena holds great potential to

enhance socially interactive robots and human-robot interaction. Inducement of certain external context effects (see Section 5.1.1) may allow us to shape the interaction without necessarily redesigning the robot itself. Moreover, given the predictable patterns of context effects, such an approach may allow us to produce *culture-neutral models* of robots and affective interaction, as argued in Chapter 4 (Bennett & Šabanović, 2015). In other words, taking advantage of context in the dynamical process of perception formation may ease the constraints for developing culturally-specific affective cues in human-robot interaction. The goal is still to design robots in culturally relevant ways, but such an approach allows us to do so in a more flexible manner (Šabanović, Bennett, & Lee, 2014; Rehm et al., 2007). In short, it may not make sense to design robots *in toto* for specific cultures (especially since culture itself is dynamic and constantly in flux), but rather to design robots that are sensitive and adaptive to particular cultural factors.

Chapter 6

A Month in the Museum: Interaction Patterns with a Robotic Face in the Wild

The previous several chapters (Chapters 3-5) have focused on the development of an empirically-grounded face, and the effects that culture and/or context might have on human perceptions thereof, through lab-based experiment. In this chapter, we explore how people interact with such a robotic face in naturalistic "in-the-wild" settings.

Abstract. We report on a long-term human-robot interaction study (spanning three weeks: 8 hours a day, 6 days a week) in a public setting (an art museum exhibit open to the general public) using an autonomous, interactive robotic face. Researchers were not present on-site. Over 700 people interacted with the robot across 300 interactions, both alone and in groups, in a free-form manner with minimal instruction. Video recordings were analyzed to see whether people exhibit common interaction patterns towards a robotic face in naturalistic settings, which could inform the development of data-driven models of robot social behavior. Clustering revealed four well-defined "interaction schemas" from the interaction behavior data. These results suggest that people often adopt specific interaction schemas when interacting with a robotic face in a free-form, naturalistic setting (outside the lab), schemas identifiable from the interaction data itself. We discuss design implications of these findings for future robot interactive behavior.

6.1 Introduction

The previous several chapters (Chapters 3-5) have focused on the development of an empirically-grounded face, and the effects that culture and/or context might have on human perceptions thereof. Those experiments, of course, were all lab-based experiments. An immediate and obvious question thus springs to mind – *what would happen if we put an interactive robotic face in a naturalistic, "in-the-wild" setting, and allowed people to interact with it in undirected, free-form ways?*

In this chapter (Chapter 6) and the next (Chapter 7), we do just that, taking another iteration of the robotic face MiRAE (described in Chapter 2) – equipped with 3D printed components, basic social interactions capabilities, and the ability to see/respond to human interactors in its environment – and putting it into a public art installation in a museum gallery. In this chapter, we focus on the analysis of human interaction patterns, attempting to identify common behavioral patterns that might inform design of future robotic behavior. In Chapter 7, we compare these naturalistic patterns from the museum with those from the lab – i.e. *is what people do in human-robot-interaction lab experiments different from what they do in natural settings*?

6.1.1 Problem

Along with a growing focus on implementing robots outside controlled environments like labs and factories, there is an increasing need to understand how human-robot interaction (HRI) occurs in naturalistic settings, where robots interact with people who have not been trained and are not guided by researchers. Studies of HRI "in the wild" allow people to approach robots voluntarily, interact in free-form ways, and ignore, explore, and address robots on their own terms. Initial research of this kind, seeking to understand how people behave towards and make sense of robots, has shown that contextual factors (such as the social and physical environment) affect the flow and success of interactions (Šabanović, Michalowski, & Simmons, 2006; Mutlu & Forlizzi, 2008). Along with naturalistic HRI studies, researchers have also started identifying interaction patterns between people and robots both in and outside the lab as a foundation for developing models of appropriate social interaction cues and

capabilities for robots (Mutlu & Forlizzi, 2008; Kahn et al, 2008). Such studies outside the lab, in particular, provide a larger degree of ecological validity and can inform robot design for real-world applications (Šabanović, Reeder, & Kechavarzi, 2014).

In this chapter, we explore the interaction patterns that emerge "in the wild" between a socially interactive robotic face and visitors to an art gallery. One of our aims is to understand how people spontaneously and voluntarily respond to a robotic face in a naturalistic setting: Are there shared models they use when they approach the robot? Are there repeating behavioral interaction patterns between the person and the robot that can inform our design of future robots and their behavior? To answer such questions, we performed clustering analysis of behavioral data collected during naturalistic human-robot interactions to extract and identify common behavioral patterns emergent from the data itself (without pre-specified class labels), and discuss how these might be integrated into future robot design.

We based our identification of interaction patterns on a long-term human-robot interaction study, which was conducted in a public setting using a fully autonomous, interactive robotic face. The robot was capable of basic social interaction: it could detect faces and motion, follow them as they moved, and react via facial expressions to external stimuli. The setting was an art museum exhibit, open to the general public, in which the robot was operational for three weeks, six days a week, eight hours a day. Researchers were not present on-site, aside from turning the robot on and off at the beginning and end of each day. Over 700 people interacted with the robot across 300 interactions over the duration of the study, both alone and in groups, in a free-form manner with little to no instruction. Interaction data was collected from multiple modalities, including both onboard and offboard video data of the human interactors as well as internal proprioceptive data of the robot's own behaviors over time. The goal was two-pronged: 1) to study how people might respond to and make sense of an interactive robotic face in naturalistic settings (i.e. do common patterns exist?), 2) to analyze how emergent interaction patterns might be used to develop data-driven models of interactive robot behavior and social robots (i.e. what are the design implications?).

6.1.2 Background and Previous Work

This chapter contributes to the growing domain of HRI studies "in the wild," particularly those done in naturalistic, public settings in which diverse users have the opportunity to interact with robots in a voluntary and open-ended fashion. Such studies have been performed in museums (e.g. Nourbakhsh, Kunz, & Willeke, 2003; Yamazaki et al, 2009), malls (e.g. Kanda et al., 2009), university campuses (e.g. Gockley et al., 2005), city streets (e.g. Weiss et al. 2010), schools (e.g. Tanaka, Cicourel, & Movellan, 2007; Leite et al., 2012), and public areas of caregiving institutions (e.g. Chang & Sabanovic, 2014). The presented work also adds to the literature on "interaction patterns" in human-robot interaction, which involves the identification and description of repeating general patterns of interactive behavior between humans and robots that can be realized in a recognizable though unique manner in different contexts (Kahn et al., 2010a).

6.1.2.1 Studying Robots in Public Spaces

The majority of HRI studies in public spaces have focused on testing the social acceptability of robots in various environments, describing user reactions to robots in public spaces, and identifying design characteristics that support social acceptance and continued use. Straub et al. (2010) studied how people interacted with a Geminoid HI-1 android at a public café in both autonomous and telepresence modes, and found that participants ascribed humanistic traits to the robot independent of the operation mode. Ruiz-del-Solar et al. (2009) evaluated Bender, a robot that could speak and express anger, sadness, and happiness through facial expressions, in three different settings (a home, a school classroom, and a university building) and showed that people could generally understand the robot's communicative attempts and were overall accepting of the robot. Studies with the "Roboceptionist," a robot operating at the entrance to Carnegie Mellon University's Robotics Institute since 2003, further showed that various interaction cues, including emotional responses and personalization, can affect people's willingness to voluntarily engage and maintain interaction with a robot (Gockley et al., 2005). Evaluations of the ACE (autonomous city explorer) robot displayed that people are willing to communicate with and help a robot

dependent on the aid of passerby to plan a route to its final destination, suggesting that a social mode of navigation is viable (Weiss et al., 2010). Mutlu and Forlizzi's (2008) ethnographic study of robot use in a hospital found that acceptance depended not on the robot's characteristics alone, but on their relative fit into the social dynamics of the work environment.

There is a long history of using robots in museums, similarly to our own, as a way to study human-robot interaction in naturalistic settings and inform robot design. Museums can be seen as particularly advantageous settings for exploratory studies of HRI – people are there to learn and experience new things, so they may be more open to novel experiences with robots. Thrun et al.'s (1999) work with the museum guide robot Minerva acknowledged the importance of interactive capabilities as well as mobility and navigation for this application area. The five-year Mobot museum robot experience described by Nourbakhsh, Kunz, & Willeke (2003) yielded a series of requirements for successful human-robot interaction in a museum setting, including the importance of the physical appearance, movement, and social awareness of the robot as an enticement to interaction. They further identified multimodal interaction design, interactive tasks, and the need for the robot to follow human social norms (including giving negative reactions to behavior that is making it difficult for the robot to perform its job, such as crowding the robot) as ways to retain visitor attention. In developing a personal rover exhibit, Nourbakhsh et al (2005) showed the importance of reliability (failing rarely despite daily use and being easy to fix), autonomy (performing without staff intervention), and a self-explanatory user interface which allows people to interact with the robot without prior training or the need for explanation. Studies with robots in museums also often have some educational or informative purpose, so the transmission of information or meeting specific learning goals are important outcomes, aside from the success of the interaction.

6.1.2.3 Developing Interaction Patterns for Robots

Studies of robots in public spaces have also been used to develop and evaluate particular models of behavior for successful HRI. An ongoing project using the Robovie platform seeks to develop HRI

capabilities for day-to-day interactions with diverse users. Observations of interactions with customers, as well as multiple studies of user acceptance, indicate that robots were able to influence the shopping habits of customers and were evaluated positively (Kanda et al., 2009). In the museum context, Yamazaki et al (2009; 2012) used ethnographic fieldwork in museums to understand how tour guides perform their job and use the resulting behavioral models in the interaction design of robot guides, which they also test out in natural settings. These studies suggest that a particular model of robotic development, which uses observation of human-human behavior in naturalistic environments to develop models for robot behavior and evaluates those in HRI "in the wild," can be useful for constructing robots for use in everyday settings.

Although HRI research commonly uses the Computers-As-Social-Actors (CASA) framework as a rationale for building interactive capabilities for robots by replicating those of humans, people do not always treat robots exactly like humans (Reeves & Nass 1996). Kahn et al. (2011) suggest robots inhabit a category between humans and machines in terms of the models that people use for interpreting and responding to their behavior. Our approach in this chapter has therefore been to explore how people behave towards an interactive robot in order to understand their initial reactions and develop further interaction capabilities for the robot. We are in particular interested in identifying repeated "interaction patterns" – "the glances, positionings, gestures" – and "sequences of behavior" that constitute face-to-face interaction (Kendon, 1990) between humans and robots as a foundation for future robot design. Such repeated behavioral sequences have been previously conceptualized as "design" or "interaction" patterns in HRI (a et al., 2008), which characterize "essential features of social interaction between humans and robots, specified abstractly enough such that many different instantiations of the pattern can be uniquely realized given different types of robots, purposes, and contexts of use" (Kahn et al., 2010a). Kahn et al have identified and used a variety of "interaction pattern" sequences, such as "Introduction" and "Walking together," in their research (e.g. Kahn et al, 2012). They also suggested a framework for validating the existence of interaction patterns in HRI (Kahn et al., 2010b), which involves establishing the effectiveness of the patterns in facilitating HRI, the ability of the pattern to account for the data, and

establishing a sensible reason for naming the pattern. Peltason and Wrede (2010a) used interaction patterns extracted from a variety of human-robot interaction scenarios to assist in the development of algorithms for robot dialogue for real-world applications. This allowed them to combine abstract task states (such as task accepted or failed) with generalized dialogue acts (such as an apology), which could be adapted to different applications and situations. Finally, common conceptual interpretations (or schemas) of robots along with interaction patterns have been identified in qualitative studies of initial and continuing interactions between people and robots (Turkle, 2006, 2011).

Our work takes advantage of the museum context as a space in which people are able to create novel interactions without research influence. We explore the resulting behavioral patterns that emerge from such open-ended initial interactions between people and robots in order to identify common "interaction schemas" that a robot might need to recognize and participate in during initial interactions with people, and to inspire the development of appropriate responses that the robot could produce to successfully elicit further continued interaction.

6.2 Methods

6.2.1 Robotic Face

The platform used in this study (MiRAE) is a minimalist robotic face shown in Figure 2.1, previously described in Chapter 2 (Bennett & Šabanović, 2014; Bennett et al., 2014; Bennett & Šabanović, 2015). It is capable of basic, non-verbal, infant-like social interaction behavior. It can detect faces and motion, respond to people, and make a variety of facial expressions (e.g. frown, smile). It has the ability to track environmental stimuli both relative to its sensory (retinotopic) and motor (spatiotopic) coordinates, and follow them using a neck mechanism. It also has a basic visual attention and affective system.

MiRAE's motor, visual, and cognitive functions are written as C++ and Python libraries, and available as open-source software online (http://www.caseybennett.com/research.html) or at the author's lab website (http://r-house.soic.indiana.edu/projects/mirae.html). Most of the computer vision aspects are

based on OpenCV (opencv.org), including use of Haar Cascade algorithms for face detection and optical flow/intensity gradient for motion detection.

MiRAE's physical construction is designed to be replicable, using easily accessible components (e.g. Arduino microcontrollers) and 3D printed facial components. It has 12 degrees-of-freedom, including 2 for its neck pan-and-tilt motion. It is also equipped with an onboard camera for computer vision purposes. Full construction details are available online (http://r-house.soic.indiana.edu/mirae/MiRAE_Construction_Manual.pdf).

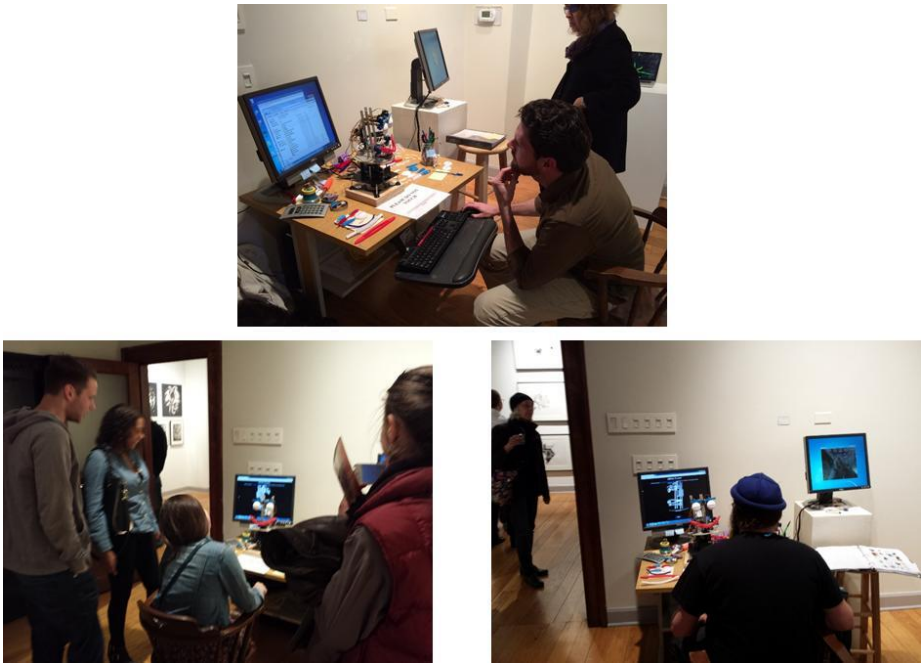
MiRAE has been experimentally validated previously in terms of affective interaction in (Bennett & Šabanović, 2014; Bennett et al., 2014; Bennett & Šabanović, 2015), showing that it is capable of producing higher, or at least comparable, identification accuracy rates for the Ekman facial expressions (Happiness, Sadness, Anger, Fear, Surprise) as a number of other robotic faces, including Kismet (Breazeal, 2003), Eddie (Sosnowski et al., 2006), and the android Geminoid-F (Becker-Asano & Ishiguro, 2011).

6.2.2 Study Setting

The study took place in an art gallery set in the downtown of a Midwestern United States city. It was open to the public. The exhibit was part of a larger art installation on computer-aided design, 3D printing, and the meshing of the digital world with our physical one. One aim of the exhibit was to educate the public about the possibilities of these novel technologies. The robot face was exhibited on a desk arrayed with various 3-D printed materials, electronic components, and tools used in the building/creation of robots (for educational purposes). These were taken from our lab and related to the project, and presented with the intent of giving visitors some idea of how the robot was produced. Visitors could also peruse a lab notebook placed beside the robot desk, which included notes and photos with inspirations for the robot's design, sketches of design iterations, details related to the robot's construction, etc. Other exhibits that shared the gallery with the robot included 3-D printed art pieces and interactive installations that responded to visitors with light and sound. People came to the gallery to see

the exhibit itself, as well as for regularly scheduled plays and art classes.

Figure 6.1: Art Museum Exhibit Setting



In the gallery, people could approach the robot as they liked, either individually or in groups, and voluntarily interact with it (as shown in Figure 6.1). The interaction was freeform and undirected – researchers were not present on-site except to turn the robot on and off at the beginning and end of each day. Minimal instruction for interacting was provided as part of the exhibit itself, consisting of a post-it note placed in front of the robot that said "Come and chat with me" and noting that the robot was capable of infant-like behavior (e.g. it could see people, detect faces, respond). Visitors were asked not to touch the robot, to decrease the possibility that the exhibit would be damaged. People were otherwise free to interact with the robotic face on their own terms, for as long as they liked, and to interpret the robot's responsive behavior however they wished. Over the course of the three-week-long exhibit, over 700 people interacted with the robot face, which functioned for a total of 143.5 hours.

6.2.3 Study Design

The study was designed as a naturalistic interaction, in which the robot (capable of autonomous interactive behavior) was placed in a public setting (see Section 6.2.2). As can be seen in Figure 6.1, the robot was placed on a desk with a chair in front of it, so that a seated person would be at eye level with the robot. A computer screen was also present displaying 3D models of the robot design (for educational purposes). Two cameras were used in the exhibit, one onboard the robot itself and another stationary offboard camera positioned in the back corner of the desk. Video from both cameras for every interaction was recorded and time-stamped for later analysis (as detailed below).

The robot was turned on/off at the beginning and end of each day, and ran for approximately 8-10 hours per day, 6 days a week (the gallery was closed on Sundays). After turning the robot on, the researcher left the building, and returned at the end of the day to check on the exhibit and turn off the robot.

The number of interactions per day varied between 21 and 5, and was roughly 12 on average. Average duration of each interaction was 65.3 seconds. Two-thirds of the interactions were group interactions with multiple people at once (group size average was 3.3 persons). The exception to the above interaction averages was the exhibit's "opening night", in which the robot was only turned on for a few hours but experienced a much higher amount and frequency of people passing by, resulting in 118 interactions. Given this difference, we separate out these two exhibit phases in much of the subsequent analysis (see Section 6.3 for details).

For the purpose of this study, we focused on making the robot's interactive behaviors reliable, autonomous, and easily understandable for participants, in accordance with Nourbakhsh et al. (2005) suggestions for museum robot installations. The autonomous interaction behavior of the robot in this study can be summed up as follows (a link to the programming code is provided in Section 6.2.1). The robot, if it detected any sort of motion in its visual field, would immediately begin tracking/following that motion (if multiple points of motion were detected, it would choose the point of largest motion). If the robot detected specific stimuli (e.g. a person or face), it would preferentially attend to that (over detected

motion). If multiple stimuli were detected, it would first attend to the closest one, then shift its attention to ones further away, based on simple attentional decay (similar to Rubi [Movellan et al., 2007]). If no motion or stimuli were detected, the robot would make random saccade motions. The affective system was also operating simultaneously. If novel stimuli were detected, the robot would express surprise or interest. Examples of novel stimuli might include a face, given that a face had not been detected recently, or stimuli suddenly increasing in size as if rapidly moving toward the robot. Otherwise, in the presence of positive stimuli (e.g. a person or face) that appeared to be interacting with the robot, the robot would express happiness (smile). If during the interaction, the person/face suddenly departed or moved out of view, the robot would express sadness (frown). For instance, a person could play peekaboo with the robot face by covering their face and then uncovering it. Finally, if a person completely stopped moving/interacting for a period of 7-8 seconds, the robot would become "bored" and begin searching for new stimuli.

Data collected for later analysis during the experiment included video from both the onboard and offboard video cameras. Motion data (optical flow and intensity gradients) sampled from a 5x5 grid across the visual field were also recorded as numerical data, at every timepoint (approximately twice per second). Internal proprioceptive data was also recorded from the robot, which included positions of its motors, internal affective states, locations of each detected stimuli (if any), and information about the current attended stimuli (if any) at every timepoint. All the data was time-stamped, so that, for instance, internal proprioceptive data could be matched later to video data, etc.

6.2.4 Analysis Approach

Due to space limitations, this chapter focuses on analysis of the offboard video data, with the aim of understanding in detail the kinds of interaction behaviors people displayed toward the robotic face. 315 total interactions occurred (over 500 total people, some who came in groups). Of those, 182 interactions (comprising 256 individuals) were included in the final analysis, about 60% of the total. The other ~40% of interactions were considered "non-interactions", as the people only looked at the robot in

those cases, but did not attempt to interact with it. The overarching question is: what do people actually do when confronted with an interactive robotic face in naturalistic settings, can any useful interaction patterns be extracted from this data, and what might that tell us about how to design the robot and/or its interactive behavior?

Video was analyzed in two phases. First, an initial pass through the full set of videos was made, identifying each interaction that took place (whether individual people or groups of people) and marking the begin and end time of each interaction. Each interaction was also categorized by a number of variables related to the context and characteristics of the interaction and its participants: Exhibit Phase (opening night vs. "regular" daily exhibit), Group Size (alone or group), Age (adult or child), Gender (male or female), Interaction Zone (proximal or distal), and Primary (yes or no). Group Size, Age, and Gender are self-explanatory, and Exhibit Phase is explained in Section 6.2.3. As for the others, Interaction Zone refers to how close the person was to the robot, with "proximal" defined as being in the chair or within a few feet of the robot (the same distance as if they were in the chair). Primary distinguished whether the person was the main interactor during the interaction – generally this was the closest person to the robot who performed most of the interaction. This distinction was developed with the understanding that when another person is in front of and interacting with the robot, it limits what another participant can do, so we wanted to separate those conditions. Interactions generally had only one primary interactor, although in a few cases there were more than one or people took turns.

A second pass was then made on the video for each interaction (excluding non-interactions, see above), which entailed annotated coding of each video clip using Anvil (http://www.anvil-software.org/) to produce coded interaction behavior data (totaling roughly 235 minutes). Video coding was performed by two independent coders, with partially overlapping sets (approximately 10% of the total data), in order to calculate inter-rater reliability. Inter-class correlation was calculated via SPSS, with a Cronbach's Alpha of 0.734, which is considered good. Deciding which behaviors to code was based on the most commonly observed behaviors during the first pass of video analysis. The final list of behavioral codes consisted of: Smiling, Frowning, Other Facial Expressions, Making Exaggerated Faces, Sticking Tongue

Out, Talking, Laughing, Mimic Robot (attempting to get the robot to mimic them, e.g. making the opposite expression of the robot), Communicative Hand Gestures (e.g. waving hello or goodbye), Attentional Hand Gestures (etc. snapping, pointing, finger-wagging), Feeding the Robot, Inspecting the Robot (examining the robot's structure rather than making direct eye contact with it or interacting with it socially). Other than the facial expressions (a person could not smile and frown at the same time), these coded behaviors were not mutually exclusive, i.e. a person could smile and talk at the same time.

Behaviors were coded for both occurrences and time spent (in seconds), allowing us to analyze variations in both the number of people and the time spent per person engaging in each behavior across variables. For analysis purposes, time spent was scaled by the total duration of the interaction, since different people interacted for different lengths of time. In essence, this converted the time spent into a unit-free "time spent per second" value (i.e. what percentage of each second did the person spend doing behavior 'x'), independent of the actual duration. This allowed us to directly compare different interactions.

We also transcribed and coded participant dialogue from the videos. The coding scheme included identification of Direct speech (e.g. "Hi, how are you doing?"), Anthropomorphism of the robot (e.g. "He's so sad," "He looked at me"), Mechanistic interpretation of the robot (e.g. "It has a camera"), ascription of Childlike behavior (e.g. "It's like a baby"), and focus on what the robot is attending to (e.g. "Does it see me?"). The codes were not mutually exclusive. 161 (out of the total of 256) individuals had utterances recorded.

The analysis presented in Section 6.3 comes in two parts. First, we present an analysis of the variables of each interaction to examine differences across age, gender, etc. in interaction behaviors. We chose to test these demographic variables in this exploratory research because they are commonly evaluated in human-robot interaction studies as sources of differentiation in social behavior and attitudes toward robots (e.g. Schermerhorn, Scheutz, Crowell, 2008; Ezer, Fisk, & Rogers, 2009). Statistical hypothesis testing was performed using independent samples t-test in SPSS. We adopt the null hypothesis as our starting point, i.e. there would be no differences in interaction behavior across variables.

Second, we analyzed the coded interaction behavior data using unsupervised clustering to examine whether different groups of people adopted identifiable "interaction schemas" while interacting the robot, and whether such interaction schemas were emergent in the data. We utilized two-step clustering in SPSS to identify such clusters (http://10.110.22.85:49801/help/topic/com.ibm.spss.statistics.algorithms/alg_2step_cluster.htm), and assigned each interacting person to one of these clusters. Differences in interaction behaviors between clusters were evaluated via ANOVA. Clusters were then analyzed for differences across variables (age, gender, group size, etc.) using Chi-squared tests in SPSS.

6.3 Results

6.3.1 Variable Analysis

Each interaction was defined by six variables: Exhibit Phase, Group Size, Age, Gender, Interaction Zone, and Primary (described in Section 6.2.4). Descriptive statistics for each of those six are shown in Table 6.1.

Table 6.1: Variable Descriptives

Exhibit Phase	N (%)	Int. Zone	N (%)	GroupSize	N (%)
Opening	93 (36.3%)	Proximal	130 (50.8%)	Group	167 (65.2)
Post	163 (63.7%)	Distal	126 (49.2%)	Alone	89 (34.8)
Gender	**N (%)**	**Age**	**N (%)**	**Primary**	**N (%)**
Female	145 (56.6)	Adult	201 (78.5%)	Yes	177 (69.1)
Male	111 (43.3)	Child	55 (21.5%)	No	79 (30.9)

Variables were compared across twelve interaction behaviors (described in Section 6.2.4), both for people count (i.e. the number of people engaging in each behavior) and time spent (scaled per second, see Section 6.2.4). Due to potential differences in the opening and regular exhibit phases, we chose to only include the "regular" phase for variables other than Exhibit Phase in further analyses. Percentage values for both are shown in Tables 6.2 and 6.3.

Table 6.2. Variable Comparison – People Count

		Facial Expressions								Hand Gestures			
Variable	Comparison	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Communicative	Attentional	Feeding	Inspecting
Exhibit Phase	Opening	65.6%	6.5%	10.8%	8.6%	4.3%	68.8%	39.8%	5.4%	30.1%	22.6%	0.0%	18.3%
	Regular	65.6%	6.7%	13.5%	18.4%	4.3%	80.4%	39.9%	5.5%	41.7%	23.3%	4.3%	17.8%
Int. Zone	Distal	56.4%	1.3%	10.3%	12.8%	1.3%	74.4%	29.5%	3.8%	34.6%	17.9%	3.8%	25.6%
	Proximal	74.1%	11.8%	16.5%	23.5%	7.1%	85.9%	49.4%	7.1%	48.2%	28.2%	4.7%	10.6%
Group Size	Alone	61.1%	11.1%	22.2%	18.5%	9.3%	63.0%	29.6%	11.1%	37.0%	27.8%	7.4%	18.5%
	Group	67.9%	4.6%	9.2%	18.3%	1.8%	89.0%	45.0%	2.8%	44.0%	21.1%	2.8%	17.4%
Age	Adult	70.2%	5.8%	13.2%	16.5%	2.5%	81.8%	47.9%	5.8%	41.3%	24.0%	5.8%	18.2%
	Child	52.4%	9.5%	14.3%	23.8%	9.5%	76.2%	16.7%	4.8%	42.9%	21.4%	0.0%	19.0%
Gender	Female	69.7%	6.7%	14.6%	22.5%	4.5%	84.3%	46.1%	4.5%	44.9%	22.5%	6.7%	11.2%
	Male	60.8%	6.8%	12.2%	13.5%	4.1%	75.7%	32.4%	6.8%	37.8%	24.3%	1.4%	27.0%
Primary	No	60.8%	2.0%	3.9%	7.8%	0.0%	84.3%	47.1%	0.0%	35.3%	13.7%	2.0%	11.8%
	Yes	67.9%	8.9%	17.9%	23.2%	6.3%	78.6%	36.6%	8.0%	44.6%	27.7%	5.4%	20.5%

Table 6.3. Variable Comparison – Time Spent

		Facial Expressions								Hand Gestures			
Variable	Comparison	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Communicative	Attentional	Feeding	Inspecting
Exhibit Phase	Opening	23.64%	0.95%	0.74%	1.35%	0.52%	12.33%	3.11%	1.15%	3.45%	2.63%	0.00%	3.18%
	Regular	14.52%	0.61%	0.89%	1.70%	0.34%	16.63%	2.12%	0.88%	4.52%	3.40%	0.97%	2.77%
Int. Zone	Distal	16.73%	0.06%	0.96%	1.27%	0.05%	12.55%	2.17%	0.72%	5.40%	2.69%	1.60%	5.01%
	Proximal	12.49%	1.11%	0.82%	2.09%	0.60%	20.38%	2.08%	1.03%	3.71%	4.04%	0.40%	0.72%
Group Size	Alone	13.63%	1.38%	1.62%	1.55%	0.64%	11.23%	2.14%	1.96%	5.88%	3.24%	2.41%	3.81%
	Group	14.96%	0.22%	0.53%	1.77%	0.19%	19.31%	2.11%	0.35%	3.84%	3.47%	0.26%	2.26%
Age	Adult	16.71%	0.45%	0.73%	0.97%	0.17%	16.98%	2.68%	0.78%	4.58%	3.54%	1.31%	2.60%
	Child	8.20%	1.07%	1.34%	3.79%	0.82%	15.63%	0.53%	1.18%	4.34%	2.98%	0.00%	3.27%
Gender	Female	15.01%	0.50%	1.06%	1.95%	0.42%	18.17%	2.80%	0.54%	4.68%	2.18%	1.48%	1.31%
	Male	13.92%	0.73%	0.68%	1.39%	0.24%	14.78%	1.31%	1.29%	4.32%	4.85%	0.36%	4.53%
Primary	No	14.55%	0.04%	0.09%	0.28%	0.00%	16.08%	2.87%	0.00%	2.23%	4.41%	0.07%	1.64%
	Yes	14.50%	0.87%	1.25%	2.34%	0.49%	16.89%	1.78%	1.28%	5.56%	2.93%	1.38%	3.29%

The raw data in Tables 6.2 and 6.3 were evaluated for statistical significance via an independent samples t-test. Significant differences are shown in Table 6.4, with p values in parentheses.

Table 6.4. Variable Comparison – Significant Differences

		Facial Expressions								Hand Gestures			
Variable	Comparison	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Communicative	Attentional	Feeding	Inspecting
Exhibit Phase[1]	People Count				2.32 (.021)		2.01 (.046)					2.70 (.008)	
	Time Spent	2.95 (.004)										2.09 (.039)	
Int. Zone	People Count	2.85 (.005)	2.84 (.005)					2.91 (.004)					3.31 (.001)
	Time Spent		2.22 (.029)			2.11 (.038)							2.56 (.012)
Group Size	People Count			2.05 (.043)			2.00 (.047)						
	Time Spent			2.04 (.045)			3.57 (.001)						
Age	People Count	2.02 (.047)						4.23 (.001)				2.71 (.008)	
	Time Spent	3.50 (.001)			2.25 (.029)			3.54 (.001)				2.09 (.038)	
Gender	People Count												2.55 (.012)
	Time Spent												2.41 (.018)
Primary	People Count		2.09 (.039)	3.06 (.003)	2.78 (.006)	2.72 (.008)			3.11 (.002)		2.16 (.033)		
	Time Spent		2.30 (.023)	3.78 (.000)	3.67 (.000)	2.46 (.015)			2.31 (.023)	2.68 (.008)			

[1]All variable comparisons other than Exhibit Phase only included the Regular phase

As can be in Table 6.4, there were a number of differences in behavior due to Primary, Age, and Interaction Zone. Primary interactors were more likely than non-primary interractors to engage in diverse behaviors towards the robot, while adults were more likely to engage in smiling, laughing, and feeding behaviors than children (some of which may be due to the adoption of a caregiver interaction schema by adults, see Section 6.3.2). In regards to interaction zone, proximal interactors were more likely to engage in smiling, frowning, and laughing, which is consistent with behaviors that might be expected in close range face-to-face communication, while distal interactors were more likely to inspect the robot. There were also a few (though less) differences due to Exhibit Phase and Group Size, e.g. people in groups were more likely to talk to the robot. Finally, there were almost no differences in interaction behavior due to Gender, aside from male participants being more likely to inspect the robot than female participants. We have noticed a similar gender differentiation trend in human-robot interaction in another study of HRI in a public space performed with an assistive robot in a nursing home (Chang & Sabanovic, In Press).

6.3.2 Clustering Analysis

Clustering was performed on the interaction behavior data, to see if clusters naturally emerged from patterns in the data, i.e. whether people adopted identifiable "interaction schemas" while interacting with the robot. Results can be seen in Table 6.5, with percentages representing the scaled "time per second" spent performing each behavior, on average.

Four clusters were found in the data. The same clusters were found independently twice, both with and without the opening Exhibit Phase data. The silhouette score on the clusters was 0.65, which is considered good. Upon considering the behaviors identified within each cluster, we named them: 1) Non-Primary, 2) Low-Intensity/Non-Specific, 3) Caregiver Interaction, and 4) Mimic-Play. ANOVA tests were used to identify the principle defining behaviors (i.e. statistically significant) for each cluster, which are highlighted. The first cluster (Non-Primary) was primarily comprised of all the non-primary interactors (see Table 6.6 below). The third cluster (Caregiver Interaction) appears to be comprised of

people who seemed to adopt a nurturing interaction pattern, smiling, laughing, and talking much more frequently, even attempting to feed the robot, and using attentional hand gestures to engage it. In short, their behaviors seemed to resemble how a caregiver might attempt to interact with an infant. The fourth cluster (Mimic-Play) was principally defined by individuals who attempted to engage the robot via diverse facial expressions, often expressing opposing expressions to those displayed by the robot (e.g. if the robot smiled, they would immediately frown). For instance, the Caregiver cluster spent about 40% of their time smiling at the robot, while those in the Mimic-Play cluster also spent nearly 40% of the time making facial expressions, but their expressions were more diverse and/or exaggerated. The final cluster (Low-Intensity/Non-Specific) was in essence comprised of people who did not appear to adopt a dominant interaction schema (i.e. either caregiver or mimic-play or anything else), but rather did a little bit of everything. This was the largest cluster, representing approximately 50% of people who encountered the robot face in the exhibit.

Table 6.5. Cluster Analysis

		1	2	3	4	*F*-value	Sign.
	N (All Phases)	74	132	17	33		
	N (Post Only)	48	84	7	24		
Facial Expressions	**Smiling**	11.6%	12.2%	40.2%	20.8%	7.57	.000*
	Frowning	0.0%	0.1%	0.0%	3.9%	11.84	.000*
	Other	0.1%	0.0%	0.0%	5.8%	72.18	.000*
	Exaggerated	0.3%	2.8%	0.6%	1.0%	3.22	.024*
	Tongue Out	0.0%	0.6%	0.0%	0.3%	1.12	.342
	Talking	12.6%	16.2%	48.0%	17.0%	4.56	.004*
	Laughing	1.5%	1.1%	18.6%	2.0%	37.22	.000*
	Mimicking	0.0%	0.0%	0.0%	6.0%	12.29	.000*
Hand Gestures	**Communicative**	2.2%	6.8%	2.4%	1.8%	2.62	.053
	Attentional	1.7%	3.0%	20.3%	3.2%	4.47	.005*
	Feeding	0.1%	0.2%	17.8%	0.7%	30.04	.000*
	Inspecting	1.7%	3.5%	0.0%	3.2%	0.75	.522
	Cluster Label	Non-Primary	Low-Intenstity/ Non-Specific	Caregiver Interaction	Mimic-Play		

Clusters were also evaluated for differences in the variables via Chi-squared tests, shown in Table 6.6. The table does not include Exhibit Phase, as we found the same clusters in both phases.

Table 6.6. Cluster Variable Analysis

Variable	Value	1	2	3	4	X^2	Sign.
Int. Zone	Proximal	45.8%	52.4%	42.9%	62.5%	3.26	.354
Group Size	Group	100.0%	60.7%	42.9%	37.5%	30.78	.000*
Age	Adult	81.3%	67.9%	100.0%	75.0%	5.46	.141
Gender	Female	62.5%	50.0%	57.1%	54.2%	1.95	.548
Primary	Yes	0.0%	100.0%	71.4%	100.0%	293.50	.000*

As can be seen in Table 6.6, the main differences in clusters in terms of interaction variables were due to Primary and Group Size. More specifically, nearly all of the non-Primary interactors ended up in the first cluster. In terms of group size, we see a steady decrease from the first to the third and fourth clusters, so that the Caregiving and Mimic-Play clusters are predominantly adopted by people interacting with the robot on their own. It therefore appears that being alone significantly increases the chance a person will adopt a dominant interaction pattern (clusters 3 and 4). We also briefly note that dialogue analysis was also conducted for each cluster, which found a much higher incidence of child-based references in the Caregiver cluster, and a higher incidence of direct comments to the robot in the Mimic-play cluster – in both cases, about twice as frequent as the incidence of similar utterances in other clusters (results omitted for brevity). These discursive patterns suggest that behavioral patterns correspond to particular conceptual schemas that the participants may have had about the robot. In discussing the results below we will therefore refer to "interaction schemas" that emerge from the interaction.

6.4 Discussion

6.4.1 General Discussion

The purpose of this study was to explore the interaction patterns that emerge between a socially-interactive technology, a robotic face, and people in a public space, i.e. "in the wild" (Šabanović,

Michalowski, & Simmons, 2006). The goal was two-fold: 1) to study how people might naturally respond to an interactive robotic face in naturalistic settings, 2) to analyze interaction patterns that emerge in the course of such interaction that can be used to develop data-driven models and design implications of future robot social behavior. Results showed significant differences in interaction behavior across age, group size, interaction zone, and whether the person was the primary interactor, but not due to gender (Section 6.3.1). Furthermore, clustering analysis revealed four well-defined "interaction schema" clusters from the interaction behavior data itself, suggesting that people often adopt specific interaction schemas when interacting with a robotic face in a free-form, naturalistic setting (outside the lab). More critically, such schemas can be derived from the interaction data without pre-specified class labels (Section 6.3.2). This suggests that our interaction patterns are valid according to Kahn et al.'s (2010b) criteria of accounting for the data, while we chose their naming based on prior social and HRI research and therefore see it as sensible. Insofar as the interactions from which we drew these patterns were successful, the effectiveness of facilitating HRI using these patterns is also established and will be further tested in future work, in which the robot will be programmed to behave in accordance with the patterns. We discuss the design implications of these findings for guiding future robot social behavior below (Section 6.4.2).

The behavioral analysis presented in this chapter identified a series of behaviors that people use when interacting with a robotic face capable of making affective expressions and basic social interaction. Some of these behaviors involved using the same mode of facial expression in response – frowning, smiling, and mimicking the robot's expressions. Others surpassed the robot's capabilities and included hand gestures, such as waving and even feeding the robot. All the primary interaction happened within the boundaries of the robot's personal space, in proxemics terms (Hall, 1966) and did not involve broader spatial movement. This is likely a result of the constraints of our context, which involved interaction with a robotic face that stayed in one place rather than spatially mobile interaction, and placed people in a seated position in front of the robot. While this is a limitation to the generalizability of our results, it also suggests it is possible to use not just the robot's characteristics, but the characteristics of the interaction context, to constrict and simplify the problem of identifying and performing appropriate behaviors. Our

set of behavioral codes, therefore, represents a set of basic "behavioral units" a face robot will need to be able to recognize and respond to in the course of naturalistic HRI in face-to-face interaction under similar conditions.

Along with the incidence of simple behaviors, the analysis also showed the existence of behavioral clusters in which particular combinations of the identified behaviors occurred repeatedly (e.g. Mimic-Play, Caregiver Interaction). These "interaction schemas" occurred naturally, even though the interactions with the robot were unguided by researchers. We suggest they can be used to design "interaction patterns" for HRI (Kahn et al., 2008, 2010b), which are discussed further in Section 6.4.2 below. The interaction schemas we identified also suggest that there are shared models through which people interpret robots even when they receive little guidance on how to do so from researchers. We therefore showed that eliciting particular interaction patterns in participants, even when researchers are not around, is possible with minimal cues and priming. This is particularly important for HRI "in the wild", since critical readings of social interactions between people and robots suggest that successful human-robot interactions demonstrated in laboratories and other experimental settings are heavily scaffolded by researchers, and therefore may not be available when the robot operates autonomously without researchers present (Suchman, 2007; Alač, 2011). While we agree that (often unconscious) scaffolding occurs in such situations, we interpret our study as showing that predictable social interaction patterns can emerge between people and robots without direct intervention by researchers. In our study, the constraints of interactive capabilities of the robot face and a brief note left on the table in front of the robot seemed to have successfully led people to treat the robot in particular ways, e.g. engaging the robot as a child, chatting with it, adopting particular interaction schemas (caregiver, mimic-play). What is interesting is that even this minimal scaffolding, which did not require researchers to be present, led to an identifiable set of common interaction patterns among visitors.

Although our robot was very simple, people engaged in interaction schemas similar to those previously identified by Turkle in relation to more complex robots, such as PARO, Furbies, and Cog (Turkle, 2005, 2011). Gallery visitors were not only willing to take a look at the display and inspect the

robot to figure out how it works (akin to what Turkle calls the "engineering style" of interaction), but also spent time interacting with and even talking to the robot (a more "relational" style). The design of the exhibit made both types of interaction possible. The behavior units and interaction schemas we identified through behavioral analysis were also related to participants' utterances to and about the robot. It is likely that people who apply different interaction schemas in HRI will have different expectations from the robot (Lee et al., 2010) and that the robot should therefore respond in different ways to interaction partners according to the behavioral patterns it identifies. How these responses should differ will be investigated in future work (see Section 6.4.2 and 6.4.4).

While there has been prior work in HRI "in the wild", it has largely focused on understanding whether robots will be acceptable to people in naturalistic space, or on using observations of human-human interaction in the wild to identify potential behaviors that robots should be using and evaluated them in naturalistic interaction. This study recognizes that people may not always treat a robot like they would a human (e.g. they would most likely not inspect a human to see how they work), and seeks to identify a set of interactive behaviors and shared interaction patterns that people naturally use when interacting with robots. This gives us an idea of what kinds of behaviors robotic faces will need to be able to recognize and respond to in daily interaction, and how to design minimal cues that can guide people to adopt particular schemas in freeform interaction.

6.4.2 Design Implications

A key takeaway from the results detailed in this chapter is the potential design implications for future robot behavior. Indeed, closer inspection of the interaction schemas discovered here (see Section 6.3.2) suggests several implications for the design of specific robot behaviors and when/how those behaviors are enabled. For instance, one design implication is that robots could recognize particular interaction schemas, and react accordingly. A corollary to this is that certain robot capabilities may be more useful for certain schemas, e.g. having a robot engage in facial expression mimicry may be more useful in a mimic-play schema averse to a caregiver paradigm, whereas having a robot capable of

responding to vocal tone and attentional gestures may be more useful in a caregiver schema, according to our observations. Much work has been done previously on these sorts of capabilities, e.g. facial expression mimicry (Boucenna et al., 2010) and vocal tone response (Breazeal, 2003). The issue that the results here suggest is that those capabilities may only be appropriate/useful when the human interactor adopts the appropriate schema – otherwise a mismatch may occur between how the robot behaves and how the human behaves. The point is that if a robot has limited interaction capabilities (which most robots do, due to cost limitations), then those capabilities should be chosen to match expected schemas (or at least the schemas robotic designers intend to elicit). Such schemas may also be impacted by other features of the robot design, such as aesthetic qualities and form factor. The interplay between those other features and adopted interaction schemas warrants further research.

Alternatively, if the robot has a wide array of interactive capabilities, and they are capable of identifying particular interaction schemas of the human, then the robot could switch its mode of interaction to match the human's. This presents another intriguing possibility – having a robot switch between these different modes in an attempt to purposely elicit certain schemas from the human interaction partner. Given the normal dynamics seen in human-human social interaction (Warren, 2006; De Jaegher & Di Paolo, 2007), it is reasonable to assume such robotic elicitation is feasible. In short, if we can elicit certain interaction schemas in humans, this may lessen the burden on the robot to respond to the myriad of potential human behaviors. One could think of this as a sort of "scaffolding" of the environment, i.e. the robot creating its own "cognitive niche" to reduce its cognitive burden (Clark, 2013). Social interaction is a dance, after all.

Furthermore, recognizing that an interaction partner has adopted a common behavioral pattern could also simplify the task of identifying which behaviors the human might be performing at any given time towards the robot. For example, in a caregiving interaction pattern, the robot might assume that people will be smiling at it most of the time, and can use that knowledge to fill in uncertainties in interpreting the behavioral cues given by the person (e.g. facial expression, gesture, speech). We describe this potential further in Section 6.4.4, as well as Chapter 8.

Finally, as we saw predominant interactive behaviors also varied by age, group size, interaction zone, and the type of interaction (primary or non-primary) of the person, robots could use any knowledge they have of the characteristics of their human interaction partners to guide their analysis of the person's actions and their chosen responses. We found groups of people to be less likely to adopt a caregiving or mimic-play interaction with the robot, so the robot could focus on responding in a manner appropriate to such interactions when there are one or two people in its vicinity. Caregiving interactions were largely performed by adult participants, so robots interacting with children may want to focus on engaging in reciprocal play responses rather than trying to elicit and respond to nurturing by the younger participants.

6.4.3 Limitations

There are also a number of limitations. First, naturalistic data is of unquestioned importance – it is after all the space that robots likely inhabit in the future – but naturalistic data is also messy and full of conflating factors. With this in mind, we had to make a number of analysis choices, which, regardless of their correctness, still engender certain assumptions about the data and human behavior (e.g. our distinction between primary and non-primary interactors, or our separation of opening and regular exhibit phases). Another limitation was that even with minimal instruction, there is still some instruction, which included references to the robot's infant-like capabilities and asking visitors to "chat" with the robot. This may have affected the interaction schemas adopted here. Rather than a limitation, however, we could also consider this a design contribution, as it shows that minimal priming can lead people to engage with the robot in particular ways in open-ended interaction. It is hard to say what people might have done otherwise, since the interaction still would have been limited by the robot's capabilities. We also must keep in mind that the study was performed during an art exhibit, which could be a context where people might be more open to trying new things and engaging in interactions with a robot in an exploratory and ludic manner. The interaction was also open-ended and did not require people to complete any specific task. In a more task-oriented context, interactions with and opinions of a face robot might differ. Finally, a limitation with any work of this kind that relies on coding of human behavior is that there are many

nuances of human behavior that elude any coding scheme. As such, there may be some important aspects of the interactions that we cannot account for.

6.4.4 Future Work

There are a number of future avenues for pursuing the work described here further. First, one potential avenue is replicating this naturalistic study in a lab setting (which we do in Chapter 7). This allows us to empirically compare what people do in the lab settings vs. naturalistic settings, to quantify if and how they might be different. For instance, do we get the same interaction schema clusters? How are the patterns different? Such work has implications for our interpretations of lab studies in HRI, and those implications may further affect how results from HRI lab studies can be applied to real-world robotic design.

Another avenue is further exploration of the temporal networks of interaction patterns (networks mapping the flow from one behavior[s] to the next over time). Such networks allow for the calculation of the statistical relationships that characterize transitions between human behaviors (e.g. in-degree/out-degree values for each behavioral node), which can then be utilized as transition probabilities from one behavioral node to the next. These transition probabilities can serve as the basis for data-driven models to guide robot interactive behavior, such as Partially Observable Markov Decision Processes (POMDPs) (Bennett & Hauser, 2013), allowing the robot to make predictions about what future human behavior it may experience and/or what interaction schema may be occurring. In other words, if we know that a person has engaged in a certain pattern of behaviors (e.g. talking then smiling then laughing), we can use these networks to calculate the probability of future behaviors the robot might expect to see (e.g. attentional hand gestures), as well as to approximate the interaction schema that may be occurring. Such applications (without human coding intervention) are dependent on advances in activity/gesture recognition using sparse visual features, largely driven by temporal modeling algorithms such as particle filtering (Mitra & Acharya, 2007). Some behaviors though, e.g. smile and frown recognition, are already readily available. Going forward, data from this study and the ongoing lab experiments can be used to

construct such models and empirically test them during future human-robot interaction studies. We discuss some preliminary work around this in Chapter 8.

6.4.5 Conclusion

This chapter presented the analysis and findings of a naturalistic human-robot interaction study in which people were able to interact in a free-form manner with an affectively expressive face robot in an art museum. We showed that people's behaviors toward the robot differed according to their proximity to the robot, the number of people involved in the interaction, and whether they were adults or children. We also identified four recognizable interaction schemas that emerged from the data, among which caregiving and mimic-play interaction were of particular interest in terms of close-up face-to-face interaction between robots and humans. Our findings suggest that future robot design can take these interaction patterns into account to simplify the recognition and production of social behaviors by robotic faces in similar, open-ended and task-free settings. We furthermore showed that even simple cues could lead people to adopt dominant interaction patterns with the robot, which is a promising approach for scaffolding future interactions between people and robots "in the wild." Finally, we detail the future directions arising from the presented work, which will focus on the development of probabilistic, temporal models of behavioral interaction based on the interaction patterns we identified.

Chapter 7

Comparing Human Interaction with a Robotic Face in-the-lab vs. in-the-wild: An Empirical Study

The previous chapter focused on analyzing and identifying common behavioral patterns of humans interacting with a robotic face in a naturalistic setting. In this chapter, we explore how such interaction patterns from a naturalistic setting compare to those from the lab. Are they the same, or are they different?

Abstract. We performed an empirical, lab-based study (n=72) of the effects of contextual factors (setting, culture) on how people socially interact with an autonomous, interactive robotic face. The empirical study was intended to replicate a previous naturalistic, "in-the-wild" study performed in a public museum exhibit using the same robot, allowing us to directly compare what people do in the lab with what people do "in-the-wild". We also performed a cross-cultural component of the lab study between the United States and Japan, allowing us to compare differences in how Japanese and American subjects interacted with the robotic face. There were a couple significant findings. First, what people did in the lab-based experiments and what people did in the naturalistic museum setting was fundamentally different. However, such differences could be quantified, potentially allowing for them to be accounted for in data-driven models of robot social behavior. Second, we found a number of similarities and differences between the Japanese and American human subjects interacting with the robotic face. This cross-cultural variation in interaction patterns suggests specific interaction behaviors that could be targeted for enhancing face-to-face robotic interaction in these cultures.

7.1 Introduction

7.1.1 Background

In the previous chapter (Chapter 6), we focused on identifying common behavioral patterns of humans interacting with a robotic face "in-the-wild", identifying a number of interaction schemas that appear to be emergent in such settings during voluntary, undirected interaction. In this chapter, we evaluate whether those findings apply across different contexts, e.g. controlled lab settings and/or cross-culturally. A principal question is whether interaction patterns found in naturalistic settings are the same as those we might observe in lab-based experiments designed to closely replicate the naturalistic setting (a public museum exhibit). In other words, *is what people do in human-robot-interaction lab experiments different from what they do in natural settings?*

This is a fundamental question – if we intend to build models to guide future robot behavior based off of human-robot interaction (HRI) data from lab settings, would those models be applicable to how people behave in the real-world? If they are different, then how? Can we quantify the difference? The answers to such questions may enable us to account for such differences in our models (if the differences follow predictable patterns). Alternatively, they may suggest that lab-based models are perhaps not the most useful for designing real-world robot interaction behaviors (Šabanović, Michalowski, & Simmons, 2006; Mutlu & Forlizzi, 2008; Walters et al., 2011).

The work described here comprised two separate studies that were meant to resemble each other, the difference being one was conducted in the lab and the other "in-the-wild" in a public museum exhibit. Both studies (museum and lab) used an iteration of the robotic face MiRAE (described in Chapter 2) – equipped with 3D printed components, basic social interactions capabilities, and the ability to see/respond to human interactors in its environment. The study setting – the desk/chair setup, props/materials on the desk, etc. (see Section 6.2.2) – was arrayed in both the museum and lab in roughly the same fashion. In both cases, explicit instruction on how to interact with the robot, or its capabilities, was kept to a minimum. The primary difference in the lab is that subjects were brought into the lab setting, given an explicit task to perform (interact/entertain the robotic face) in a pre-specified timeframe. In other words,

the lab experiments were like most lab-based experiments – structured rather than free-form, and performed in an artificial, controlled environment.

Of course, even if the naturalistic and lab-based interaction patterns were not fundamentally different, a question exists whether this would hold universally, or perhaps be culture-specific. In other words, would lab-based interaction patterns be the same across cultures? Indeed, previous work has found fundamental differences in human-robot interaction across cultures [Bartneck et al., 2007; Li, Rau, & Li, 2010; Lee & Sabanović, 2014; Sabanović, Bennett, & Lee, 2014]. To explore this question, the lab-based experiments were conducted cross-culturally (in the United States and Japan). We should note that the naturalistic setting (museum) was performed only in the United States (and as such we only compare the U.S. lab experiments to it). We thus have two primary, separate questions in this study: *1) comparing the naturalistic interaction patterns to the lab-based ones (U.S. only), and 2) comparing U.S. lab-based interaction patterns with those from Japan.* Both questions address the critical role that context plays in shaping human-robot interaction, and holds important implications for designing future robot social behavior across such contexts.

7.1.2 Related Work

A comprehensive overview of human-robot interaction studies "in-the-wild" is provided in Section 6.1.2. In short, a number of such studies have been performed in museums (e.g. Nourbakhsh, Kunz, & Willeke, 2003; Yamazaki et al, 2009), malls (e.g. Kanda et al., 2009), university campuses (e.g. Gockley et al., 2005), city streets (e.g. Weiss et al. 2010), schools (e.g. Tanaka, Cicourel, & Movellan, 2007; Leite et al., 2012), and public areas of caregiving institutions (e.g. Chang & Sabanovic, 2014). The study here extends upon that previous work, empirically exploring how lab-based experiments compare and the effects of culture on observed interaction patterns.

There is a relative dearth of studies doing direct empirical comparisons of lab and naturalistic studies in HRI, although more broadly there has been work on the subject in psychology. For instance, Neal & Wood (2009) explored the nature of habit forming in naturalistic and lab settings, finding that lab-

based experiments may alter psychological processes by directing conscious attention to what are normally subconscious, unattended processes and thus creating conscious goal-mediated actions that are not present in naturalistic settings. This held true even when there was an element of misdirection in the experiment (i.e. when the task was used as a cover for what was actually being studied). In other words, the very act of created a task-oriented lab experiment altered the underlying psychological mechanisms used to perform various daily activities (e.g. social interaction). Dunbar (2001) found similar effects in analogy-making psychological tasks, and Belsky (1980) observed the same effects in mother-infant interaction. At the same time, an open question is to what degree such effects may alter the generalizability of lab results – even if the cognitive processes are altered to some degree, the results may still be the same or similar (Anderson & Bushman, 1997; Anderson, Lindsay, & Bushman, 1999). Thus, the same questions extend to HRI, and empirical work is needed to understand how such effects may impact the results of HRI lab experiments, and how such lab experiments may compare to naturalistic, "in-the-wild" studies (Šabanović, Michalowski, & Simmons, 2006; Mutlu & Forlizzi, 2008)

7.2 Methods

7.2.1 Robotic Face

The platform used in this study (MiRAE) is a minimalist robotic face shown in Figure 2.1, previously described in (Bennett & Šabanović, 2014; Bennett et al., 2014; Bennett & Šabanović, 2015). It is capable of basic, non-verbal, infant-like social interaction behavior. It can detect faces and motion, respond to people, and make a variety of facial expressions (e.g. frown, smile). It has the ability to track environmental stimuli both relative to its sensory (retinotopic) and motor (spatiotopic) coordinates, and follow them using a neck mechanism. It also has a basic visual attention and affective system.

This is the same version of MiRAE described in the previous chapter (See Section 6.2.1 and Chapter 2 for more details). The exact same version was used in both the museum and lab settings. MiRAE's motor, visual, and cognitive functions are written as C++ and Python libraries, and available as open-source software online (http://r-house.soic.indiana.edu/projects/mirae.html). The computer vision

aspects are based on OpenCV (opencv.org). MiRAE's physical construction is designed to be replicable, using easily accessible components (e.g. Arduino microcontrollers) and 3D printed facial components, with 12 degrees of freedom including a pan & tilt neck mechanism (full construction details available online: http://r-house.soic.indiana.edu/mirae/MiRAE_Construction_Manual.pdf).

MiRAE has been experimentally validated previously in terms of affective interaction in (Bennett & Šabanović, 2014; Bennett et al., 2014; Bennett & Šabanović, 2015), showing that it is capable of producing higher, or at least comparable, identification accuracy rates for the Ekman facial expressions (Happiness, Sadness, Anger, Fear, Surprise) as a number of other robotic faces, including Kismet (Breazeal, 2003), Eddie (Sosnowski et al., 2006), and the android Geminoid-F (Becker-Asano & Ishiguro, 2011).

7.2.2 Study Setting

There are two settings for this study: 1) museum, and 2) lab. The museum exhibit setting is described in detail in Section 6.2. In short, it took place in an art gallery set in the downtown of a Midwestern United States city. It was open to the public. People could come by, either individually or in groups, and voluntarily interact with the robot (as shown in Figure 6.1). The interaction was free-form and undirected – researchers were not present on-site except to turn the robot on/off at the beginning and end of each day. Minimal instruction for interacting was provided in the exhibit itself, noting that the robot was capable of infant-like behavior (e.g. it could see people, detect faces, respond). People were free to interact with the robotic face in their own terms, for as long as they liked, and to interpret the robot's responsive behavior however they wished. The robot face exhibit was designed to look like a working roboticist's lab desk. The desk was arrayed with various materials, electronic components, and tools used in the building/creation of robots. People often used the materials as props or toys when interacting with the robot (e.g. "feeding" the robot), though that was not by intentional design as part of the research.

The lab setting took place in the r-house lab at Indiana University (http://r-house.soic.indiana.edu) and a university in Yokohama, Japan. In both places, a dedicated room was setup to resemble the museum exhibit (as shown in Figure 6.2), replete with desk, chair, computer, and various materials and components to be used as props or toys. Cameras were positioned in the same location as the museum (back right corner of the desk). Averse to the museum, subjects came to the lab, and were directed to the room. They went through the full process of completing informed consent, filling out research instruments, and being given explicit task instructions for interacting with the robot (more details provided in Section 7.2.3 below). In other words, it was a controlled, structured experiment. The lab experiment was performed in both the United States and Japan, using the exact same setup and the same physical robotic face. Subjects were given the exact same task instructions in both cases, using a script (in English or Japanese depending on the location) read by a native language speaker. Sample size was 36 in each country (total n=72) for the lab experiments. Subjects were college students. The average age was 21.5, and the gender mix was exactly 50/50 males to females.

The lab setting was in direct contrast to the museum exhibit, where participants voluntarily interacted with the robot face when passing by, completed no research forms, were given minimal instructions. In fact, based on collected video, most participants in the museum setting did not appear to be aware they were involved in a research study, or that they were even being video recorded (no indication of those were given in the museum exhibit). In short, subjects in the lab setting were acutely aware that they were being observed, and that there was some research "goal" in mind during their participation, whereas those in the museum setting were generally unaware of either of those.

7.2.3 Experimental Design

In this section, we will focus on describing the lab experiments. The museum exhibit is described in Section 6.2.3.

After informed consent was obtained, subjects were given several forms/instruments to complete. Subjects were administered the Negative Attitudes toward Robots Scale (NARS, prior to each

experiment) (Nomura & Kanda, 2003) and Godspeed instruments (after each experiment) (Bartneck et al., 2009). Explicit instructions were given about the task – that 1) the subject was to engage/entertain the robot for the next several minutes, 2) that the robot was capable of infant like social interaction, 3) that the robot could see/hear/react to their presence, and 4) detect motion and faces. They were not told explicitly *what to do* (e.g. play peekaboo with the robot), but rather given a general sense of the robot's capabilities and what was expected of them. How to entertain/engage the robot was still left up to their terms, as well how they interpreted interpreted the robot's responsive behavior. The interactions were timed (unlike the museum exhibit), so that each person interacted with the robotic face for approximately 3 minutes.

The autonomous interaction behavior of the robot in this lab experiments was the same as the museum exhibit (described in Section 6.2.3), and can be summed up as follows (a link to the programming code is provided in Section 7.2.1). The robot, if it detected any sort of motion in its visual field, would immediately begin tracking/following that motion (if multiple points of motion were detected, it would choose the point of largest motion). If the robot detected specific stimuli (e.g. a person or face), it would preferentially attend to that (over detected motion). If multiple stimuli were detected, it would first attend to the closest one, then shift its attention to ones further away, based on simple attentional decay (similar to Rubi [Movellan et al., 2007]). If no motion or stimuli were detected, the robot would make random saccade motions. The affective system was also operating simultaneously. If novel stimuli were detected, the robot would express surprise or interest. Examples of novel stimuli might include a face, given that a face had not been detected recently, or stimuli suddenly increasing in size as if rapidly moving toward the robot. Otherwise, in the presence of positive stimuli (e.g. a person or face) that appeared to be interacting with the robot, the robot would express happiness (smile). If during the interaction, the person/face suddenly departed or moved out of view, the robot would express sadness (frown). For instance, a person could play peekaboo with the robot face by covering their face and then uncovering it. Finally, if a person completely stopped moving/interacting for a period of 7-8 seconds, the robot would become "bored" and begin searching for new stimuli.

Data collected for later analysis during the experiment included video from both the onboard and offboard video cameras. Motion data (optical flow and intensity gradients) sampled from a 5x5 grid across the visual field were also recorded as numerical data, at every timepoint (approximately twice per second). Internal proprioceptive data was also recorded from the robot, which included positions of its motors, internal affective states, locations of each detected stimuli (if any), and information about the current attended stimuli (if any) at every timepoint. All the data was time-stamped, so that, for instance, internal proprioceptive data could be matched later to video data, etc.

7.2.4 Analysis Approach

The analysis of the data collected from the museum exhibit and lab-based experiments is broken into two parts, based on our two principle questions here: *1) comparing the naturalistic interaction patterns to the lab-based ones (U.S. only), and 2) comparing U.S. lab-based interaction patterns with those from Japan.* Similar to the previous chapter (Chapter 6), we focus on analysis of the offboard video data, with the aim of understanding in detail the kinds of interaction behaviors people displayed toward the robotic face.

The analysis of the museum exhibit data is described in Section 6.2.4. For the lab-based experiments, we followed a similar procedure. Only a single pass on coding the video data was required, since in the lab we had controlled, 3-minute interactions with clear/enforced beginnings and ends. Annotated coding of each interaction video was performed using Anvil (http://www.anvil-software.org/) by the same two independent coders as performed the museum exhibit coding. Based on overlapping coding sets (roughly 10% of the total video data), Inter-class correlation was previously calculated via SPSS, with a Cronbach's Alpha of 0.734, which is considered good. The same basic annotation scheme was used for the lab as the museum, with a couple minor tweaks based on what was observed in the museum. Attentional Hand Gestures were broken out into 3 subcategories, which allowed us to look more closely at what kind of gestures were being performed, but could still be summed back up to directly compare to the museum exhibit.

Coded behaviors included: Smiling, Frowning, Other Facial Expressions, Making Exaggerated Faces, Sticking Tongue Out, Talking, Laughing, Evasion (attempting to evade the robot's field of view, moving in and out of the field of view), Mimic Robot (making opposing expressions in an attempt to get the robot to mimic them), Communicative Hand Gestures (e.g. waving hello or goodbye), Attentional Hand Gestures (etc. snapping, pointing, finger-wagging), Feeding the Robot, Inspect Robot (examining the robot's structure rather than making direct eye contact with it). Attentional Hand Gestures were further broken down into 3 sub-types: Clapping, Peekaboo, and Attentional Other (i.e. anything not clapping or Peekaboo). Other than the facial expressions and Attentional Hand Gesture sub-types (e.g. a person could not smile and frown at the same time), these coded behaviors were not mutually exclusive, i.e. a person could smile and talk at the same time.

Behaviors were coded for both occurrences and time spent (in seconds). For analysis purposes, time spent was scaled by the total duration of the interaction, since different people interacted for different lengths of time. In essence, this converted the time spent into a unit-free "time spent per second" value (i.e. what percentage of each second did the person spend doing behavior 'x'), independent of the actual duration. This allowed us to directly compare different interactions, and interactions between the lab and museum that may have lasted for varying durations.

In the next section, we first compare the lab (U.S. only) interaction data with the museum exhibit interaction data, and then second compare the U.S. lab interactions with the Japan lab interactions. In both cases, statistical hypothesis testing was performed using independent samples *t*-test in SPSS. We adopt the null hypothesis as our starting point, i.e. there would be no differences in interaction behavior across country or lab vs. museum setting. For the cross-cultural comparisons, we also evaluated subject perceptions of the interaction via the NARS and Godspeed instruments; however, no statistically significant differences were found, and these will not be discussed further in the analysis.

Additionally, in the analysis of the lab vs. museum data, we analyzed the coded interaction behavior data using unsupervised clustering to examine whether the lab participants adopted similar identifiable "interaction schemas" while interacting with the robot that were discovered in the naturalistic

museum setting (described in Section 6.3.2), We utilized the same two-step clustering in SPSS (http://10.110.22.85:49801/help/topic/com.ibm.spss.statistics.algorithms/alg_2step_cluster.htm) used previously to identify such clusters (Section 6.2.4), and evaluated differences in the behavioral patterns between the interaction schemas observed in the naturalistic museum setting, and those seen in the lab. In short, do human interactors in the lab adopt similar interaction schemas?

7.3 Results

7.3.1 Museum vs. Lab Interaction

Museum interactions and lab-based interactions (U.S. only) were compared across twelve interaction behaviors (described in Section 7.2.4), both for people count (i.e. the number of people engaging in each behavior) and time spent (scaled per second, see Section 7.2.4). Since the museum interactions were previously coded without using the sub-types of attentional hand gestures, those were omitted from the comparison. Percentage values for both people count and time spent are shown in Tables 7.1 and 7.2, respectively.

Table 7.1: Museum vs. Lab – People Count

	Facial Expressions								Hand Gestures			
Setting	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Communicative	Attentional	Feeding	Inspecting
Museum	66.0%	7.0%	13.0%	15.0%	4.0%	76.0%	40.0%	5.0%	38.0%	23.0%	3.0%	18.0%
Lab	94.0%	22.0%	33.0%	39.0%	17.0%	97.0%	86.0%	6.0%	56.0%	89.0%	0.0%	3.0%

Table 7.2: Museum vs. Lab – Time Spent

	Facial Expressions								Hand Gestures			
Setting	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Communicative	Attentional	Feeding	Inspecting
Museum	17.83%	0.73%	0.83%	1.57%	0.41%	15.07%	2.48%	0.98%	4.13%	3.12%	0.62%	2.92%
Lab	22.97%	0.46%	2.00%	3.99%	0.32%	34.42%	4.04%	0.16%	4.56%	43.23%	0.00%	0.03%

The raw data in Tables 7.1 and 7.2 were evaluated for statistical significance via an independent samples t-test. Significant differences are shown in Table 7.3, with p values in parentheses.

Table 7.3: Museum vs. Lab – Significant Differences

	Facial Expressions								Hand Gestures			
Comparison	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Communicative	Attentional	Feeding	Inspecting
People Count	5.90 (.000)	2.17 (.037)	2.53 (.015)	2.82 (.007)		5.47 (.000)	7.01 (.000)		2.08 (.038)	11.1 (.000)	2.68 (.008)	4.23 (.000)
Time Spent						4.68 (.000)				7.73 (.000)		5.46 (.000)

As can be seen from Table 7.3, there are numerous differences in nearly every category regarding the percentage of people who engaged in various behaviors. There were fewer differences in time spent, though those differences turned out to be major ones. People in the lab were much more likely to engage in attentional hand gestures (people count: 89% vs. 23%, time spent: 43% vs. 3%) and Talking (people count: 97% vs. 76%, time spent: 34% vs. 15%), while people in the museum were more likely to engage in feeding and inspecting behaviors (feeding behaviors were in fact no observed at all in the lab experiments). There were also differences in communicative hand gestures, laughing behaviors, and various facial expressions. In short, what people did in the lab-based experiments and what people did in the naturalistic museum setting was fundamentally different.

In general, people in the lab engaged in a greater variety and frequency/intensity of behaviors. This can be seen in Table 7.4. The intensity/frequency of behaviors was over double in the lab compared to the museum (total behaviors per second), with people engaging in 1.16 behaviors every second in the lab averse to 0.51 in the museum. In other words, people were in essence constantly engaged in behaviors (on average) during the lab experiment, whereas people in the museum had pauses in between behaviors, and a more deliberate/paced approach. This is also reflected by the total number of behaviors each person engaged in, with people in the lab engaging in roughly 5-6 behaviors during the interaction (out of the total 12 possible), and individuals in the museum engaging in about 3 different behaviors. In short, people in the lab engaged in a wider variety of behaviors, and their interaction was more intense.

Table 7.4: Museum vs. Lab – Total Behavior Comparison

Setting	Total Behaviors per Second	Total # Behaviors
Museum	0.51	3.07
Lab	1.16	5.42
t-test	9.15 (.000)	8.26 (.000)

We equate these differences in the types of (Table 7.3) and variety/intensity (Table 7.4) of interaction behaviors to the fact that the lab-based experiments had an explicit task, which altered their behavior in fundamental ways. Subjects in the lab setting were acutely aware that they were being observed, and that there was some research "goal" in mind during their participation, whereas those in the museum setting were generally unaware of either of those.

Finally, we evaluated the lab interaction data to see whether we could recover the same or similar interaction paradigms we saw in the previous chapter via clustering analysis (Section 6.3.2). In particular, we were interested in seeing whether the caregiver and mimic-play "interaction schemas" that were spontaneously observed during the naturalistic museum interactions would replicate themselves in the lab-based experiments. Results are shown in Table 7.5. The museum clusters are replicated from Table 6.5 in the previous chapter. Lab interaction data is in the column on the right.

Table 7.5: Museum vs. Lab – Cluster Analysis

		Museum				Lab
		1	2	3	4	
	N (All Phases)	74	132	17	33	
	N (Post Only)	48	84	7	24	
Facial Expressions	Smiling	11.6%	12.2%	40.2%	20.8%	23.0%
	Frowning	0.0%	0.1%	0.0%	3.9%	0.5%
	Other	0.1%	0.0%	0.0%	5.8%	2.0%
	Exaggerated	0.3%	2.8%	0.6%	1.0%	4.0%
	Tongue Out	0.0%	0.6%	0.0%	0.3%	0.3%
	Talking	12.6%	16.2%	48.0%	17.0%	34.4%
	Laughing	1.5%	1.1%	18.6%	2.0%	4.0%
	Mimicking	0.0%	0.0%	0.0%	6.0%	0.2%
Hand Gestures	Communicative	2.2%	6.8%	2.4%	1.8%	4.6%
	Attentional	1.7%	3.0%	20.3%	3.2%	43.2%
	Feeding	0.1%	0.2%	17.8%	0.7%	0.0%
	Inspecting	1.7%	3.5%	0.0%	3.2%	0.0%
	Cluster Label	Non-Primary	Low-Intenstity/ Non-Specific	Infant-Like Caregiver Interaction	Mimic-Play	

In short, the clustering analysis was unsuccessful at recovering those paradigms, and was in fact unable to produce any meaningful clusters from the lab interaction data. In other words, there were no discernible patterns in the lab experiments, and all of the subjects formed a single amorphous cluster. We can see in Table 7.5 that this single lab cluster does not fit into any of the interaction schemas seen in the museum. This reinforces the above point that what people do in lab-based experiments and what people do in the naturalistic settings appears to be fundamentally different.

7.3.2 Cultural Differences in Lab-Based Interaction

Identical lab-based experiments were conducted, involving interaction with the robotic face, in both the U.S. and Japan. Interaction data was compared across fifteen interaction behaviors, including the sub-types for attentional hand gestures (described in Section 7.2.4), both for people count (i.e. the number of people engaging in each behavior) and time spent (scaled per second, see Section 7.2.4). Percentage values for both people count and time spent are shown in Tables 7.6 and 7.7, respectively.

Table 7.6: Japan vs. US – People Count

	Facial Expressions									Hand Gestures					
										Communicative	Attentional				
Country	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Evasion	Communicative	Clapping	Peekaboo	Attentional Other	Feeding	Inspecting
Japan	94.0%	17.0%	22.0%	22.0%	0.0%	89.0%	56.0%	0.0%	69.0%	83.0%	33.0%	25.0%	81.0%	0.0%	19.0%
USA	94.0%	22.0%	33.0%	39.0%	17.0%	97.0%	86.0%	6.0%	44.0%	56.0%	17.0%	22.0%	89.0%	0.0%	3.0%

Table 7.7: Japan vs. US – Time Spent

	Facial Expressions									Hand Gestures					
										Communicative	Attentional				
Country	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Evasion	Communicative	Clapping	Peekaboo	Attentional Other	Feeding	Inspecting
Japan	24.4%	0.4%	0.4%	2.6%	0.0%	9.9%	1.5%	0.0%	6.0%	16.7%	3.4%	1.1%	39.1%	0.0%	0.7%
USA	23.0%	0.5%	2.0%	4.0%	0.3%	34.4%	4.0%	0.2%	5.3%	4.6%	0.9%	1.1%	41.3%	0.0%	0.0%

The raw data in Tables 7.6 and 7.7 were evaluated for statistical significance via an independent samples t-test. Significant differences are shown in Table 7.8, with p values in parentheses.

Table 7.8: Japan vs. US – Significant Differences

										Communicative	Attentional				
Comparison	Smiling	Frowning	Other	Exaggerated	Tongue Out	Talking	Laughing	Mimicking	Evasion	Communicative	Clapping	Peekaboo	Attentional Other	Feeding	Inspecting
People Count					2.65 (.012)		2.99 (.004)		2.18 (.032)	2.65 (.010)					2.30 (.026)
Time Spent						5.65 (.000)				3.33 (.002)					2.27 (.009)

As can be seen in the tables, Japanese and US subjects were remarkably similar in many ways (e.g. smiling, peekaboo, and other attentional hand gestures), but also showed some stark differences in specific behaviors. Japanese subjects were more likely to engage in communicative hand gestures (i.e. waving), evasive body movements, and inspecting the robot. US subjects engage in laughing, talking, and sticking their tongue out. Of particular note, Japanese spent much less time engaged in verbal behavior (9.9%) compared to the American subjects (34.4%), even though roughly the same percentage of people in both countries did talk at least once during the experiment (89% vs. 97%). It is possible that the Japanese subjects' greater use of communicative hand gestures was intended as a non-verbal substitute for such verbal social behaviors towards the robot.

Both the similarities and differences between Japanese and US subjects interacting with the robotic face raise interesting possibilities for interpretation. We explore these possibilities further in Section 7.4.

7.4 Discussion

This study addressed research questions related to differences in how people interact with robotic faces in different contexts, e.g. naturalistic settings vs. lab-based experiments and across different cultures. We focused on two primary questions here: *1) comparing the naturalistic interaction patterns to the lab-based ones (U.S. only), and 2) comparing U.S. lab-based interaction patterns with those from Japan.* There were several significant findings. First, we found that what people did in the lab-based experiments and what people did in the naturalistic museum setting was fundamentally different, and were able to quantify those differences. We also found that interaction patterns in lab settings did not fit into "interaction schema" clusters that had been previously identified in-the-wild. Second, we found a number of similarities and differences between the Japanese and American human subjects interacting with the robotic face. This cross-cultural variation in interaction patterns suggests specific interaction behaviors that could be targeted for enhancing face-to-face robotic interaction in these cultures.

These results have a number of potential implications. First, the fact that there are fundamental differences in interaction patterns between the lab and naturalistic, "in-the-wild" settings suggests that models for guiding robot social behavior based on lab data may not be applicable to real-world settings, at least not without modification. As many research labs around the world, including our own, attempt to data-driven design models of robot behavior from lab experiments, this warrants caution. Studies, such as the one presented here, may ameliorate this issue by quantifying the differences between lab and naturalistic settings. Quantifying the differences allows us to account for them in our models, for instance adjusting the probability values in transition models used in probabilistic, temporal frameworks used in action/decision-making reasoning (e.g. Markov Decision Processes [MDPs]) described in the previous chapter (Section 6.4.4, Bennett & Hauser, 2013).

At the same time, the inability to recover the same "interaction schemas" clusters in the lab experiments as seen in the naturalistic museum setting is problematic. It suggests that there are critical aspects (observable phenomena) of human-robot social interaction that occur in naturalistic settings that are not observable at all in the lab. Quantifying the differences does not reconcile this problem. And indeed, as we saw in Chapter 6, such phenomena as people adopting common paradigms for social interaction may be critical to understanding important aspects of what is occurring. For instance, adoption of different interaction schemas may underlie different social expectations of the robots and their behavior by the humans, influencing the way those humans identify with the robots (Lee et al., 2010), and thus necessitate different behavioral responses from the robot to successfully engage human interactors. The derivation of and successful application of these kinds of "interaction patterns" may be challenging in some lab-based experiments (Kahn et al., 2010b).

Finally, there were distinct patterns of similarities and differences in interaction patterns with the robotic face across cultures (Japan vs. United States). In particular, there seemed to be greater use of non-verbal (e.g. communicative hand gestures, evasive body movements) in Japanese subjects, and a greater use of verbal behaviors (e.g. talking, laughing) in U.S. subjects. It is possible that the Japanese subjects' greater use of communicative hand gestures was intended as a non-verbal substitute for such verbal social

behaviors towards the robot. At the same time, some interaction behaviors, such as smiling, peekaboo, and other attentional hand gestures, showed remarkable similarities between the two cultural groups. Such cross-cultural variation in interaction patterns suggests specific interaction behaviors that could be targeted for enhancing face-to-face robotic interaction in culturally-specific ways. Critically, the presence of *both* similarities and differences between cultures indicate that sensitivity and adaptivity to nuances in social cues and behaviors may be sufficient for robots to engage in meaningful social interaction across cultures, rather than outright re-design of such robots for specific cultures (Šabanović, Bennett, & Lee, 2014).

In conclusion, how people socially interact with an autonomous, interactive robotic face depends on a number of contextual factors, e.g. lab vs. naturalistic setting, cultural background of the subject. *What people do in the lab is fundamentally different than what people do in naturalistic, "in-the-wild" settings*. On the other hand, understanding the effects of such contextual factors allows us to explicitly account for them, whether that be in designing models of future robot social behavior for real-world settings or adapting robots for different cultures.

Chapter 8

Future Directions

So far, we have discussed completed studies related to the dynamics of social interaction between people and robotic faces and the development of an empirically-grounded robotic face. In this chapter, we discuss ongoing, related work not completed at the time of writing: creating probabilistic, temporal models for guiding robot behavior, making sense of visual data the robot "sees" during social interaction, and clues from borderline personality disorder towards creating robotic face "personalities".

Abstract. None

8.1 Introduction

So far throughout the preceding chapters, we have described a number of completed studies related to the dynamics of social interaction between people and robotic faces and the development of an empirically-grounded robotic face. This chapter is a departure from that, focusing on ongoing, related work not completed at the time of writing:

1) creating probabilistic, temporal models for guiding robot behavior
2) making sense of visual data the robot "sees" during social interaction
3) clues from borderline personality disorder towards creating robotic face "personalities"

These three efforts are at various stages. #1 has been currently at a data analysis stage, focusing on previously collected interaction data, but no implementation has yet been attempted on the robot during real-time interaction. #2 is further along in the analysis stage, with some trial implementation and preliminary results available. Finally, #3 is only conceptual at this point, focusing on existing literature and hypotheses about how that might apply to the functioning/programming of a robotic face. All three of these topics represent potentially exciting developments stemming from the work up to this juncture.

Given the incomplete nature of these efforts at the current time, we provide a brief, informal introduction to each one below, before launching into our conclusion in the final chapter.

8.2 Temporal Dynamics (e.g. rhythmicity, synchronicity) in Human-Robot Social Interaction: Towards Developing Future Models to Guide Interactive Robot Behavior

One challenge in applying much of what has been learned in preceding chapters, such as the museum exhibit studies in Chapters 6 and 7, is that while we can obviously glean behavioral patterns during human-robot social interaction from observing/analyzing those interactions, it is not necessarily clear what we "do" with such patterns. How do we then apply that to a robot or robotic face's functioning? How do we inject such patterns into its programming? Even simpler, what questions should

we ask? For example, is it necessary for a robotic face to know that specific interaction behaviors are being performed by a human interactor, or is simply being able to distinguish whether a human is attempting to engage in interaction sufficient to ameliorate human-robot interaction?

We discuss some of this in Section 6.4, discussing the potential derivation of transition probabilities of one interaction behavior to another using network analysis techniques, which could then be fed into Partially Observable Markov Decision Processes (POMDPs), a common algorithm used for rational decision-making over time in both robotics and artificial intelligence (AI) applications (Bennett & Hauser, 2013). POMDPs (and their simpler, fully observable MDP cousins) are probabilistic, temporal frameworks that allow us to calculate optimal actions even in the face of uncertainty over the absolute state of the world (we can make sensory observations, but must infer the underlying environment indirectly through them) and the effects of actions we may take. POMDPs do this by operating in the realm of *belief states*, which allows us to perform efficient Bayesian inference of future states based on probabilistic relationships between observations, actions, and the actual state of the world. It can also take into account costs and rewards of particular actions and choices a robot or AI system may take. Moreover, MDPs/POMDPs can also be designed as online AI agents – determining an optimal policy at each timepoint (t), taking an action based on that optimal policy, then re-determining the optimal policy at the next timepoint (t+1) based on new information and/or the observed effects of performed actions (Littman, 2009; Russell & Norvig, 2010).

The application of such probabilistic, temporal models is still a work in progress. The first stage is developing temporal networks of interaction patterns (networks mapping the flow from one behavior[s] to the next over time), as described in Section 6.4.4. Such networks allow for the calculation of the statistical relationships that characterize transitions between human behaviors (e.g. in-degree/out-degree values for each behavioral node), which can then be utilized as transition probabilities from one behavioral node to the next. Examples of such temporal networks can be seen in Figures 8.1 and 8.2 for Clusters 3 (caregiver interaction) and Clusters 4 (mimic-play) that were described in Chapter 6 (see Section 6.3.2).

Figure 8.1: Temporal Network – Cluster 3 (Caregiver Interaction)

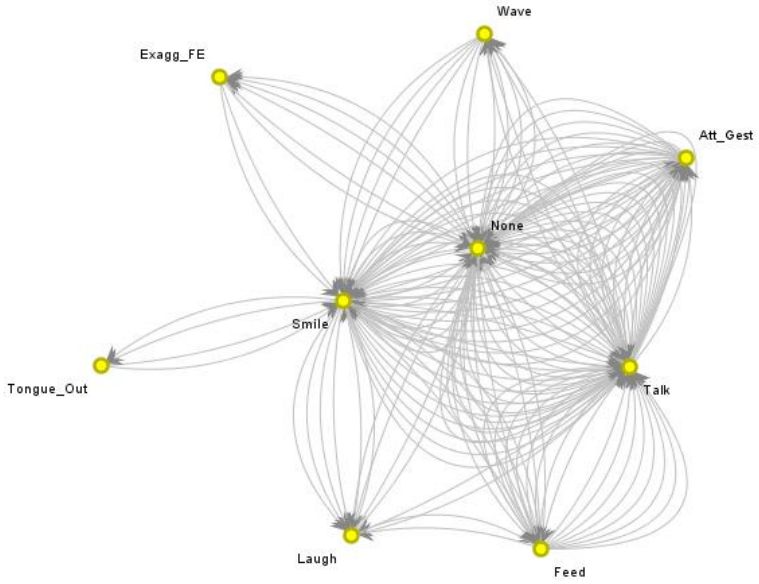
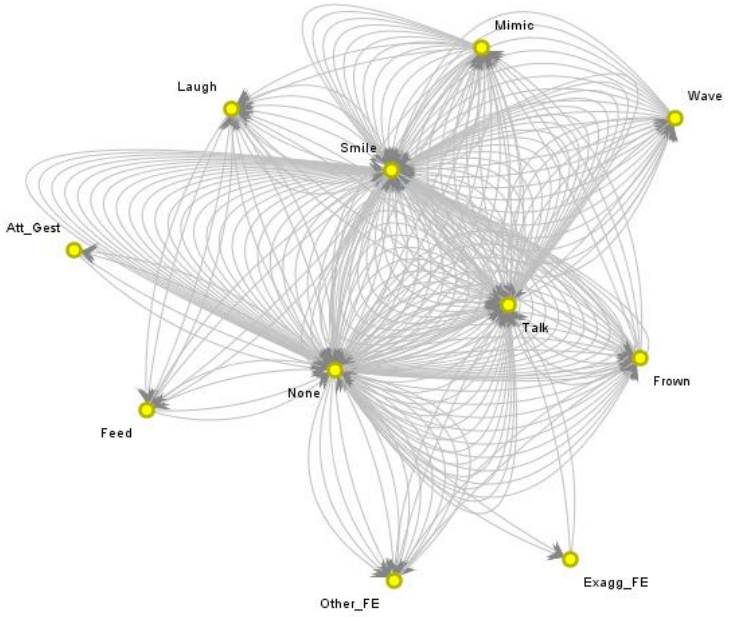


Figure 8.2: Temporal Network – Cluster 4 (Mimic-play)



As can be seen in the figures, there are notable differences in how the different behavioral nodes connect (e.g. the number of edges) as well as the centrality of certain nodes in different clusters. These reflect the differences seen in Table 6.5. The difference here is that the temporal element has been added – it is differences in how behaviors change from a given timepoint to the next that are fundamental.

The above work is ongoing. Another related topic focuses on answering an even simpler question. What if the robotic face could simply detect whether an interactor was present from the visual data, and could react accordingly? We describe this in the next section.

8.3 Making Sense of What a Robotic Face "Sees": Machine Learning and Sparse Visual Data

This section will cover several related efforts. First, we describe the preliminary implementation of a neural network on the robotic face that seeks to detect a simple dichotomy – whether an interactor is or is not present based on sparse visual data patterns of motion, rather than intensive Haar-cascade or related algorithms that attempt to detect specific objects (such as people or faces) using sophisticated but computationally demanding approaches (Lienhart & J Maydt, 2002). Second, we describe efforts to process the sparse visual data a robot sees and relate that to specific human interaction behaviors, and potentially to the robot's own behaviors (e.g. facial expressions).

Several algorithms, such as Haar-cascade algorithms, exist for detecting specific objects, e.g. people or faces or hands. The challenge with such algorithms is that they are: 1) computationally-intensive, and 2) simply detecting a person or a face does not indicate whether it is attempting to interact with you. From a social interaction perspective, what engages us with other entities is not so much what they "are", but what they "do", aspects such as synchrony and entrainment come into play. This is why we don't attempt to interact with statues, or faces of Jesus we see in a pancake. So, an even simpler question is whether we can detect a potential interactor from sparse visual data a robot might collect.

To do this, the robotic face platform captured motion data (optical flow and intensity gradient deltas) across a 5x5 grid in its visual field at every timepoint, i.e. 25 sampled data points from the roughly 20,000 pixels available at every timepoint. The philosophy behind this is that early visual systems had

limited capacity, with only a handful of photoreceptor cells that could only detect at gross levels of visual acuity (Land & Nillson, 2012), but still could perform evolutionarily important tasks, such as distinguishing between light and dark, or potential threats in the environment based on quick movement. Modern human visual systems are much more sophisticated of course, but interaction between organisms extends far back into the realm of simpler visual systems. As such, we start by emulating the simpler visual systems (in this case assuming the presence of 25 photoreceptor cells), and build from there.

A feed-forward neural network was implemented on the robotic face platform, utilizing the 25 sampled grid points as representing 25 photoreceptor cells. 77 input were used – 75 "visual neurons" (25 optical flow on the x-axis, 25 optical flow on the y-axis, 25 intensity gradient delta) and 2 "motor neurons" that took in the pan/tilt neck positions of where the robot was looking at the time. 40 hidden neurons were used (roughly half the input neurons). 2 output neurons were used representing neck pan and tilt movement bias, i.e. the output biased whether the robotic face looked toward the motion or continued random saccades. The idea was to train the robotic face so that when an interactor was present, it would attend to said interactor, when an interactor was not present (or perhaps present but not interacting), it would ignore him/her.

Rewards were based on whether the robot continued to detect a person based on traditional Haar cascade algorithms that were running concurrently with the neural network. The neural network was not privy to the information from the Haar cascade algorithms until after the fact, however (after it made its predictions and acted on them). In other words, rewards for learning purposes were based on whether a human interactor was in fact detected, and back-propogated into the neural network to adjust its future predictions. The idea was that if the robotic face attended to an interacting human, that human would engage longer, while if the robotic face ignored them, it would cause the interaction to end, or end more quickly.

To pilot test this, we used lab personnel, having them alternately move in and out of the field of the robotic face. Results from one of these trial runs can be seen in Figures 8.3 and 8.4.

Figure 8.3: Adaptive Neural Network – Positive and Negative Predictions

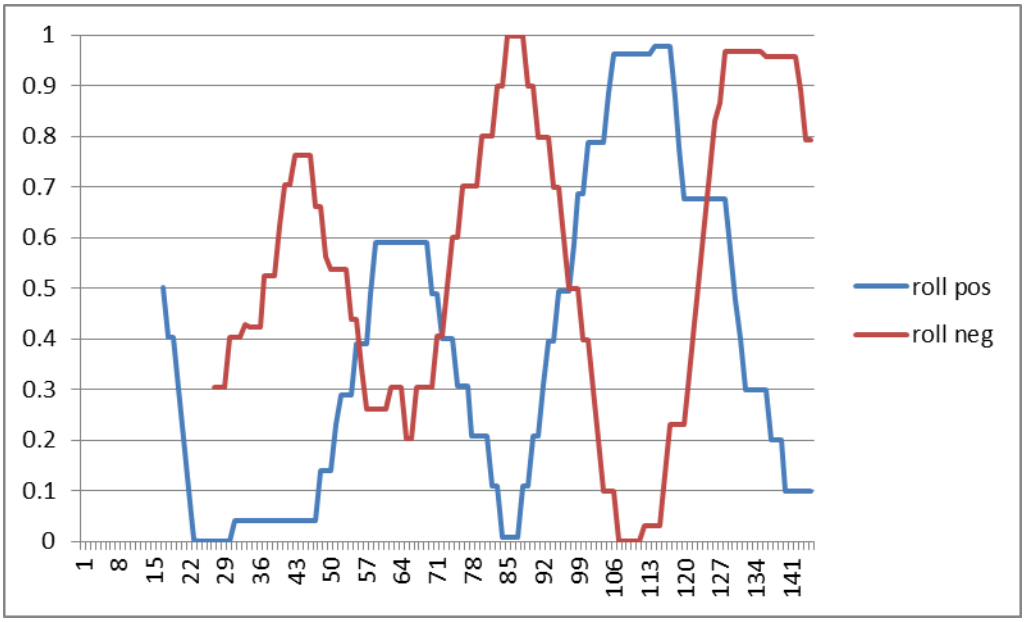
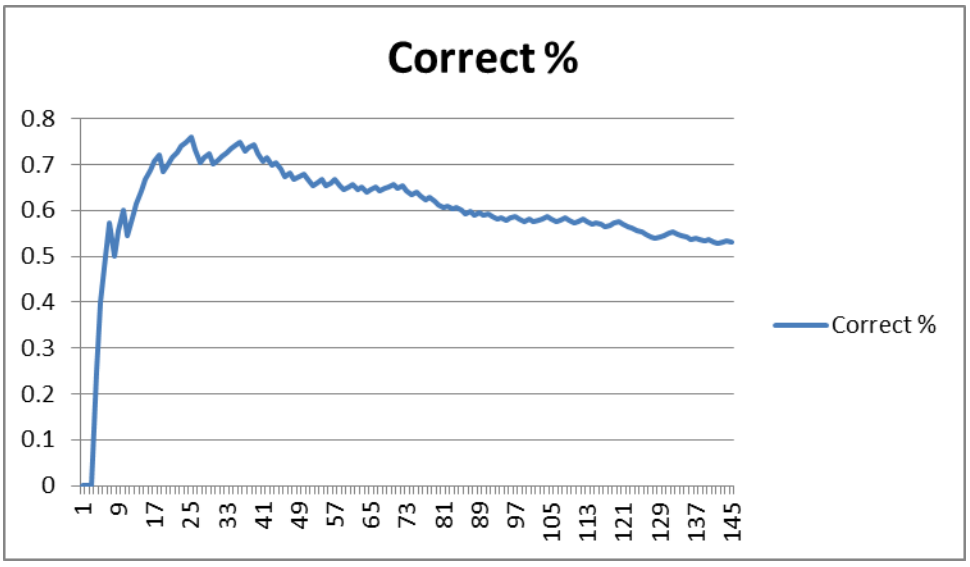


Figure 8.4: Adaptive Neural Network – Overall Accuracy Rate



At every timepoint, the neural network produces two predictions – positive (there is a human interactor present) and negative (there is NOT a human interactor present) – as probabilities on the scale 0-1. In Figure 8.3, we show the rolling positive and negative prediction error rates (by rolling we mean

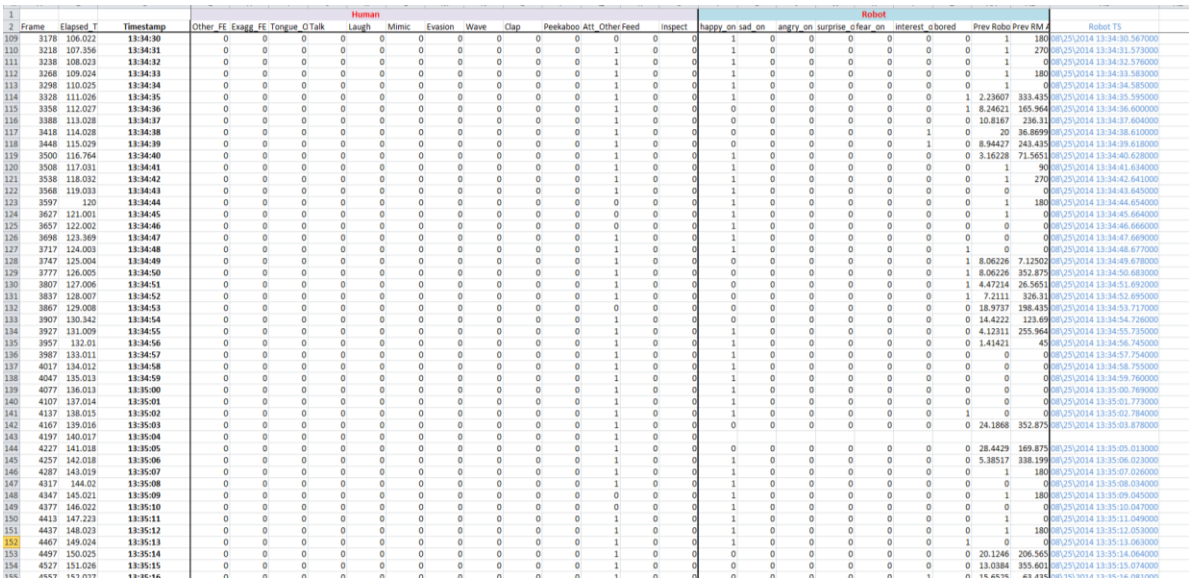
averaged over the last 10 timepoints, to reduce noise). The goal here would be to have both positive and negative predictions gradually reduce to zero. Instead, we can clearly see the points where the human interactor is moving in and out of the frame – the prediction that is true in that case slowly reduces its error rate, while the other one climbs. What appears to be happening is that – since the predictions are coming from the same neural network – that the neural network is training itself to its current situation, and forgetting whatever it learned prior. Again, this is a pure feed-forward network, it has no recurrence, and thus no memory.

The same notion can be seen in the overall accuracy rate, shown in Figure 8.4. The two predictions – positive and negative – essentially compete against each other, with the highest one in terms of probability value being the "winner." We can see after the initial training (first ~30 timepoints), in which the robotic face has only seen one condition, a steady drop off in the accuracy rate towards 0.5. Given this is a binary prediction, 0.5 essentially means that we have no predictive power. Other trial runs replicated this result.

This is all very initial work, so it is not surprising that we were unsuccessful at our initial attempts to integrate adaptive neural networks into robotic face behavior. However, it does shed light on a couple potential improvements that could ameliorate our performance. First, the possibility exists that adding some recurrence to the neural network would help prevent it from "forgetting" what it's learned. Adding such recurrence can be as simple as adding a history node to the network that modulates current predictions based on what was recently observed. Often the state of the world has a relationship to what happened prior, and such information can be useful in making better predictions. Second, another possibility exists in creating two separate neural networks, one for positive prediction and one for negative, that run concurrently and undergo their own learning processes. From a biological standpoint, these are essentially parallel networks, but the disconnect allows learning to occur in the face of dichotomous information. Such structural disconnect might even explain the presence of dichotomous thinking in natural cognition. All of the above ideas, of course, are subject to further experimentation. At the moment, they are only possibilities.

Finally in this section, we will briefly describe efforts to process and find patterns in the sparse visual data the robot "sees" and relate that to human interaction behavior as well as the robot's. This turned out to be a more laborious process than originally envisioned, and as such we don't yet have data or results to share, but we can describe the process. This entails four steps. First, frame-by-frame annotations are exported from the video coding software (Anvil) into tab-delimited files. Second, data is converted using a template into a format so that the behaviors can be manipulated via analysis software. Third, the data is further converted using a second template into a condensed format, so that we have one row per data, and a single "behavior activation" field that exists as a binary string of 1's and 0's. Finally, the coded behavior data is added to data from the robot's own behavioral logs. The result is something along the lines shown in Figure 8.5.

Figure 8.5: Aggregated Coding Behavior Data



				Human															Robot									
2	Frame	Elapsed_T	Timestamp	Other	FE_Exagg	FE_Tongue_O	Talk	Laugh	Mimic	Evasion	Wave	Clap	Peekaboo	Att_Other	Feed	Inspect	happy_on	sad_on	angry_on	surprise_o	fear_on	interest_o	bored	Prev Robo	Prev RM.4	Robot TS
109	3178	106.022	13:34:30	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	180	08\25\2014 13:34:30.567000
110	3218	107.356	13:34:31	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	270	08\25\2014 13:34:31.573000
111	3238	108.023	13:34:32	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	08\25\2014 13:34:32.576000
112	3268	109.024	13:34:33	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	180	08\25\2014 13:34:33.583000
113	3298	110.025	13:34:34	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	08\25\2014 13:34:34.585000
114	3328	111.026	13:34:35	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	2.23607	333.435	08\25\2014 13:34:35.595000
115	3358	112.027	13:34:36	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	8.24621	165.964	08\25\2014 13:34:36.600000
116	3388	113.028	13:34:37	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	10.8167	236.31	08\25\2014 13:34:37.604000
117	3418	114.028	13:34:38	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	20	36.8699	08\25\2014 13:34:38.610000
118	3448	115.029	13:34:39	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	8.94427	243.435	08\25\2014 13:34:39.618000
119	3500	116.764	13:34:40	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	3.16228	71.5651	08\25\2014 13:34:40.628000
120	3508	117.031	13:34:41	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	90	08\25\2014 13:34:41.634000
121	3538	118.032	13:34:42	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	270	08\25\2014 13:34:42.641000
122	3568	119.033	13:34:43	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:34:43.645000
123	3597	120	13:34:44	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	180	08\25\2014 13:34:44.654000
124	3627	121.001	13:34:45	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	08\25\2014 13:34:45.664000
125	3657	122.002	13:34:46	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:34:46.666000
126	3698	123.369	13:34:47	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:34:47.669000
127	3717	124.003	13:34:48	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	08\25\2014 13:34:48.677000
128	3747	125.004	13:34:49	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	8.06226	7.12502	08\25\2014 13:34:49.678000
129	3777	126.005	13:34:50	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	8.06226	352.875	08\25\2014 13:34:50.683000
130	3807	127.006	13:34:51	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	4.47214	26.5651	08\25\2014 13:34:51.692000
131	3837	128.007	13:34:52	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	7.2111	326.31	08\25\2014 13:34:52.695000
132	3867	129.008	13:34:53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18.9737	198.435	08\25\2014 13:34:53.717000
133	3907	130.342	13:34:54	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	14.4222	123.69	08\25\2014 13:34:54.726000
134	3927	131.009	13:34:55	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	4.12311	255.964	08\25\2014 13:34:55.735000
135	3957	132.01	13:34:56	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1.41421	45	08\25\2014 13:34:56.745000
136	3987	133.011	13:34:57	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:34:57.754000
137	4017	134.012	13:34:58	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:34:58.755000
138	4047	135.013	13:34:59	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:34:59.760000
139	4077	136.013	13:35:00	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:35:00.769000
140	4107	137.014	13:35:01	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:35:01.773000
141	4137	138.015	13:35:02	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	08\25\2014 13:35:02.784000
142	4167	139.016	13:35:03	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	24.1868	352.875	08\25\2014 13:35:03.878000
143	4197	140.017	13:35:04	0	0	0	0	0	0	0	0	0	0	1	0	0										
144	4227	141.018	13:35:05	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	28.4429	169.875	08\25\2014 13:35:05.013000
145	4257	142.018	13:35:06	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	5.38517	338.199	08\25\2014 13:35:06.023000
146	4287	143.019	13:35:07	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	180	08\25\2014 13:35:07.026000
147	4317	144.02	13:35:08	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:35:08.034000
148	4347	145.021	13:35:09	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	180	08\25\2014 13:35:09.045000
149	4377	146.022	13:35:10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	08\25\2014 13:35:10.047000
150	4413	147.223	13:35:11	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	08\25\2014 13:35:11.049000
151	4437	148.023	13:35:12	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	180	08\25\2014 13:35:12.053000
152	4467	149.024	13:35:13	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	08\25\2014 13:35:13.063000
153	4497	150.025	13:35:14	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	20.1246	206.565	08\25\2014 13:35:14.064000
154	4527	151.026	13:35:15	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	13.0384	355.601	08\25\2014 13:35:15.074000

Note that to the far right in Figure 8.5 (in the light blue font), are timestamps for each row. This allows us to match each row to the sparse visual data at that exact same timepoint. Such data could then be added to the dataset shown in Figure 8.5, and pulled into analysis or machine learning software, to hunt for patterns in the data in an automated fashion. For instance, is there any correspondence between

the robot's behaviors at a given time and the human's? Can we detect the human's behaviors from the sparse visual data alone? All of this remains to be done, but we have the data and the process pipeline to determine the answers to such questions going forward.

8.4 Robotic Face Personality and a "Sense of Self": Clues from Human Borderline Personality Disorder

One of the primary issues with creating robots for social interaction purposes is how to create natural seeming behavior from what is otherwise pre-programmed, prescriptive computer code. Naturally intelligent organisms, such as humans, don't all behave the same way. There may be some underlying chemical or instinctive "programming" at play, but the fruition of that is idiosyncratic in terms of individual behavior. In laymen's terms, we often speak about people as having some "sense of self", a personality. The question is how we might replicate such idiosyncratic personalities in robots and robotic faces in an emergent yet replicable way (without explicitly programming different "selves" into different robots).

A useful potential starting place for searching to an answer to that question comes from psychological investigation into human personalities, in particular pathological cases of such personalities, otherwise known as *personality disorders*. Studying pathological cases – places where normal personality manifestations break down – can provide useful clues to where "personality" and/or a "sense of self" comes from in natural cognitive systems. Indeed, those break downs, and their developmental etiology, point to critical aspects of personality and the formation thereof. One particularly informative personality disorder towards this end is known as *borderline personality disorder* (BPD).

BPD is a complex disorder, a combination of genetic/neurobiological factors and environmental triggers. BPD develops across a person's life, with primary causative factors associated with early childhood, and a displacement of the attachment system that underlies many later normal emotional reactions and behaviors, such as relationships, bonding, flight or fight responses, and impulsive behavior

regulation (e.g. drug/alcohol abuse, hypersexuality). According to James Masterson (1998), the main issue in BPD is failure of proper individuation from the primary caretaker during the pre-oedipal stage, due to a variety of reasons (e.g. trauma or sexual abuse). This stage is fundamental toward the development of self, the understanding by the child that they can separate from the caretaker without being abandoned. Young children will often test this, by leaving for short periods of time but "checking in" on a regular basis. For a child at this stage, abandonment equals death. Children who fail to properly individuate at this critical stage have difficulty seeing themselves as a separate person, exhibiting disordered attachment and thinking that alternates between enmeshment/clinging behaviors and avoidant ones, often in chaotic ways. They carry this core wound through life. Indeed people with BPD (pwBPD) show marked changes in neurobiological functioning as adults: changes to amygdala functioning (affecting memory formation and emotional reactivity), neurotransmitter levels (e.g. dopamine, serotonin, epinephrine), and neural activity communication patterns within the neocortex. In short, the child never properly develops a "sense of self", and in its place develops a set of maladaptive, primitive defensive/coping mechanisms (e.g. projection, disassociation, mirroring), which are exhibited to much lesser degrees in non-personality disordered individuals. BPD however, in its exaggeration of normal personality defense mechanisms, provides a unique window into the construction of personality and self. pwBPD, in fact, have no sense of self, only a false self, which shifts based on the personalities of others around them (Masterson, 1998; Fonagy, 2008). They are often described as emotionally hypersensitive, or, as Dr. Marsha Linehan put it, "People with BPD are like people with third degree burns over 90% of their bodies. Lacking emotional skin, they feel agony at the slightest touch or movement" (Mason & Kreger, 2010).

The name for BPD itself is a bit of a misnomer. Early psychologists, confused by the odd presentation of symptoms, initially thought that BPD lied on the "borderline" of psychosis and neurosis, as the patients showed clear neurotic traits, with only fleeting glimpses of psychotic episodes. In other words, they were neither neurotic nor psychotic, but somewhere in between. The current Diagnostic and

Statistical Manual of Mental Disorders (DSM IV, 2000) defines BPD as a spectrum disorder (like autism) in people exhibiting at least 5 of the following 9 traits, in a persistent manner across time:

1) **Frantic efforts to avoid real or imagined abandonment**
2) **A pattern of unstable and intense interpersonal relationships** characterized by alternating between extremes of idealization and devaluation
3) **Identity disturbance**, such as a significant and persistent unstable self-image or sense of self
4) **Impulsivity** in at least two areas that are potentially self-damaging (e.g., spending, sex, substance abuse, reckless driving, binge eating)
5) **Recurrent suicidal behavior, gestures, or threats, or self-mutilating behavior**
6) **Emotional instability** due to significant reactivity of mood (e.g., intense episodic dysphoria, irritability, or anxiety usually lasting a few hours and only rarely more than a few days)
7) **Chronic feelings of emptiness**
8) **Inappropriate, intense anger or difficulty controlling anger** (e.g., frequent displays of temper, constant anger, recurrent physical fights)
9) **Transient, stress-related paranoid thoughts** or severe dissociative symptoms

Such symptoms may manifest as outright rages towards a significant other or family member, or in passive-aggressive behaviors, or other acting-out behaviors such as marital infidelity. pwBPD often do not remember doing such behaviors after the fact (known as disassociation). Disassociation can be thought of as an extreme form of "zoning out". For instance, driving somewhere and arriving, but being unable to remember the trip, is a form of mild disassociation. Moods can also shift frequently and rapidly in pwBPD, with dramatic swings from high to low, happy to sad. They suffer persistent paranoid thoughts, such as a partner cheating even though there may be no evidence thereof. They project their own maladaptive behaviors onto others, a defense against painting themselves as "bad". They alternately split others, seeing them as all good or all bad (since they have no sense of self, they fail to perceive others as whole selves that can be simultaneously *both* good and bad, i.e. black or white thinking).

Relationships, even close friendships, with pwBPD often follow a pattern of idealization and devaluation that ends in such splitting behavior. They can sometimes seem chameleon-like, shifting their personality and interests based on those around them. When triggered into severe emotional dysregulation, they often engage in bouts of destructive impulsive behaviors, such as promiscuous sexual activity, verbal/physical abuse, or alcohol binges. pwBPD are typically not aware of such maladaptive behavior – the reality they experience is the only one they know. Most people without personality disorders can experience some of these BPD traits for short periods of time, but they are not persistent. To a non-BPD person, such behaviors and reactions can seem extremely confusing, chaotic, and/or bizarre.

The presentation of symptoms may only be apparent in close, intimate relationships (intimacy is often a trigger for BPD behavior). In fact, many pwBPD may be what is loosely termed "high-functioning" – they may be highly intelligent, able to hold down professional jobs (lawyers, teachers), and seem otherwise normal, if perhaps slightly quirky. Estimates put the prevalence of BPD in the general population at 2-4%, with roughly 75% of those being females. In other words, a conservative estimate would be that at least 1 out of every 50 people suffer from BPD (Lenzenweger et al., 2007; Grant et al. 2009).

There are multiple theories for how BPD manifests from a neurobiological sense, as a product of both nature and nurture. Current research points to the likelihood of multiple manifestations leading to the same array of symptoms that are characteristic of the disorder. One commonality is that there appears to be some temporal aspect to all the manifestations, some disruption to the normal process of memory formation and their functioning in relation to emotional activation/regulation. This leads to a disconnect between external sensory information and internal perceptual states, i.e. the pwBPD inappropriately perceives and reacts (or overreacts) to environmental stimuli. Fuchs' (2007) details how BPD is really a failure in the ability to integrate past, present, and future into one coherent stream. Like experiencing life as a series of disconnected fragments. Hence the lack of sense of self, the impulsivity, identity disturbance, disassociation, etc. Interestingly, emotion is heavily tied to memory formation, via the amygdala. Emotions serve an important evolutionary purpose, in that they are associated with specific

memories, and allow for quick reactions to certain situations. The neo-cortex in humans (our logical thinking center) is intended to mitigate those initial reactions, if necessary. In BPD however, these normal processes have gone haywire. The amygdala goes into overdrive, and there is a lack of executive control.

The temporal aspect of BPD described above (e.g. deficits in memory formation and associated emotion activation/regulation) points to critical clues in understanding the formation of personality. One of these is that the linking of external sensory information to internal perceptual states, integrated over time, serves as the critical cog in a sense of self, autonomy. The world is defined by such perceptual states. Indeed, in the extreme case of a pwBPD, <u>feelings are facts</u>.

For a non-BPD individual, this is still true, though to a lesser degree. Non-BPD integrate information over time, allowing for contradictory information to exist simultaneously. Without this (in BPD world), the aforementioned coping mechanism known as splitting is triggered, essentially classifying things into binary states: good vs. bad, black vs. white, true vs. false. Our standard machine learning classifiers fall into such a splitting paradigm. However, the key to natural intelligence is that perception (in a healthy individual) supersedes those binary states. The binary perceptual states are projected onto the world, via emotional responses that trigger certain behaviors, leading to new perceptions. The validity of the external state of the world to those internal perceptual states can never be absolutely confirmed. Rather, *we project in order realize*. We assume the state of the world fits our internal reality, perform behaviors consistent with that projection, then observe whether the world changes in response as predicted, integrated over time. Feelings are facts, ones that we resolve into an essence of our "sense of self" in the world. We need no internal model of the state of the world, only a model of emotion and appropriate emotional-behavioral response.

BPD is a pathological case of this, showing us aberrantly how such projections can go wrong, and how, even in the face of this, we still assume our internal reality is correct. We revise the facts (the world) to fit our feelings (our internal reality), not the other way around. We learn by adjusting our

emotional-behavioral responses via memory formation and recalibration. Situations which we find problematic in this regard lead to cognitive dissonance.

Temporal fragmentation of such feelings and perceptions in pwBPD provides us a unique window into the criticality of time in these processes, into how we make sense of the world and our place within it. It is not simply a matter of being able to detect or classify (e.g. is a chair vs. is not a chair) – it is about being able resolve contradictory information.

So how can we use this information gleaned from personality disorders like BPD to help us develop robotic face personalities? Here, we lay out one potential method of constructing a robotic face personality based around this information. The critical aspect is, rather than developing some sort of system that detects states of the world and bases behavioral responses on them, we instead link sensory information to projections of the robot's internal reality (in a simple case, a robot's internal "emotional" state), which is then projected onto the world. For instance, if the robot is "happy", then the world is "happy", and the robot reacts accordingly. Learning takes place in the space of the emotion-experience connections, i.e. associating experiences/memory (sensory input) with internal perceptual states. Importantly, the robot never verifies that the external world agrees with its projection – it looks for incongruences in the sensory feedback to adjust the emotion-experience connections. The robot "assumes" that its projection is correct. <u>The robot's "feelings" are fact</u>.

This approach is all conceptual at this point, but let us look at one hard example. We can extend the simple case of the robot being happy from above. Let us say we have a human interactor in the robotic face's field of view, performing some interaction behaviors. The robot has some method of perceiving the human interactor and its activity levels (e.g. motion via optical flow in its field of view). At the initial point of detection, the robot might be, say, happy that it has detected a new interactor in field of view. The happiness triggers a related behavior (a happy facial expression). The robot projects this internal perceptual state on the world, and the related behavior is performed (probabilistically). On the next timestep, the robot could look for validation from its sensory input. One simple way might be the activity levels, perhaps they increase (i.e. total absolute optical flow increases). The robot takes this as

validation that its internal perceptual state is, in fact, a valid representation of the world. The connection between the emotional state and the sensory input might be strengthened (e.g. the probability between the two could be increased within whatever probabilistic framework was being employed). In the future, such sensory states would be more likely to trigger the emotional response, *but also* be more likely to be expected during subsequent projection. The robot would continue to project its internal emotional state onto the world, and perform associated behaviors, until the sensory input shifts enough to trigger a change in internal perceptual state (an open question would be how large a triggering shift would be necessary, machine learning techniques would likely need to be employed to learn the difference between a triggering shift and small incongruence). This would lead to different projections, and different associated behaviors. For instance, if the activity levels suddenly drop in some significant way, this may trigger the robot to shift into a "sad" state.

There are a couple key points to understanding the subtle change here from previous approaches to creating models of interactive robot behavior. One key point here is that, even though we are using labels like "happy", we have to disentangle those during implementation. Sensory input is related to internal perceptual states, which relates to behaviors. However, none of that is determined *a priori*. The approach to learning here is backwards. The robot starts by "assuming" whatever it senses is related to the projection of its internal perceptual state (which initially might be randomized, or perhaps primed with some prior interaction data … presumably such priming might be the purpose of much instinctive behavior in natural organisms). In other words, the robot assumes its feelings are facts. Incongruences are accounted for by adjusting how those facts relate to internal feelings. The feelings do not change, only what "experiences" are associated with them. If they don't agree, then there is something wrong about the experience, something wrong with the sensory input. This differs from previous models of artificial emotion, such as Kismet (Breazeal, 2003), where the starting point is trying to discern states of the world (i.e. facts are facts) and have the robot learn how to respond. The second point is that these internal perceptual states are intimately connected to embodied behavior. We can see this clearly with facial expression behavior, but it pervades through many other forms of behavior as well. Without some

sort of embodied behavior, there would be no need for these internal perceptual states, or their subsequent projections, because they are only important in that they drive behavioral responses to a world which is not directly accessible to us. Hence, these internal perceptual states derive from such embodied behavior. In the simplest case, these perceptual states can be visualized as binary polar opposites, perhaps in multidimensional form. Indeed, if one glances back at the 3D circumplex model of emotion in Figure 5.1 (Chapter 5), we can see such a model for human emotion.

The fundamental point here is that emotions are not only *responses* to sensory information about the world, they are also *projections* of the state of the world over short-term time intervals. This is all conceptual at this point, but raises interesting ideas about how we might create robotic face personalities based on such ideas. Since the robots will be learning their own internal perceptual state model based on their experiences, and each will have slightly different experiences, the end result may be robotic faces that respond differently to similar stimuli. In short, each one would have its own "personality".

Chapter 9

Conclusion

9.1 General Conclusion

The purpose of this dissertation is two-fold: 1) to develop an empirically-based design for an interactive robotic face, and 2) to understand how dynamical aspects of social interaction may be leveraged to design better interactive technologies and/or further our understanding of social cognition. In the preceding chapters, we explored the above questions across a range of studies, including lab-based experiments, field observations, and placing autonomous, interactive robotic faces in public spaces. We also discussed future work (see Chapter 8), how this research relates to making sense of what a robot "sees", creating data-driven models of robot social behavior, and development of robotic face personalities.

The findings showed the minimal features necessary to communicate emotion and engage in affective interaction via a robotic face, as well as the factors (e.g. added neck motion) that contribute to that. We also showed that previous theories about cross-cultural differences in facial expression recognition did not seem congruent with empirical data, and that context effects were capable of over-riding any differences that were present. We found that context congruency (the alignment of the facial emotion with emotion triggered by the contextual environment) had a significant effect on human perceptions, and that this effect varied by the emotional valence of the context and facial expression. Moreover, these effects occurred regardless of the cultural background of the participants. The results suggested a form of *projection.* Emotions perceived in the faces of others – including robots – appeared to be an internal construct in the mind of the perceiver, based on a number of perceptual and cognitive processes (Bennett & Šabanović, 2015; Barrett, Mesquita, & Gendron, 2011).

The autonomous, interactive robotic face was also deployed in a public museum exhibit, and allowed to interact with people in naturalistic interactions without the researchers present. Clustering revealed four well-defined “interaction schemas” from the interaction behavior data. These results suggest

that people often adopt specific interaction schemas when interacting with a robotic face in a free-form, naturalistic setting (outside the lab), schemas identifiable from the interaction data itself. Such findings hold design implications for future robot interactive behavior. Finally, we compared the interaction data from the museum and lab-based experiments. The key finding was that what people did in the lab-based experiments and what people did in the naturalistic museum setting was fundamentally different. However, such differences can be quantified, potentially allowing for them to be accounted for in data-driven models of robot social behavior.

The above details all of our major findings, from the humble beginnings of trying to validate that the robotic platform even worked, to setting it loose in the world to see what happened without the constraints of scientific assumption and lab sterility. But can we derive from the above a single conclusion? What principle aspect, above all others, would stand out? For us, it is the following principle:

What occurs in the course of social interaction, and social cognition alike, is the product of not merely the interaction itself, nor the interactors (human or robot). Rather, social interaction is really the confluence of many different components, a phenomenon that takes root within some contextual environment. And not simply a physical one, but a sum total of experiences of the interactors (e.g. their cultural background) and the world itself. And much of this experience is projection, a reality that only exists in our heads. We see this point illustrated not only in our work here (Chapters 4 and 5), but also echoed in human psychology (Section 8.4). This lies at the heart of the nature of *sociality*.

Teasing apart these aspects leads to this simple conclusion: if we want to design interactive technologies, we are really designing the system, not the technology itself. But that interaction – what we see, what we experience – is a projection of our internal reality. Thus the system we are designing is as much about manipulating reality, as it is manipulating illusion.

9.2 Moving Forward

Moving forward, there remain a number of exciting avenues to further this research, some of it ongoing currently. Many of these avenues are laid out previously in various individual chapters, as well as Chapter 8. We refer the reader back to Chapter 8 for an in-depth review of ongoing work around temporal dynamics, making sense of what the robot "sees", and development of robotic face personalities. Here, we summarize some broader future directions of the research, focused on culture-neutral models of social interaction, aesthetic design work, and some potential applications of interactive robotic faces like the ones used here.

First, taking advantage of context effects around social interaction may ease the constraints for developing culturally-specific affective cues in human-robot interaction, opening the possibility to create culture-neutral models of robots and affective interaction and/or social interaction. Given that culture is dynamic and constantly in flux, it may not make sense to design robots *in toto* for specific cultures, but rather to design robots that are sensitive and adaptive to particular cultural factors, temporal ones included. We suggest this can be done in two ways: 1) by making the robot design process more culturally reflexive and inclusive of the perspectives of diverse stakeholders, and 2) by designing robots to be sensitive and adaptable to salient cultural cues – an approach we have termed *Culturally Robust Robotics* (Šabanović, Bennett, & Lee, 2014).

Second, work is currently underway to explore various aesthetic design features of robotic faces using modern 3D printing technology, which allows for rapid prototyping. Such aesthetic features include the design of facial components, e.g. lips, eyes, including aspects such as shape, color, and texture. However, the focus is also on aspects outside these explicit facial components, e.g. the shape of the face, what materials it is made out of (plastic vs. metal), size, schema-approach (animal-like vs. abstract). Such research is a long-term agenda, literally an agenda unto itself, with millions of potential questions that could be asked. Nonetheless, a simple question lies at its heart: what should an interactive robotic face *look* like? Potential participatory design approaches with human users are envisioned for some of this research, alongside the rapid prototyping approaches listed above.

Finally, a relevant question is what sorts of applications might be possible for interactive robotic faces like the ones used here (MiRAE) in the future. Beyond simply being a novelty, or for entertainment, how could they be useful? A generic answer would be that any robot operating in a human social environment could benefit from more natural social interaction. Moreover, many of the findings from this work could apply to any interactive technology (e.g. clinical AI, smartphones, clinical decision support systems) – indeed that was the initial impetus behind much of the work. All of this is true, and worthwhile for further pursuit. However, we also feel that the potential exists for specific applications for interactive robotic faces themselves. One of these is in the healthcare realm, particularly as a socially-assistive robot (SAR) for elderly people experiencing chronic physical and mental illness (Šabanović, Bennett, & Piatt, 2014). Recent trends in healthcare have seen a shift toward prevention and patient-centered care, toward in-home settings and away from institutionalized ones (e.g. hospitals, nursing homes). There is a significant opportunity for novel technologies, such as socially-assistive companion robots, to help facilitate that shift, in particular with older adults having chronic health issues. Interactive robotic faces such as MiRAE hold potential to contribute to such efforts.

10. References

1. Adams B, Breazeal C, Brooks RA, and B Scassellati (2000) Humanoid robots: a new kind of tool. *IEEE Intelligent Systems and their Applications*. 15(4): 25–31.
2. Alač M, Movellan J and F Tanaka (2011) When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics. *Social Studies of Science*. 41(6): 893–926.
3. Allison B, Nejat G, and E Kao (2009) The design of an expressive humanlike socially assistive robot. *Journal of Mechanical Robotics*. 1(1): 011001.
4. Anderson CA and BJ Bushman (1997) External validity of" trivial" experiments: The case of laboratory aggression. *Review of General Psychology*. 1(1): 19.
5. Anderson CA, Lindsay JJ, and BJ Bushman (1999. Research in the Psychological Laboratory Truth or Triviality? *Current Directions in Psychological Science*. 8(1): 3-9.
6. Anderson K and PW McOwan (2006) A real-time automated system for the recognition of human facial expressions. *IEEE Trans Syst Man Cybern B Cybern*. 36(1): 96–105.
7. Arizpe J, Kravitz DJ, Yovel G, and CI Baker (2012) Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PLoS ONE*. 7(2): e31106.
8. Aronoff J, Woike BA, and LM Hyman (1992). Which are the stimuli in facial displays of anger and happiness? Configurational bases of emotion recognition. *J Pers Soc Psychol*. 62(6): 1050–1066.
9. Asada M, Hosoda K, Kuniyoshi Y, Ishiguro H, Inui T, Yoshikawa Y, et al. (2009) Cognitive developmental robotics: A survey. *IEEE Transactions on Autonomous Mental Development*. 1(1): 12–34.
10. Auvray M, Lenay C, and J Stewart (2009) Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*. 27(1): 32–47.

11. Barrett LF, Mesquita B, and M Gendron (2011). Context in emotion perception. *Current Directions in Psychological Science*. 20(5): 286–90.
12. Barsalou LW, Breazeal C, and LB Smith (2007) Cognition as coordinated non-cognition. *Cogn Process*. 8(2):79–91.
13. Barrett LF and EA Kensinger (2010) Context is routinely encoded during emotion perception. *Psychological Science*. 21(4): 595-599.
14. Barrett LF, Mesquita B, and M Gendron (2011) Context in emotion perception. *Current Directions in Psychological Science*. 20(5): 286–290.
15. Bartneck C and M Okada (2001) Robotic user interfaces. *Proceedings of the Human and Computer Conference (HC2001)*, pp.130–140.
16. Bartneck C, Suzuki T, Kanda T, and T Nomura (2007) The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society*. 21(1-2), 217–30.
17. Bartneck D, Kulic E, Croft M, and S Zoghbi (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*. 1(1): 71-81.
18. Bazo D Vaidyanathan R, Lentz A, and C Melhuish (2010) Design and testing of a hybrid expressive face for a humanoid robot. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.5317–5322.
19. Becker-Asano C and H Ishiguro (2011) Evaluating facial displays of emotion for the android robot Geminoid F. *Proceedings of the IEEE Workshop on Affective Computational Intelligence* (WACI), pp.1–8.
20. Bechara A, Damasio H, and AR Damasio (2000) Emotion, decision making and the orbitofrontal cortex. *Cereb Cortex.* 10(3):295–307.
21. Beer RD (1995) A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*. 72(1–2): 173–215.

22. Beer RD (2000) Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*. 4(3): 91–99.

23. Belsky J (1980) Mother-infant interaction at home and in the laboratory: A comparative study. *The Journal of Genetic Psychology*. 137(1): 37-47.

24. Bennett CC and K Hauser (2013) Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*. 57(1): 9-19.

25. Bennett CC and S Šabanović (2013). Perceptions of affective expression in a minimalist robotic face. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.81-82.

26. Bennett CC and TW Doub (2014) Temporal modeling in clinical artificial intelligence, decision-making, and cognitive computing: Empirical exploration of practical challenges. *Proceedings of the 3rd SIAM Workshop on Data Mining for Medicine and Healthcare (DMMH)*.

27. Bennett CC and S Šabanović (2014) Deriving minimal features for human-like facial expressions in robotic faces. *International Journal of Social Robotics*. 6(3): 367-381.

28. Bennett CC, Šabanović S, Fraune M, and K Shaw (2014) Context congruency and robotic facial expressions: Do effects on human perceptions vary across culture? *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp.465-470.

29. Bennett CC and S Šabanović (2015) The effects of culture and context on perceptions of robotic facial expressions. *Interaction Studies*. In Press.

30. Berns K and J Hirth (2006) Control of facial expressions of the humanoid robot head ROMAN. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.3119–3124.

31. Biehl M, Matsumoto D, Ekman P, Hearn V, Heider K, Kudoh T, and V Ton (1997) Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*. 21(1): 3-21.

32. Blais C, Roy C, Fiset D, Arguin M, and F Gosselin (2012) The eyes are not the window to basic emotions. *Neuropsychologia*. 50(12): 2830–2838.

33. Blow M, Dautenhahn K, Appleby A, Nehaniv CL, and DC Lee (2006) Perception of robot smiles and dimensions for human-robot interaction design. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp.469–474.

34. Boiger M, and B Mesquita (2012) The construction of emotion in interactions, relationships, and cultures. *Emotion Review*. 4(3): 221–229.

35. Bosse T, Pontier M, and J Treur (2010) A computational model based on Gross' emotion regulation theory. *Cognitive Systems Research*. 11(3): 211-230.

36. Boucenna S, Gaussier P, Andry P, and L Hafemeister (2010) Imitation as a communication tool for online facial expression learning and recognition. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.5323-5328.

37. Breazeal C (2003) Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies.* 59(1–2): 119– 155.

38. Breazeal C (2009) Role of expressive behaviour for robots that learn from people. *Philos Trans R Soc Lond B Biol Sci*. 364(1535):3 527–3538.

39. Bryson JJ and EAR Tanguy (2010) Simplifying the design of humanlike behaviour: emotions as durative dynamic state for action selection. *Int Journal of Synthetic Emotions*. 1(1): 30–50.

40. Calder AJ and AW Young (2005) Understanding the recognition of facial identity and facial expression. *Nat Rev Neurosci*. 6(8): 641–651.

41. Cañamero D (1997) Modeling motivations and emotions as a basis for intelligent behavior. *Proceedings of the 1st ACM International Conference on Autonomous Agents*, pp.148–155.

42. Canamero L (2005) Emotion understanding from the perspective of autonomous robots research. *Neural Networks*. 18(4): 445-455.

43. Canamero L and J Fredslund (2001) I show you how I like you - can you read it in my face? *IEEE Transactions on Systems Man and Cybernetics A: Systems and Humans*. 31(5): 454 –459.

44. Caporael LR (1997) The evolution of truly social cognition: The core configuration model. *Personality and Social Psychology Review*. 1(4): 276–298.

45. Chaminade T, Zecca M, Blakemore S-J, Takanishi A, Frith CD, Micera S, et al. (2010) Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. PLoS ONE 5(7): e11577.

46. Chang WC and S Šabanović (2014). Observational study of naturalistic interactions with the socially assistive robot PARO in a nursing home. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 294-299.

47. Chang WC and S Šabanović (2015). Interaction expands function: Social shaping of the therapeutic robot PARO in a nursing home. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, In Press.

48. Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*. 36(3): 181–204.

49. Cohn JF (2010) Advances in behavioral science using automated facial image analysis and synthesis. *IEEE Signal Process Magazine*. 27(6): 128–133.

50. Darwin C (1872) *The Expression of the Emotions in Man and Animals*. John Murray: London, UK.

51. Dautenhahn K (2007) Socially intelligent robots: dimensions of human-robot interaction. *Philos Trans R Soc Lond B Biol Sci*. 362(1480): 679–704.

52. Davies IR, Sowden PT, Jerrett DT, Jerrett T, and GG Corbett (1998) A cross-cultural study of English and Setswana speakers on a colour triads task : A test of the Sapir-Whorf hypothesis. *British Journal of Psychology*. 89(1): 1–15.

53. De Gelder B (2009) Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philos Trans R Soc Lond B Biol Sci*. 364(1535): 3475–3484.

54. De Jaegher H and EA Di Paolo (2007) Participatory sense-making: An enactive approach to social cognition. *Phenomenology in Cognitive Science*. 6(4): 485-507.

55. De Jaegher H, Di Paolo E, and S Gallagher (2010) Can social interaction constitute social cognition? *Trends in Cognitive Sciences*. 14(10): 441–447.

56. Delaherche E, Chetouani M, Mahdhaoui, A, Saint-Georges C, Viaux S, and D Cohen (2012) Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*. 3(3): 349-365.

57. Delaunay F, De Greeff J, and T Belpaeme (2009) Towards retro-projected robot faces: An alternative to mechatronic and android faces. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp.306–311.

58. DiSalvo CF, Gemperle F, Forlizzi J, and S Kiesler (2002) All robots are not created equal: the design and perception of humanoid robot heads. *Proceedings of the 4th ACM Conference on Designing Interactive Systems*, pp.321–326.

59. Dolan RJ (2002) Emotion, cognition, and behavior. *Science*. 298(5596): 1191–1194.

60. DSM- IV (2000) Diagnostic and statistical manual-text revision (DSM-IV-TRim, 2000). *American Psychiatric Association*.

61. Dunbar K (2001) The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. *The Analogical Mind: Perspectives from Cognitive Science*, pp.313-334.

62. Ekman P (1971) Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*. 19: 207–283.

63. Ekman P (2009) Darwin's contributions to our understanding of emotional expressions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 364(1535): 3449–3451.

64. Ekman P, Friesen WV (2003) *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Malor Books: Los Altos, CA, USA.

65. Elfenbein HA (2013) Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*. 5(1): 90–96.

66. Embgen S, Luber M, Becker-Asano C, Ragni M, Evers V, and KO Arras (2012) Robot-specific social cues in emotional body language. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp.1019–1025.

67. Esau N, Kleinjohann B, Kleinjohann L, and D Stichling (2003) MEXI: machine with emotionally extended intelligence. In: Abraham A, Köppen M, Franke K (eds) *Design and Application of Hybrid Intelligent Systems*. IOS Press, Amsterdam, pp.961–70.

68. Ezer N, Fisk AD, and WA Rogers (2009) More than a servant: Self-reported willingness of younger and older adults to having a robot perform interactive and critical tasks in the home. *Proceedings of the Human Factors and Ergonomics Society Annual Meetin*g, 53(2): 136-140.

69. Faber F, Bennewitz M, Eppner C, Gorog A, Gonsior C, Joho D, et al. (2009) The humanoid museum tour guide Robotinho. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp.891–896.

70. Fonagy P (2000) Attachment and borderline personality disorder. *Journal of the American Psychoanalytic Association*. 48(4): 1129-1146.

71. Fong T, Nourbakhsh I, and K Dautenhahn (2003) A survey of socially interactive robots. *Rob Auton Syst*. 42(3–4): 143–66.

72. Friesen WV (1973) Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules. *Dissertation Abstracts International*. 33(8-B): 3976-3977.

73. Froese T and T Ziemke (2009) Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*. 173(3-4): 466–500.

74. Froese T and EA Di Paolo (2011) The enactive approach: theoretical sketches from cell to society. *Pragmatics & Cognition*. 19(1): 1-36.

75. Fuchs T (2007) Fragmented selves: Temporality and identity in borderline personality disorder. *Psychopathology*. 40(6): 379-387.

76. Fugate, JMB (2013) Categorical Perception for Emotional Faces. *Emotion Review*. 5(1): 84-89.

77. Gadanho SC and J Hallam (2001) Robot learning driven by emotions. *Adaptive Behavior*. 9(1): 42–64.

78. Geraci R M (2006) Spiritual robots: Religion and our scientific view of the natural world. *Theology and Science*. 4(3): 229-246.

79. Gibson JJ (1979) *The Ecological Approach to Visual Perception*. Houghlin Mifflin: Boston, MA, USA.

80. Gockley R, Bruce A, Forlizzi J, Michalowski M, Mundell A, Rosenthal S, et al. (2005). Designing robots for long-term social interaction. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.1338-1343.

81. Grant BF, Chou SP, Goldstein RB, Huang B, Stinson FS, Saha TD, et al. (2008). Prevalence, Correlates, Disability, and Comorbidity of DSM-IV Borderline Personality Disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry*. 69(4): 533–545.

82. Gratch J, Rickel J, Andre E, Cassell J, Petajan E, and N Badler (2002) Creating interactive virtual humans: some assembly required. *Intelligent Systems*. 17(4): 54-63.

83. Gross JJ and RW Levenson (1995) Emotion elicitation using films. *Cognition and Emotion*. 9(1): 87–108.

84. Hall ET (1977) *Beyond Culture*. Anchor Books: New York, NY, USA.

85. Hall ET (1966) *The Hidden Dimension*. Anchor Books: New York, NY, USA

86. Henson RN, Goshen-Gottstein Y, Ganel T, Otten LJ, Quayle A, and MD Rugg (2003) Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cereb Cortex*. 13(7): 793-805.

87. Hermans D, De Houwer J, and A Eelen (1994) The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion*. 8(6): 515-533.

88. Huang CM and B Mutlu (2012) Robot behavior toolkit: generating effective social behaviors for robots. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 25-32.

89. Ikegami T and K Suzuki (2008) From a homeostatic to a homeodynamic self. *BioSystems*. 91(2): 388–400.

90. Ishiguro H (2005) Android science — toward a new cross-interdisciplinary framework. *ICCS/CogSci Workshop: Toward Social Mechanisms of Android Science*, pp.1–6.

91. Jack RE, Blais C, Scheepers C, Schyns PG, and R Caldara (2009) Cultural confusions show that facial expressions are not universal. *Curr Biol*. 19(18): 1543-1548.

92. Jack RE, Garrod OG, Yu H, Caldara R, and PG Schyns (2012) Facial expressions of emotion are not culturally universal. *Proceeedings of the National Acadamy of Sciences USA*. 109(19): 7241–7244.

93. Kahn PH Jr, Freier NG, Kanda T, Ishiguro H, Ruckert JH, Severson RL, and SK Kane (2008) Design patterns for sociality in human-robot interaction. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.97-104.

94. Kahn PH Jr, Ruckert JH, Kanda T, Ishiguro H, Reichert A, Gary H, and S Shen (2010a) Psychological intimacy with robots: Using interaction patterns to uncover depth of relation. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.123-124

95. Kahn PH Jr, Gill BT, Reichert AL, Kanda T, Ishiguro H, and JH Ruckert (2010b) Validating interaction patterns in HRI. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.183-184.

96. Kahn PH Jr, Reichert AL, Gary HE, Kanda T, Ishiguro H, Shen S, et al. (2011) The new ontological category hypothesis in human-robot interaction. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.159-160

97. Kahn PH Jr, Kanda T, Ishiguro H, Gill BT, Ruckert JH, Shen S, Gary HE, et al. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 33-40.
98. Kanda T, Shiomi M, Miyashita Z, Ishiguro H, and N Hagita (2009). An affective guide robot in a shopping mall. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 173-180.
99. Kaplan F (2004) Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*. 1(3): 465-480.
100. Kelso JAS (2012) Multistability and metastability: understanding dynamic coordination in the brain. *Philos Trans R Soc Lond B Biol Sci*. 367(1591): 906–918.
101. Kendon A (1990). *Conducting Interaction*. Oxford University Press: New York, NY, USA.
102. Kirby R, Forlizzi J, and R Simmons (2010) Affective social robots. *Rob Auton Syst*. 58(3): 322–332.
103. Kitayama S, Mesquita B, and M Karasawa (2006) Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology*. 91(5): 890–903.
104. Kleinsmith A, De Silva PR, and N Bianchi-Berthouze (2006) Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*. 18(6): 1371–89.
105. Ko SG, Lee TH, Yoon HY, Kwon JH, and M Mather (2011) How does context affect assessments of facial emotion? The role of culture and age. Psychological Aging. 26(1): 48–59.
106. Koda T, Ruttkay Z, Nakagawa Y, and K Tabuchi (2010) Cross-cultural study on facial regions as cues to recognize emotions of virtual agents. In: *Lecture Notes in Computer Science (Culture and Computing)*, T. Ishida, Ed. Springer: Berlin, pp.16–27.
107. Kozima H, Michalowski M., and C Nakagawa (2009) Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics*. 1(1): 3–18.

108. Kret ME and B de Gelder (2010) Social context influences recognition of bodily expressions. *Experimental Brain Research*. 203(1): 169–180.

109. Kret ME, Stekelenburg JJ, Roelofs K, and B de Gelder (2013) Perception of face and body expressions using electromyography, pupillometry, and gaze measures. *Frontiers in Psychology*. 4: 28.

110. Krumhuber EG, Kappas A, and ASR Manstead (2013) Effects of dynamic aspects of facial expressions: A review. *Emotion Review*. 5(1): 41–46.

111. Kwon DS, Kwak D, Keun Y, Park JC, Chung MJ, Jee ES, et al. (2007) Emotion interaction system for a service robot. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp.351-356.

112. Land MF and DE Nilsson (2012) *Animal Eyes*. Oxford University Press: Oxford, UK.

113. Lee MK, Kiesler S, Forlizzi J, Srinivasa S, and P Rybski (2010) Gracefully mitigating breakdowns in robotic services. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.203-210.

114. Leite I, Castellano G, Pereira A, Martinho C, and A Paiva (2012) Modeling empathic behavior in a robotic game companion for children: An ethnographic study in real-world settings. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.367-374.

115. Lee TH, Choi JS, and YS Cho (2012) Context modulation of facial emotion perception differed by individual difference. *PLoS ONE*. 7(3): e32987.

116. Lee HR and S Sabanović (2014) Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.17–24.

117. Lienhart R and J Maydt (2002) An extended set of haar-like features for rapid object detection. *IEEE International Conference on In Image Processing (ICIP)*, pp. I-900.

118. Lenzenweger MF, Lane MC, Loranger AW, and RC Kessler (2007) DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biological Psychiatry*. 62(6): 553-564.

119. Li D, Rau PP, and Y Li (2010) A cross-cultural study: effect of robot appearance and task. *International Journal of Social Robotics*. 2(2): 175-186.

120. Lindquist KA and M Gendron (2013) What's in a word? Language constructs emotion perception. *Emotion Review*. 5(1): 66–71.

121. Littman ML (2009) A tutorial on partially observable Markov decision processes. *Journal of Mathematical Psychology*. 53(3): 119–125.

122. MacDorman KF & H Ishiguro (2006) The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*. 7(3): 297–337.

123. MacDorman KF, Vasudevan SK, and CC Ho (2009) Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society*. 23(4): 485-510.

124. MacDorman KF, Green RD, Ho C-C, and CT Koch (2009) Too real for comfort? Uncanny responses to computer generated faces. *Comput Hum Behav*. 25(3): 695–710.

125. Mason PT and R Kreger (2010) *Stop Walking on Eggshell*. New Harbinger Publications: Oakland, CA, USA.

126. Masterson JF (1988) *The Search for the Real Self: Unmasking the Personality Disorders of our Age*. Taylor & Francis: New York, NY, USA.

127. Masuda T and RE Nisbett (2001) Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*. 81(5): 922-934.

128. Masuda T, Ellsworth PC, Mesquita B, Leu J, Tanida S, and E Van de Veerdonk (2008) Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*. 94(3): 365–381.

129. Matsumoto D (1992) American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*. 23(1): 72–84.

130. Matsumoto N, Fujii H, and M Okada (2006) Minimal design for human-agent communication. *Artificial Life and Robotics*. 10(1): 49–54.

131. Mayer C, Sosnowski S, Kuhnlenz K, and B Radig (2010) Towards robotic facial mimicry: system development and evaluation. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 198–203.

132. McGann M, De Jaegher H, and E Di Paolo (2013) Enaction and psychology. *Review of General Psychology*. 17(2): 203–209.

133. Merleau-Ponty M (1945) *Phénoménologie de la Perception*. Gallimard: Paris.

134. Michalowski MP, Sabanovic S, and H Kozima (2007) A dancing robot for rhythmic social interaction. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.89–96.

135. Mitra S and T Acharya (2007) Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. 37(3): 311-324.

136. Mori M (1970) Bukimi no tani [The uncanny valley]. *Energy*. 7(4): 33–35. http://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley

137. Movellan JR, Tanaka F, Fortenberry B, and K Aisaka (2005) The RUBI/QRIO project: Origins, principles, and first steps. *4th IEEE International Conference on Development and Learning (ICDL)*, pp.80-86.

138. Mutlu B and J Forlizzi (2008) Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.287-294.

139. Neal DT and W Wood (2009) Automaticity in situ and in the lab: The nature of habit in daily life. *Oxford Handbook of Human Action*, pp. 442-457.

140. Nelson NL and JA Russell (2013) Universality revisited. *Emotion Review*. 5(1): 8-15.

141. Nisbett RE (2003) *The Geography of Thought: How Asians and Westerners Think Differently*. Free Press: New York, NY, USA.

142. Nisbett RE, Peng K, Choi I, and A Norenzayan (2001) Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*. 108(2): 291–310.

143. Nomura T and T Kanda (2003). On proposing the concept of robot anxiety and considering measurement of it. *Proceedings of IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*, pp.373–378.

144. Nomura T, Suzuki T, Kanda T, and K Kato (2006) Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3): 437–454.

145. Nourbakhsh IR, Kunz C, and T Willeke (2003) The mobot museum robot installations: a five year experiment. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.3683-3641.

146. Nourbakhsh IR, Hamner E, Porter E, Dunlavey B, Ayoob E, Hsiu T, et al. (2005) The design of a highly reliable robot for unmediated museum interaction. *IEEE International Conference on Robotics and Automation (ICRA)*, pp.3225-3231.

147. Ogino M, Watanabe A, and M Asada (2008) Detection and categorization of facial image through the interaction with caregiver. *7th IEEE International Conference on Development and Learning (ICDL)*, pp.244-249.

148. Pantic M (2009) Machine analysis of facial behaviour: naturalistic and dynamic behaviour. *Philos Trans R Soc Lond B Biol Sci*. 364(1535): 3505–3513.

149. Pantic M and MS Bartlett (2007). Machine analysis of facial expressions. In: *Face Recognition*. I-Tech Education and Publishing: Vienna, Austria, pp.377–416.

150. Peltason J and B Wrede (2010) Modeling human-robot interaction based on generic interaction patterns. *AAAI Fall Symposium: Dialog with Robots*.

151. Perlovsky L (2009) Language and emotions: emotional Sapir-Whorf hypothesis. *Neural Networks*. 22(5-6): 518–526.

152. Peterson MF and MP Eckstein (2012) Looking just below the eyes is optimal across face recognition tasks. *Proceeedings of the National Acadamy of Sciences USA*, 109(48): e3314–3323.

153. Pfeiffer R and J Bongard (2007) *How the Body Shapes the Way We Think*. MIT Press: Boston, MA, USA.

154. Picard RW (1997) *Affective Computing*. MIT Press: Boston, MA, USA.

155. Pierno AC, Mari M, Lusher D, and U Castiello (2008) Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*. 46(2): 448-454.

156. Pollack ME, Brown L, Colbry D, Orosz C, Peintner B, Ramakrishnan S, et al. (2002) Pearl: A mobile robotic assistant for the elderly. *AAAI Workshop on Automation as Eldercare*, pp.85-91.

157. Powers A and S Kiesler (2006) The advisor robot: tracing people's mental model from a robot's physical attributes. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.218–225.

158. Reeves B and C Nass (1996) *The Media Equation: How People Treat Computers, Television and New Media like Real People and Places*. Cambridge University Press: New York, NY, USA.

159. Rehm M, Bee N, Endrass B, Wissner M, and E André (2007) Too close for comfort? Adapting to the user's cultural background. *ACM Proceedings of the International Workshop on Human-Centered Multimedia*, pp.85-94

160. Righart R and B de Gelder (2008) Recognition of facial expressions is influenced by emotional scene gist. *Cognitive Affective & Behavavioral Neuroscience*. 8(3): 264–272.

161. Robinson P, and R El Kaliouby (2009) Computation of emotions in man and machines. *Philos Trans R Soc Lond B Biol Sci*. 364(1535): 3441–3447.

162. Ruiz-del-Solar J, Mascaró M, Correa M, Bernuy F, Riquelme R, and R Verschae (2009) Analyzing the human-robot interaction abilities of a general-purpose social robot in different naturalistic environments. In: *RoboCup 2009: Robot Soccer World Cup XIII*, pp.308-319.

163. Russell, JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bulletin*. 115(1): 102-141.

164. Russell JA and JM Fernández-Dols (1997). *The Psychology of Facial Expression*. Cambridge University Press: Cambridge, UK.

165. Russell S and P Norvig (2010) *Artificial Intelligence: A Modern Approach, 3rd Ed*. Prentice Hall: Upper Saddle River, NJ, USA.

166. Ruttkay Z (2009) Cultural dialects of real and synthetic emotional facial expressions. *AI & Society*. 24(3): 307–315.

167. Šabanović S (2010) Emotion in robot cultures: Cultural models of affect in social robot design. *Proceedings of the Conference on Design & Emotion (D&E2010)*, Chicago, IL.

168. Šabanović S (2014) Inventing Japan's "robotics culture": The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science*. In Press. Available online: http://sss.sagepub.com/content/early/2014/01/17/0306312713509704.abstract

169. Šabanović S, Michalowski MP, and R Simmons (2006) Robots in the wild: Observing human-robot social interaction outside the lab. *IEEE International Workshop on Advanced Motion Control*, pp.596-601.

170. Šabanović S, Bennett CC, and H Lee (2014) Towards culturally robust robots: A critical social perspective on robotics and culture. *Proceedings of the ACM/IEEE HRI Workshop on Culture Aware Robotics (CARS)*. In Press.

171. Šabanović S, Bennett CC, Piatt JA, Chang W, Hakken D, Kang S, and D Ayer (2014) Participatory design of socially assistive robots for preventive patient-centered healthcare. *IEEE/RSJ Assistive Robotics Workshop at International Conference on Intelligent Robots and Systems (IROS)*. In Press.

172. Šabanović S, Reeder S, and B Kechavarzi (2014) Designing robots in the wild: In situ prototype evaluation for a break management robot. *Journal of Human-Robot Interaction*. 3(1): 70-88.

173. Saldien J, Goris K, Vanderborght B, Vanderfaeillie J, and D Lefeber (2010) Expressing emotions with the social robot Probo. *International Journal of Social Robotics*. 2(4): 377–389.

174. Saint-Aimé S, Le Pévédic B, Duhaut D (2009) First evaluation of EMI model of interaction. *Proceedings of the 14th IASTED International Conference on Robotics and Applications*, pp.263–270.

175. Saint-Aime S, Le-Pevedic B, Duhaut D, and T Shibata (2007) EmotiRob: Companion robot project. *Proceedings of IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*, pp.919–924.

176. Scassellati B (2000) How developmental psychology and robotics complement each other. *Technical Report - Massachusetts Institute of Technology*. CSAIL, Cambridge, Massachusetts.

177. Scassellati B, Admoni H, and M Matarić (2012) Robots for use in autism research. *Annu Rev Biomed Eng*. 14: 275–94.

178. Scheeff M, Pinto J, Rahardja K, Snibbe S, and R Tow (2002) Experiences with Sparky, a social robot. In: Dautenhahn K, Bond A, Cañamero L, Edmonds B (eds) *Socially Intelligent Agents*. Springer, USA, pp.173–180.

179. Schermerhorn P, Scheutz M, and CR Crowell (2008) Robot social presence and gender: do females view robots differently than males? *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.263-270.

180. Schiano DJ, Ehrlich SM, Rahardja K, and K Sheridan (2000) Face to interface: facial affect in (hu)man and machine. *SIGCHI Conference on Human Factors in Computing Systems*, pp.193–200.

181. Shore B (1996) *Culture in Mind: Cognition, Culture, and the Problem of* Meaning. Oxford University Press: Oxford, UK.

182. Sidner CL, Lee C, Morency LP, and C Forlines (2006) The effect of head-nod recognition in human-robot conversation. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.290-296.

183. Sidner CL and C Lee (2007) Attentional gestures in dialogues between people and robots. In: Nishida T (ed) *Conversational Informatics: An Engineering Approach*. John Wiley & Sons, West Sussex, UK, pp.103–115.

184. Sosnowski S, Bittermann A, Kuhnlenz K, and M Buss (2006) Design and evaluation of emotion-display EDDIE. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.3113–3118.

185. Sterelny K (2007) Social intelligence, human intelligence and niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 362(1480): 719-730.

186. Straub I, Nishio S, and H Ishiguro (2010) Incorporated identity in interaction with a teleoperated android robot: A case study. *Proceedings of IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*, pp. 119-124.

187. Suchman L (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press: Cambridge, UK.

188. Tanaka F, Cicourel A, and JR Movellan (2007) Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the USA*, 104(46): 17954-17958.

189. Thrun S, Bennewitz M, Burgard W, Cremers AB, Dellaert F, Fox D, et al. (1999) Experiences with two deployed interactive tour-guide robots. *Proceedings of the International Conference on Field and Service Robotics (FSR'99)*.

190. Thrun S, Maren B, Burgard W, Cremers AB, Dellaert F, Dieter F, et al. (1999) MINERVA: A tour-guide robot that learns. *Proceedings of the 23rd Annual German Conference on Artificial Intelligence*, pp.14-26.

191. Trovato G, Kishi T, Endo N, Zecca M, Hashimoto K, and A Takanishi (2013) Cross-cultural perspectives on emotion expressive humanoid robotic head: Recognition of facial expressions and symbols. *International Journal of Social Robotics*. 5(4): 515-527.

192. Turkle S (2005) Relational artifacts/children/elders: The complexities of cybercompanions. P*roceedings of Toward Social Mechanisms of Android Science*, pp. 62-73.

193. Turkel S (2011) *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books: New York, NY, USA.

194. Van Breemen A, Yan X, and B Meerbeek (2005) iCat: an animated user-interface robot with personality. *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*, pp.143–144.

195. Warren WH (2006) The dynamics of perception and action. *Psychological Review*. 113(2): 358–389.

196. Walters ML, Oskoei MA, Syrdal DS, and Dautenhahn (2011) A long-term human-robot proxemic study. *Proceedings of IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*, pp.137-142.

197. Weiss A, Igelsböck J, Tscheligi M , Bauer A , Kuhnlenz K, Wollherr D, and M Buss (2010) Robots asking for directions – The willingness of passers-by to support robots. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.23-30.

198. Yamazaki K, Yamazaki A, Okada M, Kuno Y, Kobayashi YH, Pitsch K, et al. (2009) Revealing Gauguin: Engaging visitors in robot guide's explanation in an art museum. *ACM Conference on Human Factors in Computing Systems (CHI)*, pp.1437-1446.

199. Yamazaki A, Yamazaki K, Ohyama T, Kobayashi Y, and Y Kuno (2012) A techno-sociological solution for designing a museum guide robot: Rregarding choosing an appropriate visitor. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.309-316.

200. Yoshikawa M, Matsumoto Y, Sumitani M, and H Ishiguro (2011) Development of an android robot for psychological support in medical and welfare fields. *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp.2378–2383.

201. Yuki M, Maddux WW, and T Masuda (2007) Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *J Exp Soc Psychol*. 43: 303-311.

202. Zhang J and AJ Sharkey (2011) Contextual recognition of robot emotions. In: Groß R, Alboul L, Melhuish C, Witkowski M, Prescott TJ, Penders J (eds) *Towards Autonomous Robotic Systems*. Springer, Berlin, pp.78–89.

11. Curriculum Vitae

2. BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Bennett, Casey	Senior Research Fellow
eRA COMMONS USER NAME	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Western Kentucky University, Bowling Green, KY	B.A.	2003	Anthropology
Indiana University, Bloomington IN	M.A.	2005	Biological Anthropology
Indiana University, Bloomington IN	Ph.D.	2015(exp.)	Biomedical Informatics/Artificial Intelligence

A. Personal Statement

Casey Bennett, MA, is a research fellow in the Dept. of Informatics at Centerstone Research Institute (CRI), and part of the School of Informatics and Computing at Indiana University. His work focuses on artificial intelligence and biomedical informatics, including the areas of robotics, human-robot interaction, data warehousing, machine learning, clinical decision support, and personalized medicine. He was the lead designer for Centerstone's award-winning organization-wide data warehouse and analytics platform in healthcare (2010 TDWI Best Practices Award). He also served as the lead data architect for CRI's federal grant projects – providing database design and technical guidance for studies funded by AHRQ, SAMHSA, NIH, and the CDC – and the PI on studies related to the implementation of patient-response outcome systems across Indiana and Tennessee. His work has been featured as part of IBM's "Smarter Planet" campaign.

He is currently working on projects utilizing robots for therapeutic purposes in assisted living facilities, exploring the minimal features for robotic faces that can display human-like facial expressions, multi-agent simulation modeling of diabetes, and artificial intelligence for simulating clinical decision-making in chronic illness.

Technical skills include **Databases** (SQL, Oracle, Postgres, MySQL, MS Access, OLAP, ODBC/JDBC, Toad, SQL Plus, ETL Design and Implementation, Kettle), **Programming** (Perl, C/C++, Python, Visual Basic, PHP, HTML, XML, XSLT), **Robotics** (Arduino, ROS, OpenCV), **Business Intelligence** (Jasper Reporting Engine, Pentaho Reporting Engine, Crystal Reports, Qlikview, iReport, Mondrian), **Platforms**

(Windows, Mac, Linux/Unix), **Statistics** (SPSS, SAS, R statistical software), **Genetic Software** (Mega2, Phylip, DNAsp), **Data Mining** (Weka, Knime, Clementine, RapidMiner, BayesiaNet, C4.5), **Networks** (LAN's, VPN, Remote Drives, FTP file servers, Apache, SSH), **GIS** (ArcGIS, Quantum), and **Genetics** (Genetic sequence analysis, PCR (standard and real-time), Gel electrophoresis, primer design, microarrays, population genetics, phylogenetics)

B. Positions and Honors

Positions and Employment

2003-2005	Database Designer, Indiana University - Indiana Molecular Biology Institute, Bloomington, IN
2004-2004	Computer Technical Support Staff, Kiva Networking, Bloomington, IN
2005-2006	Data Analyst , Inviva, Louisville, KY
2006-Present	Research Fellow/Data Architect, Dept. of Informatics - Centerstone Research Institute, Nashville, TN
2011-Present	Associate Instructor, Indiana University – School of Informatics and Computing (SOIC), Bloomington, IN

Other Experience, Honors & Professional Memberships

1999-2003	Award of Excellence Scholarship and Western Kentucky University
2003	*Magna Cum Laude*, Western Kentucky University
2004-2005	Fellow at the Center for the Study of Global Change, Indiana University
2006-Present	Member, The Data Warehousing Institute (TDWI)
2010	TDWI Best Practices Award, The Data Warehousing Institute, www.tdwi.org
2010	CARF "Exemplary" Status for Clinical Analytics, Commission on Accreditation of Rehabilitation Facilities, www.carf.org
2010-2012	Work featured in IBM's "Smarter Planet" Campaign
2010-Present	Member, American Medical Informatics Association (AMIA)
2013-Present	NSF IGERT Associate – Cognitive Science, Indiana University
2013	Summer Research Award – NSF IGERT Program, Cognitive Science, Indiana University
2014-Present	Member, Institute of Electrical and Electronics Engineers (IEEE)

C. Selected Peer-Reviewed Publications and Presentations

1. Bennett CC and TW Doub (2015) "Expert Systems in Mental Healthcare: AI Applications in Decision Making and Consultation." In: David D. Luxton (ed.) *Artificial Intelligence in Mental Healthcare*. Elsevier Press. In Press.

2. Bennett CC and S Sabanovic (2015) "The effects of culture and context on perceptions of robotic facial expressions." *Interaction Studies*. In Press.

3. Bennett CC, Sabanovic S, Fraune M, & K Shaw (2014) Context congruency and robotic facial expressions: Do effects on human perceptions vary across culture? *IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*. Edinburgh, Scotland. pp. 465-470.

4. Sabanovic S, Bennett CC, Piatt JA, et al. (2014) "Participatory design of socially assistive robots for preventive patient-centered healthcare." *IEEE/RSJ Assistive Robotics Workshop at International Conference on Intelligent Robots and Systems (IROS)*. In Press.

5. Bennett CC and S Sabanovic (2014) "Deriving minimal features for human-like facial expressions in robotic faces." *International Journal of Social Robotics*. 6(3): 367-381.

6. Bennett CC and TW Doub (2014) "Temporal modeling in clinical artificial intelligence, decision-making, and cognitive computing: Empirical exploration of practical challenges." *Proceedings of the 3rd SIAM Workshop on Data Mining for Medicine and Healthcare (DMMH)*. Philadelphia, PA, USA.

7. Sabanovic S, Bennett CC, and HR Lee (2014) "Towards culturally robust robots: A critical social perspective on robotics and culture." *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction (HRI) Workshop on Culture-Aware Robotics (CARS)*. Bielefeld, Germany.

8. Sabanovic S, Bennett CC, Chang WL, and L Huber (2013) "PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia." *Proceedings of the 13th International Conference on Rehabilitation Robotics (ICORR)*. Seattle, Washington. pp. 1-6. PMID: 24187245.

9. Bennett CC and S Sabanovic (2013) "Perceptions of affective expression in a minimalist robotic face." *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction (HRI)*. Tokyo, Japan. pp. 81-82.

10. Bennett CC and K Hauser (2013) "Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach." *Artificial Intelligence in Medicine*. 57(1): 9-19. PMID: 23287490

11. Bennett CC (2012) "Utilizing RxNorm to support practical computing applications: Capturing medication history in live electronic health records." *Journal of Biomedical Informatics*. 45(4): 634-641. PMID: 22426081

12. Bennett CC, Doub TW, and R Selove (2012) "EHRs connect research and practice: Where predictive modeling, artificial intelligence, and clinical decision support intersect." *Health Policy and Technology*. 1(2): 105-114.

13. Bennett, CC, Doub, TW, Bragg, AD, et al. (2011) "Data mining session-based patient reported outcomes (PROs) in a mental health setting: Toward data-driven clinical decision support and personalized treatment." *Proceedings of the IEEE Health Informatics and Systems Biology Conference*. pp. 229-236.

14. Bennett, CC (2011) "Clinical productivity system: A decision support model." *International Journal of Productivity and Performance Management*. 60(3): 311-319.

15. Bennett, CC and TW Doub (2010) "Data mining and electronic health records: Selecting optimal clinical treatments in practice." *Proceedings of the 6th International Conference on Data Mining*. pp. 313-318.

D. Research Support

Ongoing Research Support

#1343940 (Hauser/Grannis) 1/1/2014-12/31/2016
National Science Foundation – Smart and Connected Health
Intelligent Clinical Decision Support with Probabilistic and Temporal EHR Modeling
This project is a collaboration between Indiana University, Regenstrief Institute, Marshfield Clinic, and Centerstone Research Institute, exploring temporal modeling of clinical data for use in decision support systems (CDSS).
Role: Key Personnel

Completed Research Support

#1143712 (Sabanovic) 8/1/2011-7/31/2014
National Science Foundation - Division of Information & Intelligent Systems
EAGER: Cultural models in social robotics - Comparative studies with users in the US and Japan
This project evaluates the use of robots (Paro) for therapeutic purposes in assisted living facilities, while exploring the differences in how users perceive, make sense of, and interact with social robots across cultures.
Role: Investigator/Data Analysis

TI018870 (Blakely/Hardy) 9/30/2007-9/29/2013
SAMHSA – Center for Substance Abuse Treatment
Targeted Capacity – Co-Occurring Disorders Treatment and HIV/AIDS Services
This project is expanding and enhancing access to integrated dual disorders treatment for individuals who are released from prisons and jails who are abusing substance and at-risk for HIV/AIDS.
Role: Database Implementation

SM057010 (Outlaw) 9/30/2005-9/29/2011
SAMHSA: Center for Mental Health Services
Mule Town Family Network
This project formalizes the infrastructure to plan, implement, and evaluate wraparound services that respond to the needs of children and youth (birth to 21 years) with serious emotional disturbances and their families.
Role: Lead Data Architect

(Trivedi, Daly, Doub) 10/01/2007 – 9/30/2010
Agency for Healthcare Research and Quality
Using Information Technology to Provide Measurement Based Care for Chronic Illness
This project is testing implementation of measurement based care (MBC) in an ambulatory care setting with an integrated clinical decision support system (CDSS) and an electronic health record (EHR-CDSS). The proposal focuses on the use of MBC to improve the quality of care for patients with major depressive disorder (MDD).
Role: Data Architect

(Bennett) 1/01/2010-12/31/2010
Funding: Ayers Foundation; Centerstone of Tennessee, Centerstone of Indiana
Practice-Based Evidence Outcomes Pilot Study – CDOI
Implementation and analysis of the effects of a client-directed clinical outcome measure in Tennessee and Indiana.
Role: Principal Investigator

(Bennett) 6/01/2008-6/01/2009
Funding: Ayers Foundation
Clinical Productivity System – A Decision Support Model (2009) – Designed and implemented a clinical productivity system designed on a decision support model. Increased revenues by 30%, treatment plan completion 25%, case management eligibility 20%, clinical percentage 10%, as well as improvements in compliance issues and outcomes collections.
Role: Principal Investigator

(Bennett) 1/01/2008-12/31/2008
Funding: Centerstone of Tennessee
Case Management – Interactive Clinician Tools (2008) –Tools and supporting infrastructure to improve consistency of clinical care in the case management domain.
Role: Lead Data Architect

TI17232 (Outlaw) 8/15/2005-8/14/2008
SAMHSA – Center for Substance Abuse Treatment
TCE Rural Populations: Methamphetamine Evidence-based Treatment & Healing (METH)
This program targeted adults ages 18+ who are abusing methamphetamine and other emerging drugs in six rural counties. Utilizing the Matrix Model, support services (outreach, assessment, case management),

and community education, the Rural METH Initiative expanded access to structured, culturally competent care for a diverse rural population of individuals using stimulants.
Role: Lead Data Architect

SM-56910 (Doub, Moran) 10/01/2004 –9/30/2007
SAMHSA – Center for Mental Health Services
Implementation of the IMPACT Model to Treat Depression in Older Adults
This project evaluates the effectiveness of the IMPACT model, for mental health outreach, treatment, and prevention services in a primary care setting for older adults in Davidson and Williamson Counties.
Role: Lead Data Architect