

NEURAL CORRELATES OF TRUSTWORTHINESS EVALUATIONS IN
CROSS-CULTURAL INTERACTIONS

by

Leanne R. Young

APPROVED BY SUPERVISORY COMMITTEE:

Daniel C. Krawczyk, Chair

James C. Bartlett

Alice J. O'Toole

Amy E. Pinkham

Copyright 2016
Leanne R. Young
All Rights Reserved

NEURAL CORRELATES OF TRUSTWORTHINESS EVALUATIONS IN
CROSS-CULTURAL INTERACTIONS

by

LEANNE R. YOUNG, BS, MA

DISSERTATION

Presented to the Faculty of
The University of Texas at Dallas
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY IN
COGNITION AND NEUROSCIENCE

THE UNIVERSITY OF TEXAS AT DALLAS

December 2016

ACKNOWLEDGMENTS

This research was funded under a contract with the Defense Advanced Research Projects Agency (DARPA). I would like to thank the program manager, Dr. Doug Weber, and his assistant, Gretchen Knaack, for their support. Thank you also to my advisor, Dr. Daniel Krawczyk, for his assistance and support throughout this study. It has been a pleasure to work with and learn from him. Thank you also to my other committee members, Dr. Alice O'Toole, Dr. James Bartlett, and Dr. Amy Pinkham, whose contributions to the study design and insights on the analytical methods were invaluable.

Ms. Amanda Hahn is owed a particular debt of gratitude for adapting her multivariate pattern analysis software to this project. It is a testimony to the quality and thoroughness of her work that a relative Matlab novice could easily read, understand and make further modifications to the code. This research could not have been completed as quickly or as smoothly without the tireless efforts of Tiffany Jantz, Jelena Rakic, and Jameson Miller in recruiting, scheduling and scanning participants. Ms. Linda Nguyen's expertise with the Implicit Association Test and her assistance in collecting, editing and evaluating the stimuli videos were also invaluable, and I look forward to seeing her continue and expand upon the research started herein. Thank you to the staff of the Advanced Imaging Research Center for their professionalism.

Finally, I would like to thank Dr. Geoff Ling, who first suggested the idea of using a novel approach to advance the field of social neuroscience. His enthusiasm was both inspirational and contagious.

August 2016

NEURAL CORRELATES OF TRUSTWORTHINESS EVALUATIONS IN
CROSS-CULTURAL INTERACTIONS

Publication No. _____

Leanne R. Young, PhD
The University of Texas at Dallas, 2016

Supervising Professor: Daniel C. Krawczyk, PhD

Studies researching the neural correlates of implicit appearance-based trust evaluations have implicated a number of brain regions, including the amygdala, prefrontal cortices, posterior superior temporal sulcus and the fusiform gyrus. In cross-cultural studies of explicit trust judgments, the anterior cingulate cortex has been implicated, and in cross-cultural studies of emotion recognition, the cuneus has been implicated. In this study, I applied multivariate pattern analysis using a linear discriminator to demonstrate that the neural activation patterns in these regions in response to passively viewing videos of trustworthy and untrustworthy-appearing East Asian and Caucasian men are dissociable. I show that the dissociation of brain patterns in response to trustworthy and untrustworthy stimuli is not dependent upon magnitudes of activation, and it is similar for both East Asian and Caucasian participants.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF ABBREVIATIONS	xi
CHAPTER 1 INTRODUCTION	
1.1 Background.....	1
1.2 Objectives	9
CHAPTER 2 METHODS	11
2.1 Participants	11
2.2 Experimental Procedure	13
2.3 Analytics	18
CHAPTER 3 RESULTS	22
3.1 Cultural Bias Implicit Association Test.....	22
3.2 Trust Discrimination Based upon Neural Activations.....	22
3.3 Comparison of MVPA and Explicit Trust Predictions.....	26
3.4 Dependence of Trust Discrimination on Magnitudes of Activation.....	28
3.5 Race Discrimination Based upon Neural Activations	28
CHAPTER 4 DISCUSSION.....	33
4.1 Dissociating on Trust.....	33
4.2 Comparison of East Asian and Caucasian Discriminability of Trustworthy Stimuli	35
4.3 Discriminability with and without Magnitudes of Activation.....	35
4.4 Comparison of MVPA and Explicit Judgment Discriminability.....	36
4.5 Race Dissociation	37

4.6	Limitations and Recommendations for Future Studies	38
CHAPTER 5 CONCLUSION		40
APPENDIX A CROSS-CULTURAL QUESTIONNAIRE		42
REFERENCES		45
BIOGRAPHICAL SKETCH		52
CURRICULUM VITAE		

LIST OF FIGURES

Figure 1. Functional MRI 1-Back Task Design.....	16
Figure 2. Sample screen shots from the Cultural Bias IAT. Participants use keys to sort pictures and words into categories listed in the upper left and right corners of the screen.....	18
Figure 3. Discriminability of Trustworthy and Untrustworthy–Appearing Stimuli based upon neural patterns of activation in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).	23
Figure 4. Discriminability of Trustworthy and Untrustworthy-Appearing Stimuli based upon neural patterns of activation during the first two TRs of stimulus presentation in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).....	25
Figure 5. Discriminability of Trustworthy and Untrustworthy –Appearing Stimuli based upon neural patterns of activation during the last two TRs of stimulus presentation in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).....	25
Figure 6. Comparison of explicit judgments by participants in this study and the ratings used to select stimulus videos for use in the study. (The x-axis labels indicate a video identification; error bars represent +/- one standard error.).....	26
Figure 7. Comparison of Discriminability of Explicit Judgments and Discriminability in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC). A “*” indicates that the d’ score for that region is statistically significantly greater or less than that of the d’ for the explicit judgments.	27
Figure 8. Mean discriminability between trustworthy and untrustworthy-appearing stimuli with magnitudes of activation and with pattern information only, in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC),	

fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).	29
Figure 9. Discriminability of East Asian and Caucasian Stimuli based upon neural patterns of activation in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).	30
Figure 10. Discriminability of East Asian and Caucasian Stimuli based upon neural patterns of activation during the first two TRs of stimulus presentation in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).	30
Figure 11. Discriminability of East Asian and Caucasian Stimuli based upon neural patterns of activation during the last two TRs of stimulus presentation in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).	31
Figure 12. Mean discriminability between East Asian and Caucasian stimuli with magnitudes of activation and with pattern information only across all four TRs in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).	32

LIST OF TABLES

Table 1. Study Participants	12
Table 2. Cross-cultural Experience.....	12
Table 3. Regions of Interest Coordinates.....	21
Table 4. Correlation (Pearson's rho) of d's in the regions of interest.....	24
Table 5. Correlation of discriminability (d') based upon neural correlations with discriminability of post-scan explicit trust judgments	28

LIST OF ABBREVIATIONS

ACC – Anterior Cingulate Cortex

AMG – Amygdala

BOLD – Blood Oxygenation Level Dependent

CUN – Cuneus

d' – discriminability

D – IAT effect

DLPFC – Dorsolateral Prefrontal Cortex

FFA – Fusiform Face Area

fMRI – Functional Magnetic Resonance Imaging

GLM – General Linear Modeling

IAT – Implicit Association Test

LDA – Linear Discriminator Analysis

M – Mean

mm – millimeter

MNI – Montreal Neurological Institute

MRI – Magnetic Resonance Imaging

ms – milliseconds

MVPA – Multivariate Pattern Analysis

OFA – Occipital Face Area

PCA – Principal Components Analysis

pSTS – Posterior Superior Temporal Sulcus

rTMS – Repetitive Transcranial Magnetic Stimulation

SD – Standard Deviation

SPM – Statistical Parametric Mapping

TE – Echo Time

TR – Repetition Time

VLPFC – Ventrolateral Prefrontal Cortex

WFU – Wake Forest University

CHAPTER 1

INTRODUCTION

Cross-cultural social interactions are an increasingly prevalent aspect of modern academic, business, military and political life. In social interactions, our perceptions and evaluations of faces and facial expressions are critical for recognizing identity, age, ethnicity and gender, and making inferences about the mental states, emotional states, and intentions of others (Said, Haxby, & Todorov, 2011). Faces are also used to infer personality traits, such as dominance, competence, trustworthiness, intelligence and likeability (Fiske, Cuddy, & Glick, 2007; Santos & Young, 2010; Todorov, Mandisodza, Goren, & Hall, 2005; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Zebrowitz & Montepare, 2015). Although most studies have shown that these trait inferences are not correlated with actual personality traits, they are ubiquitous, consistent and automatic (Olivola & Todorov, 2010; Santos & Young, 2010).

One of the most studied traits inferred from faces is trustworthiness. This study applied multivariate pattern analysis (MVPA) to investigate the underlying neural activations associated with passively viewing trustworthy and untrustworthy same and other-race stimuli. Using dynamic, naturalistic stimuli, I explore how neural activations in several brain regions dissociate trustworthy from untrustworthy-appearing and East Asian from Caucasian stimuli.

1.1 Background

1.1.1 Neural Correlates of Face-Based Trustworthiness Evaluations

Face-based trustworthiness evaluations are thought to arise, at least in part, from resemblance of neutral faces to faces expressing emotions, with positive emotions connoting trustworthiness and negative emotions connoting untrustworthiness (Stirrat & Perrett, 2010; Todorov, 2008). Neural studies of face-based trust evaluations reinforce this concept, with many of the brain regions identified in emotional expression also identified in association with face-

based trustworthiness evaluations, including the amygdala and the core face perception system. In addition, various studies have identified other cortical structures, such as the ventrolateral prefrontal cortex, the superior temporal cortex, the frontopolar cortex, the inferior and middle frontal gyri and both the inferior and superior parietal lobules with increased Blood Oxygenation Level Dependent (BOLD) signals associated with either very trustworthy or very untrustworthy faces.

Amygdala. The importance of the amygdala to evaluating trust is well-established (Adolphs, Tranel, & Damasio, 1998; Engell, Haxby, & Todorov, 2007; Engell, Todorov, & Haxby, 2010; Todorov & Duchaine, 2008; Todorov & Engell, 2008). Although early research suggested a linear response in the amygdala, with increased activation associated with highly untrustworthy faces, most current research suggests a quadratic response, with increased activation associated with both highly trustworthy and highly untrustworthy faces. The quadratic response in the amygdala is consistent with two separate, but compatible hypotheses regarding why trustworthiness evaluations are made from facial stimuli, particularly considering that these evaluations are not necessarily accurate (Olivola & Todorov, 2010). The “overgeneralization” theory proposes that trait evaluations result from overgeneralizing perceptions of facial expressions signaling whether to approach or avoid a person (Montepare & Dobish, 2003; Zebrowitz & Montepare, 2005). To explore this hypothesis, Oosterhof and Todorov (2008) used happy-to-angry judgments on a 9-point scale by 19 participants viewing 72 computer-generated neutral expression faces that varied along the trustworthiness dimension. The findings of this study revealed that trustworthiness evaluations are, indeed, sensitive to features resembling emotional expressions. Alternatively, Valentine (1991) proposed the “multi-dimensional face space model,” which suggests that the response in the amygdala is a function of geometric distance from an average face (Mattavelli, Andrews, Asghar, Towler, & Young, 2012). Thus, as face dimensions become increasingly atypical, the amygdala responds with a stronger activation. In both computer-generated faces (Oosterhof & Todorov, 2008; Said et al., 2011) and in the morphed real faces used by Mattavelli et al. (2012), typicality and trait evaluations are confounded. In the context of cross-racial trustworthiness evaluations, this confounding is likely to be significant, since other-race faces are usually perceived as atypical.

Core Face Processing System. Various studies have also identified the core face perception system, including the occipital and fusiform face areas and the posterior superior temporal sulcus. In fMRI studies, the fusiform face area (FFA) and, more generally, the fusiform gyri, were found to have bilaterally greater responses to untrustworthy and trustworthy faces (Said et al., 2011; Winston, Strange, O'Doherty, & Dolan, 2002). In an fMRI study, Winston et al. (2002) found the right pSTS to be recruited for *explicit* judgments of trustworthiness, with larger activations associated with untrustworthy faces. In a repetitive transcranial magnetic stimulation (rTMS) study, Dzhelyova, Ellison, and Atkinson (2011) explored the impact of transient lesions in the occipital face area (OFA) and posterior superior temporal sulcus (pSTS) on the ability of participants to make gender and trustworthiness judgments. Although the pSTS did not appear to have a role in gender discrimination, rTMS delivered over either the left or right pSTS increased the reaction times associated with making trustworthiness judgments for male faces, but not female faces. Overall, reaction times and accuracies were lower for judging gender and trustworthiness for untrustworthy female faces and, in fact, it was only for longer exposure times that rTMS over the OFA had an impact of increasing reaction times for trustworthiness evaluations. This finding suggests that when trustworthiness evaluations are difficult, there may be additional recruitment of the OFA.

Other Cortical Regions. Although the vast majority of the literature related to trustworthiness evaluations on neutral faces has focused on the amygdala, there have been a few functional neuroimaging studies that have looked at other brain regions supporting trait evaluations. Using functional neuroimaging, Winston et al. (2002) identified regions associated with both implicit and explicit trustworthiness judgments. Whether the task was to judge trustworthiness (explicit task) or judge age (implicit task), increased BOLD signals were associated with trustworthy faces in the dorsolateral prefrontal cortex and the frontopolar cortex. In a functional neuroimaging study for patients with schizophrenia and autism spectrum disorders Pinkham, Hopfinger, Pelphrey, Piven, and Penn (2008) observed activation bilaterally in the amygdala, posterior superior temporal sulcus, and ventrolateral prefrontal cortex (VLPFC) in both healthy controls and patient populations during a task requiring that participants rate faces as either trustworthy or untrustworthy. This finding was consistent with findings by

Cunningham, Johnson, Gatenby, Gore, and Banaji (2003) that showed an increase in activation in the VLPFC during judgments involving complex evaluative information.

In addition to the dorsolateral and ventrolateral prefrontal cortices and the frontopolar cortex, researchers have found evidence of both linear and quadratic responses to faces varying in trustworthiness in the inferior and middle frontal gyri, the superior temporal cortex, and both the inferior and superior parietal lobules (Cunningham et al., 2003; Dzhelyova et al., 2011; Pinkham et al., 2008; Said et al., 2011; Winston et al., 2002).

Although the central role of the amygdala is apparent, many of the other regions have been identified in just one or two trust evaluation studies so far, or the regions of interest have been very large, lacking localization of trust evaluations therein. Furthermore, studies attempting to characterize the response in these regions as either linear or quadratic have yielded inconsistent results (Mattavelli et al., 2012; Rule, Krendl, Ivcevic, & Ambady, 2013; Said et al., 2011; Winston et al., 2002). Thus, not only does the neural network for trustworthiness evaluations remain uncertain, but also the nature of the response in key regions is not yet understood.

1.1.2 Cross-Race Face-based Trustworthiness Evaluations

Comparatively few studies have attempted to characterize the neural correlates of trustworthiness judgments in a cross-cultural or cross-race context. These studies have identified regions such as the amygdala, anterior cingulate cortex (ACC), striatum, VLPFC and the DLPFC in association with trustworthiness judgments in a cross-cultural context. These regions are familiar from research into the neural correlates of cross-race face evaluations, emotional expression recognition and within-race trustworthiness evaluations (Cunningham et al., 2004; Said et al., 2011). Cross-cultural studies of emotional expression recognition have shown that we are generally better at recognizing emotions in same-race faces than other face faces, and doing so involves more activation in the fusiform face area and cuneus (Chiao et al., 2008; Derntl et al., 2009). Given the close relationship between emotional expression recognition and trustworthiness trait evaluations, it is plausible that enhanced activation in these regions is also associated with cross-race trustworthiness evaluations.

Behaviorally, studies with American and Asian (Walker, Jiang, Vetter, & Sczesny, 2011), Korean (McArthur & Berry, 1987; Zebrowitz, Montepare, & Lee, 1993), Japanese (Rule et al., 2010), Chinese (Albright et al., 1997) and the isolated Bolivian Tsimane people (Zebrowitz et al., 2012) have found strong agreement on judgments of traits of dominance, strength, warmth and honesty. With the exception of the Tsimane people, there was also cross-cultural agreement on attractiveness and babyfacedness. Based upon these results, we might expect same and cross-cultural trustworthiness evaluation networks to be similar. Although there is some degree of universality in trustworthiness, attractiveness and babyfacedness evaluations, some cross-cultural variability is expected, as well (Zebrowitz et al., 2012). For example, because experience / exposure to faces at an early age impacts the formation of a “typical” or “average” face, and because attractiveness is typically associated with faces that are near to this “typical” face, some cultural variability is likely on attractiveness. Furthermore, because attractiveness tends to create a “halo” effect, causing attractive faces to be perceived as having more positive traits, differences in valuation of traits across cultures may be impacted by how the halo effect influences trait evaluations. In fact, differences in the valuation of attractiveness across cultures may impact the strength of the halo effect (Zebrowitz et al., 2012). Finally, studies have shown with a variety of populations and ingroup/outgroup distinctions that information regarding and outgroup’s traits will bias the perceived facial appearance toward having facial features corresponding with the bias (Dotsch, Wigboldus, & van Knippenberg, 2011; Dotsch, Wigboldus, & Knippenberg, 2013; Ratner, Dotsch, Wigboldus, van Knippenberg, & Amodio, 2014). Thus, racial or cultural bias can alter perception.

Underlying most studies of cross-cultural trustworthiness is the assumption that all cultures engage in automatic trait evaluations and that these are relatively consistent, albeit with some cultural variability. However, Na and Kitayama (2011) conducted a study in which participants were primed with a previously studied facial photo implying a personality trait, then presented a word or a pseudo-word, and asked to indicate whether the word was an English word. When European Americans processed an antonym of the implied personality trait in the priming picture, they elicited an N400 electrophysiological signal, indicating semantic conflict. Asian Americans did not elicit this signal. Although the Asian Americans provided explicit trait

evaluations when instructed to do so, Na and Kitayama (2011) concluded from this study that Asian Americans do not make automatic trait evaluations. Na and Kitayama hypothesize that the absence of automatic trait evaluations may be related to interdependent self-construal in Asian cultures, which emphasizes situational constraints over internal motivations for understanding and interpreting behaviors. Whether this hypothesis proves to be correct or not, Na and Kitayama's research highlights not only the importance of expanding the current body of cross-cultural trustworthiness research to a broader range of cultures and races, but also acquiring a better understanding of the neural processes that underlie trustworthiness evaluations.

In sum, behavioral studies suggest that the basic cognitive processes associated with trustworthiness evaluations are likely to be similar under both same and cross-cultural conditions. But, differences in facial structure, attractiveness and typicality standards and the presence of social norms and cognitive biases may influence the outcome of trustworthiness evaluations. The impact of these differences is not well understood, but a few studies have begun to look at specific neural responses in the amygdala, striatum, ACC, VLPFC and DLPFC.

1.1.3 Outstanding Questions in Trustworthiness Evaluations

Existing face-based trustworthiness evaluation research has been mostly limited to review of static images of unfamiliar faces, presented in the absence of cranial or facial hair, glasses or other sources of variation in appearance, and in the absence of any external knowledge or experience with the individuals in the images. However, a few studies have shown that physical features, such as facial hair, the absence or presence of cranial hair, and eyeglasses have an impact on perceived trustworthiness, with beards and rimless glasses, in particular, increasing perceived trustworthiness (Bakmazian, 2014; Hellström & Tekle, 1994; Leder, Forster, & Gerger, 2011; Muscarella & Cunningham, 1996; Wogalter & Hosie, 1991).

Regarding the use of static images, a number of studies have shown that recognition of emotional expressions is enhanced with the use of dynamic stimuli, and leads to recruitment of additional brain regions, including the premotor and supplemental motor areas, the striate and extrastriate (Kilts, Egan, Gideon, Ely, & Hoffman, 2003; Recio, Sommer, & Schacht, 2011; Sato, Kochiyama, Yoshikawa, Naito, & Matsumura, 2004; Trautmann, Fehr, & Herrmann, 2009). Given the close relationship between emotional expression recognition and face-based

trustworthiness evaluations it is, perhaps, not surprising that facial dynamics have a significant effect on behaviors in a trust game, and factors such as smile onset duration and head tilt impact trustworthiness ratings (Krumhuber, Manstead, & Kappas, 2007; Krumhuber, Manstead, Cosker, et al., 2007). The neural underpinnings of trustworthiness evaluations conducted on stimuli that are in motion are as yet unknown, but are likely to resemble those identified in dynamic emotional expression studies.

Perhaps the least studied aspect of trustworthiness trait evaluations is the importance of “person” knowledge. Anecdotally, we know that trust perceptions can be strongly influenced by previous interactions or observations of an individual, or information about an individual’s reputation or group associations. Researchers studying cooperation and trust using ultimatum/bargaining games have shown that priming participants about their counterpart with descriptions such as “friend,” “opponent,” “partner,” or “foe” influences trust behaviors (Burnham, McCabe, & Smith, 2000). Similarly, “cheap talk” (bragging, threatening) prior to an ultimatum bargaining game influences behaviors during the game (Croson, Boles, & Murnighan, 2003). Gobbini and Haxby (2007) hypothesized that regions such as the anterior paracingulate cortex and the posterior superior temporal sulcus that are associated with “theory of mind” processing are responsible for the automatic and spontaneous encoding of “person” knowledge, such as traits, intentions, and mental states. Likewise, the anterior temporal cortex is associated with biographical knowledge, and the precuneus and posterior cingulate are associated with episodic memories. Together, these regions provide feedback to the core face perception system, such that recognition is more about “person” perception than mere “face” perception. The concept behind “person” perception is that it involves not only incoming sensory information, but also top-down processes that provide supplementary information about that person (Quadflieg et al., 2011). In the absence of personal knowledge, stereotypes can provide the supplemental “person knowledge.” However, in an fMRI study Todorov, Gobbini, Evans, and Haxby (2007) demonstrated that briefly associating unfamiliar faces with verbal descriptions of behaviors (aggressive, disgusting, neutral and nice) is sufficient to result in increased activity in the anterior paracingulate cortex and posterior superior temporal sulcus. Given that trustworthiness trait evaluations outside the laboratory will often involve at least some minimal

basis for “person” knowledge, it is plausible that trustworthiness evaluations in real life will involve regions such as the anterior paracingulate cortex, anterior temporal cortex, posterior superior temporal sulcus, precuneus and posterior cingulate.

Thus far, studies of trustworthiness evaluations, both same race and cross-cultural, have clearly identified the amygdala as a key region of activation. Most studies show a quadratic response in the amygdala, with increased activation in response to highly trustworthy or highly untrustworthy faces. However, since most of these studies have used computer-generated neutral faces and static images, the response of the amygdala to more naturalistic stimuli, such as a video of a person talking and showing a range of facial expressions, is unknown.

Other regions that are likely to have some role in neural networks supporting trust include the core face processing regions, the ventrolateral prefrontal cortex, and the insula. Evidence exists for both quadratic and linear responses in these regions, again based upon static images. The response in these regions to an individual talking and showing a range of facial expressions is unknown. However, studies have shown improved emotional recognition with dynamic stimuli, which suggests that there could be an increased sensitivity to variations in features influencing trustworthiness evaluations, as well. Given that emotional expression recognition is better for same-race faces and appears to involve more activation in the face processing regions and the cuneus, it is plausible that cross-race trait evaluations might do so, as well.

Although trait evaluations are known to be automatic and implicit, much of the trustworthiness research has depended upon either explicit judgments or, in the case of trust games, decisions or actions based upon implicit judgments. In both cases, it is difficult to dissociate the automatic response from cognitive processes that moderate that response. This issue is likely to be exacerbated in the context of cross-cultural evaluations, where behavioral and neural studies have shown that amygdala activation in response to other race faces correlates with implicit, but not explicit, measures of racial biases. In cross-cultural trustworthiness evaluations, the anterior cingulate cortex, the striatum, the ventrolateral prefrontal cortex and the dorsolateral prefrontal cortex also are likely to contribute. However, not only is the linearity of the responses in these regions unknown, but the extent to which the responses are automatic and implicit is unclear.

1.2 Objectives

The objective of this study is to use multivariate pattern analysis (MVPA) to elucidate the neural network that underlies implicit, appearance-based trustworthiness trait evaluations in a cross-cultural context. Although a number of brain regions have been identified as being likely to have a role in trustworthiness assessments, it would be an exaggeration to say that a neural network has been identified for trustworthiness evaluations, particularly under cross-cultural conditions.

There is currently no evidence that the neural network supporting appearance-based trust evaluations differs for trustworthy versus untrustworthy stimuli. However, there is evidence that brain regions such as the ventrolateral and dorsolateral prefrontal cortices and fusiform gyrus respond linearly to variations in appearance of trustworthiness. Although the evidence of linear responses remains controversial, these observations suggest a hypothesis that it may, in fact, be possible to dissociate the brain activations for trustworthy and untrustworthy stimuli, but that magnitudes of activation would be required for such dissociation.

For this study, I selected a subset of regions, based upon the literature, within which I anticipate the ability to discriminate trustworthy from untrustworthy stimuli based upon magnitudes and patterns of activation. The amygdala was selected, largely because a strong response in the amygdala to highly trustworthy and untrustworthy stimuli is consistently noted throughout the trust literature. In addition, the amygdala is often observed to have a heightened response to other-race faces. Thus, in the context of a cross-cultural study on trust evaluations, a dual role for the amygdala is anticipated. Likewise, as key components of the face processing system, the fusiform gyrus and posterior superior temporal sulcus were selected because these regions are frequently identified in both studies of trustworthiness evaluations and studies of cross-race faces. The anterior cingulate cortex has been identified in cross-race studies of emotional expression and *explicit* trust evaluations. It was included in this study because its role in discriminating trustworthy from untrustworthy stimuli in an *implicit* context has not been verified. Similarly, dorsolateral and ventrolateral prefrontal cortices have been implicated in a number of trust studies, but are usually associated with a level of conscious processing not required for implicit evaluations. These regions were included in this study because the role of

these regions when stimuli are viewed passively, but for a relatively long period of time, has not been verified. Finally, the cuneus was included in this study because it has been identified in cross-race studies of emotion recognition, and emotion recognition and trust are closely related. Additionally, as part of the *basic* visual system, a role for the cuneus in cross-race trust evaluations would be somewhat surprising and, therefore, interesting.

A second hypothesis of this study is that dissociation of neural responses to trustworthy and untrustworthy stimuli will be similar across races. This hypothesis is supported by the universality of trait evaluations observed in behavioral studies (Walker, Jiang, Vetter, & Sczesny, 2011; McArthur & Berry, 1987; Zebrowitz, Montepare, & Lee, 1993; Rule et al., 2010; Albright et al., 1997; Zebrowitz et al., 2012).

Although the focus of this study was on discrimination of trustworthy from untrustworthy-appearing stimuli, the use of both East Asian and Caucasian stimuli and participants offered an opportunity to explore the discriminability of race based upon the neural patterns of activation in the anterior cingulate cortex, amygdala, cuneus, dorsolateral and ventrolateral prefrontal cortices, fusiform gyrus and superior temporal sulcus. Based upon previous cross-race neural studies, particularly those involving faces varying in appearance of trustworthiness, these regions are likely to be discriminatory for both Caucasians and East Asians.

CHAPTER 2

METHODS

2.1 Participants

A total of 50 participants were recruited, including 25 Euro-Caucasian and 25 East Asian healthy, English-speaking males between 18 and 35 years of age. None of the participants had formal training in cross-cultural interactions. Participants were recruited using flyers and social media distributed on The University of Texas at Dallas campus and the Dallas community. Exclusion criteria for this study were:

- History of psychological or neurological condition – The purpose of this study was to understand how a *healthy* brain processes social information to make trait evaluations in same-race and cross-cultural contexts.
- Standard exclusion criteria per MRI safety protocols, such as presence of medical device implants or objects, or a history of claustrophobia.
- Lack of fluency in English – Participants viewed full videos including auditory content in English. If an individual is not fluent in English, he may not be able to understand the content of the videos.
- Participants other than Euro-Caucasian and East Asian males were excluded so as to ensure statistically significant results with the proposed sample size. To understand cross-cultural effects with more races and across genders, the sample size would have to be significantly increased.

Out of the initial 50 participants, one Euro-Caucasian and three East Asian participants were excluded from the study due to excessive motion in the scanner. Three other East Asian participants were excluded from the study due to software failures during scans. Table 1 provides some basic statistics on the remaining participants. In addition to collecting basic demographics, we also collected information about each participant's cross-cultural experience (Table 2).

Among the Caucasians, none spoke an East Asian language, and only three had visited East Asian countries. Most reported that they socialized with East Asian “sometimes” or “often,” and about half indicated that they had one or more East Asians that are “close” friends. None of them had East Asians in their immediate family.

Table 1. Study Participants

DEMOGRAPHICS	ALL	CAUCASIAN	EAST ASIAN
Number of Participants	43	24	19
Race			
East Asian			17
East Asian / Hispanic			1
Slavic		1	
White / Caucasian		21	1*
White / Caucasian / Hispanic		1	
White / Caucasian / Native American		1	
Age Mean (Standard Deviation)	23.1 (3.0)	21.9 (2.3)	24.1 (3.1)
* Participant registered as East Asian and was of Taiwanese descent, but indicated White/Caucasian in the demographics questionnaire.			

Table 2. Cross-cultural Experience

DEMOGRAPHICS	ALL	CAUCASIAN	EAST ASIAN
Country where born			
United States	33	23	10
Canada	1	0	1
Poland	1	1	0
China	3	0	3
Malaysia	1	0	1
South Korea	3	0	3
Taiwan	1	0	1
Citizenship			
United States	39	23	16
United States and Poland	1	1	0
China	2	0	2
South Korea	1	0	1
Country where attended High School			
United States	41	24	17
China	2	0	2
Native Language			
English	29	23	6

DEMOGRAPHICS	ALL	CAUCASIAN	EAST ASIAN
English and another European language	1	1	0
English and another East Asian language	5		5
Chinese	5		5
Korean, Vietnamese, Taiwanese	3		3
Number of parents born in the US			
None	17	1	16
One	4	2	2
Two or more	22	21	1
Number of grandparents born in the US			
None	17	1	16
One	1	1	0
Two	5	3	2
Three	0	0	0
Four or More	19	19	1

Most of the East Asian participants were first generation Americans (neither parent was born in the United States). Among the East Asians, 13 indicated that they “mostly” or “always” watch English/Euro television and movies, while only 4 indicated that they “mostly” or “always” watch Asian television and movies. Seven indicated that they socialize with Euro-Americans “mostly” or “always,” and only two indicated that they “rarely” or “never” socialize with Euro-Americans. Only four of the East Asians had Euro-Americans in their immediate or extended family. One East Asian participant stood out among the rest, indicating that most of his close friends are Euro-American, he has two Euro-Americans in his immediate family, and there are over 50 Euro-Americans in his extended family. Another participant registered as an East Asian (Taiwanese), yet self-identified as White/Caucasian in the demographics survey.

2.2 Experimental Procedure

Following the consent briefing, each participant was provided training on the 1-back task that would be used in the scanner. Each participant was then placed in a mock scanner to view 16 videos, averaging 1.25 minutes in length, of individuals describing movies that they either very much liked or disliked. Upon completion of the video set, participants were placed in a Magnetic Resonance Imaging (MRI) scanner, where they performed three 1-back tasks in an event related

design. The stimuli in the 1-back tasks were 8 second video clips (no audio) from the previously viewed videos. After completing the three 1-back tasks, they underwent a structural MRI. Following scanning, participants rated the trustworthiness of the stimuli from the short videos, completed an assessment of implicit cultural biases (Greenwald, McGhee, & Schwartz, 1998), and responded to a demographics and cross-cultural experience questionnaire (Appendix A). Details regarding each aspect of the experimental procedure follow.

2.2.1 Stimuli

Sixteen videos were used as the stimuli in this study. The videos averaged 1.25 minutes in length and included both video and audio content. The videos were selected from a set of videos of 30 East Asian men and 30 Caucasian men from 18 to 35 years of age describing movies that they either really liked or disliked. The appearance of the individuals in the videos was rated on a visual analog scale from “not at all trustworthy” to “extremely trustworthy” using a group of 50 raters of mixed ages, races and genders, recruited from the University of Texas at Dallas student population. None of the participants in this study were involved in the initial rating of the videos. The rating procedure was similar to that used in other appearance-based trait evaluation studies (Todorov, 2008). However, because I was using videos, rather than static images, in the scanner, we deviated from the standard methodology by basing the ratings on short video clips, rather than static images. The clips used for rating were the first 15 seconds of videos of the individuals describing a movie.

The video set used for imaging included four videos each of East Asians who had been previously rated as appearing trustworthy, East Asians who had been previously rated as appearing untrustworthy, Caucasians who had been previously rated as appearing the most trustworthy and Caucasians who had been previously rated as appearing the most untrustworthy. This study was conducted in conjunction with a study on deception detection. Therefore, the individuals in half of the videos from each group were describing the movies in a way consistent with their true opinion – if they liked the movie, they described how much they liked it, and if they disliked the movie, they described how much they disliked it. The other half of the individuals described the movies in a way consistent with the *opposite* of their true opinion.

Participants viewing these videos were not informed that the individuals in the videos were selected based upon having appearances that had been previously rated as appearing either highly untrustworthy or highly trustworthy during the stimulus selection process. They were also not informed that some of the opinions expressed in the videos were inauthentic.

2.2.2 Mock Scanner

Prior to beginning the scanning, participants viewed the full set of the 16 full-length videos while positioned in a mock scanner. This allowed participants to hear what the individuals in the videos were saying, thus providing context for the clips viewed subsequently during scanning. This initial viewing in the mock scanner represented a compromise between the desire to have ecologically valid stimuli, which would include both visual and auditory content, the risk of poor sound quality in the loud MRI environment, and the desire to minimize extraneous neural activations in the auditory cortex. By viewing the full videos with auditory content, I sought to provide increased ecological validity and leverage the extended *person* perception system (Gobbini & Haxby, 2007). By doing so, I anticipated that the video views prior to scanning would lead to a stronger response to the stimuli during scanning. The risk associated with showing the full videos prior to scanning, rather than during scanning, was that the participants might not clearly recall the individuals in the videos, the video content, or impressions that might have been formed during the initial viewing. To minimize this risk, the participants viewed the videos in the mock scanner, immediately prior to scanning, thus providing a similar environment to the scanner for the initial view and minimizing the time between viewing the videos in the mock scanner and scanning.

2.2.3 Imaging

Neuroimaging was conducted at the Advanced Imaging Research Center at the University of Texas Southwestern (UTSW) Medical Center. Imaging was performed on a 3-Tesla scanner (Philips MR systems, Acheiva Release 2.5.3.0). Functional images were acquired with an echo-planar image sequence sensitive to the blood oxygenation level dependent (BOLD) contrast. The volume covered the whole brain with an 80 by 80 matrix and 38 transverse 4mm

thick slices with no gap, so that the voxel size was 3mm x3mmx4mm. The repetition time (TR) was 2.0 seconds, the echo time (TE) was 20 milliseconds (ms) and the flip angle was 70 degrees.

The fMRI experiment was a 1-back design consisting of three runs presented as an event-related design in which the participants viewed 8 second clips extracted from the videos previously viewed in the mock scanner (Figure 1). Each run used a different set of short video clips, and in each run the sequence of video clips was randomized. In each run each video clip was presented twice. Participants were asked to press a button when they observed the same clip in immediate succession. This 1-back task served to sustain participant attention and provide behavioral data to verify that attention was successfully maintained throughout the experiment. By performing multiple runs, each with different clips extracted from the full videos, we mitigated the risk that a particular clip viewed during scanning was in some way not representative of the entire video.

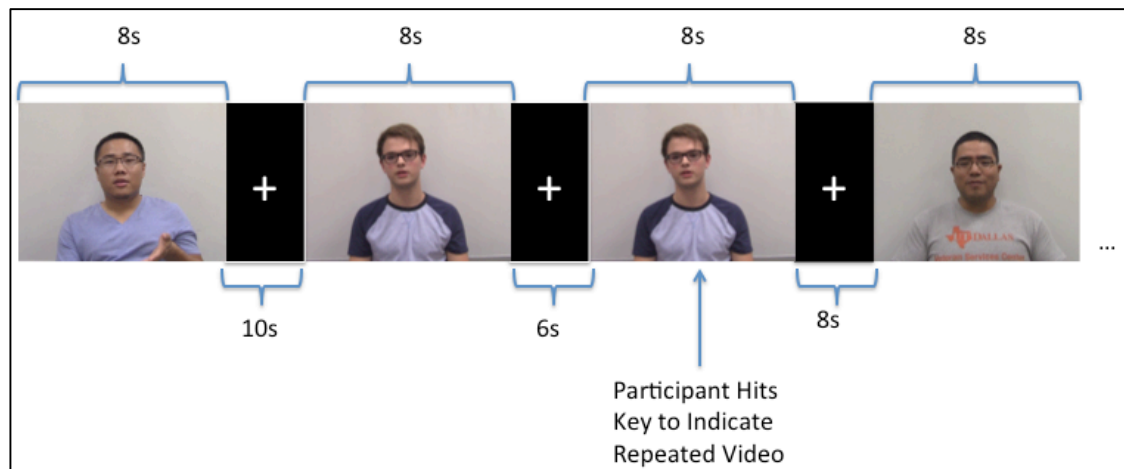


Figure 1. Functional MRI 1-Back Task Design

The clips were 8 seconds in length to provide sufficient time for engagement of top-down control processes, and a jittered interstimulus interval (ISI) of 6, 8 or 10 seconds was used between clips. A total of 269 volumes were collected in each run, and the total time for the three 1-back runs was just under 30 minutes.

Following the functional imaging, structural images of individual subjects were acquired to serve as template images onto which the functional images could be mapped. Whole brain structural scans included a T1-weighted spin-echo image sequence with 256 transverse slices and

a magnetization-prepared rapid access gradient-echo image sequence with 160 sagittal slices. The field of view was 204mmx256mmx160mm, the scan resolution was 256mmx256mm and the slice thickness was 1mm, with no gap.

2.2.4 Behavioral Data

Explicit Trust Evaluations

The ratings used for down-selecting to the videos set used in the scanner were based upon short video clips and, thus, contained rich dynamic visual content. However, because the participants in this study not only viewed the full-length videos, but also heard the audio content, there was some risk that the additional information would lead to differences in their evaluation of the trustworthiness of the stimuli,

relative to the original ratings. Therefore, after scanning was completed, participants were again presented with short clips, this time with audio content, extracted from the full-length videos, and asked to rate the individuals in clips on a visual analog scale from “not at all” trustworthy to “extremely” trustworthy. The clips used in soliciting these explicit judgments were all taken from the middle 8 seconds of the full-length videos. If these ratings were statistically significantly different from those obtained during the stimulus down-selection process, it would imply that the audio content of the full videos contributed to the participants’ trustworthiness evaluations.

Implicit Cultural Bias

To assess implicit cultural bias, I used the Implicit Association Test (IAT), one of the most widely used social science tools for implicit measurements (Greenwald et al., 1998; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). The IAT measures the response time of participants tasked with sorting stimuli into one of two categories or one of two attributes. Faster response times are associated with an implicit association of a category/attribute pair. Most commonly, the IAT is used to detect racial/cultural bias. Traditionally, the targets in these racial/cultural IATs are faces or partial faces of two different races/cultures, and the words have “Pleasant” and “Unpleasant” connotations (Greenwald et al., 1998; Nosek, Banaji, & Greenwald, 2002). In this study, my interest in racial or cultural bias is specific to how such a bias might

influence perceptions of trustworthiness. Thus, I used a variation on the traditional racial IAT in which the targets were East Asian and Euro-Caucasian faces and the attributes were words with “Trustworthy” (Truthful, Reliable, Honest, Honorable) and “Untrustworthy” connotations (Deceitful, Devious, Unreliable, Sneaky), rather than “Pleasant” and “Unpleasant” (Figure 2). To ensure that the associations identified by the IAT were racial/cultural, rather than reflecting a trait evaluation of a target face, the faces used as stimuli in the IAT were “neutral” face (rated between 2.43 and 2.57 on a 5-point trustworthiness scale) drawn from a database of East Asian and Caucasian female faces that had been previously rated on appearance of trustworthiness (Blanz & Vetter, 1999; Liang, 2015; Walker & Vetter, 2009).

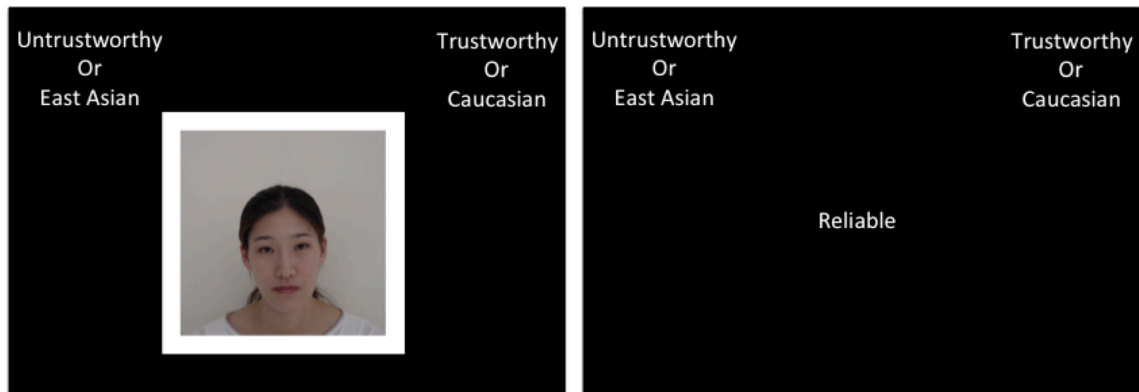


Figure 2. Sample screen shots from the Cultural Bias IAT. Participants use keys to sort pictures and words into categories listed in the upper left and right corners of the screen.

2.3 Analytics

2.3.1 Explicit Trust Ratings

The explicit trust ratings were compared to the ratings used to select the videos to be used in the scanner, using t-tests. In addition, each participant’s discriminability between trustworthy and untrustworthy videos was analyzed using signal detection theory. The discriminability measure (d') was correlated with the discriminability obtained from applying a pattern classifier to the imaging data.

2.3.2 Cultural Bias IAT

Implicit bias is defined in an implicit association test as an “IAT effect,” which is measured as a difference, called “ D ,” between response times in congruent and incongruent

conditions, normalized by the overall standard deviation. D scores were computed for each participant. A D score less than zero indicates a pro-Caucasian bias.

2.3.3 Imaging Data Analysis

Imaging data were time-slice corrected and realigned to the mean. The analysis was performed in native space on unsmoothed images. Because I observed that several participants were reasonably still *within* a run, but moved *between* runs, each run was preprocessed separately. All preprocessing was performed in Statistical Parametric Mapping (SPM) version 12.0 (Ashburner et al., 2014).

The fMRI data were analyzed using a combination of principal components analysis (PCA) and a Linear Discriminator Analysis (LDA) classifier. The purpose of the PCA is to reduce the parameter space, thus reducing the risk of overfitting. Using PCA, a set of weighted principal components can be obtained and used as input into the classifier. The goal of multivariate pattern analysis (MVPA) is to establish a statistical mapping between patterns of neural activity and associated cognitive processes. For neuroimaging analysis, this equates to being able to predict a condition variable (e.g., trustworthy face vs. untrustworthy face) from the pattern of BOLD signals recorded at the time the stimulus is present.

In general, MVPA uses a pattern classifier to learn the relationship between training data (patterns of neural activity) and condition variables (stimuli). To test the generalizability of the statistical mapping, cross-validation methods are used to score the algorithm with test data not used for the training. In this study, each of the runs for each participant was analyzed independently. Each run contained 16 stimuli, each presented twice. The BOLD signal in each voxel was averaged across the two presentations of a given stimulus. The pattern classifier was then run 16 times, each time removing a different trial as test data, and training the classifier on the remaining 15 trials. From each run, the success of the classifier was recorded as a hit, miss, false alarm or correct rejection. At the completion of the 16 classifications, these outcomes were used to compute the discriminability (d'), from Signal Detection Theory. For a given subject, with data from three 1-back runs there were three d' scores. These were averaged together to obtain an overall measure of how well the neural activations in that individual's brain discriminated between two conditions (e.g. trustworthy and untrustworthy-appearing). A d'

score near zero indicates poor discrimination; a d' score below zero indicates that the pattern classifier is discriminating between two conditions, but doing so incorrectly; a d' greater than zero indicates meaningful discrimination between two conditions. A negative d' in MVPA is typically an indication that the classifier is overfitting the data.

There were a variety of options as to which data could be provided to a classifier. Given that each stimulus presentation was for 8 seconds – equating to 4 brain volumes, or repetition times (TRs), the classifier could be run using data from just one brain volume (2 seconds of presentation), the average of two or more of the brain volumes, or two or more brain volumes, without averaging. Using just one brain volume has the disadvantage of excluding 3/4ths of the available data and, potentially, excluding information critical to dissociating brain patterns associated with the two conditions. Averaging across all brain volumes has the disadvantage that temporal data is lost in the averaging process. Thus, I used all four brain volumes in the classifier, without averaging. In this way, both spatial and temporal information contribute to the pattern analysis. However, I also ran the classifier using an average of just the first two volumes and an average of the last two volumes. By doing so, I could explore whether discrimination in a given region occurred primarily early or late during the stimulus presentation.

Masks of the regions of interest were created in Montreal Neurological Institute (MNI) space, then transformed into each subject's native space. The mask for the amygdala was anatomically defined using the WFU Pick Atlas (Maldjian, Laurienti, & Burdette, 2004; Maldjian, Laurienti, Kraft, & Burdette, 2003; Tzourio-Mazoyer et al., 2002). Because several centroids were identified for the ACC by Richeson et al. (2003), together covering a relatively large area, I created a region of interest composed of two 10mm diameter spheres, one located at 6, 18, 33, and the other at 12, 30, 24. All other masks were defined as 10mm diameter spheres with centroids based upon findings of previous trust studies (Table 3). Where the literature provided coordinates for only one side of the brain, symmetry was assumed to locate the region on the other side.

Table 3. Regions of Interest Coordinates

REGION OF INTEREST	LEFT COORDINATES	RIGHT COORDINATES	SOURCE
Ventral Fusiform Gyrus (Fusiform Face Area)	-48, -48, -24	44, -46, -22	Winston et al. (2002)
Ventrolateral Prefrontal Cortex	-42, 22, -9	42, 24, -20	Pinkham et al. (2008)
Dorsolateral Prefrontal Cortex		34, 52, 6	Winston et al. (2002)
Superior Temporal Sulcus	-55, -52, 7	52, -48, 10	Dzhelyova et al. (2011); Winston et al. (2002)
Cuneus	-14, -96, -6		Derntl et al. (2009)
Anterior Cingulate Cortex		6-12, 17-33, 20.5-22*	Richeson et al. (2003)

* This paper provided several centroids for regions of interest within the ACC.

CHAPTER 3

RESULTS

3.1 Cultural Bias Implicit Association Test

The mean D score for East Asian participants was -0.002 (SD=0.31), which is not statistically significantly different from zero ($p=.97$), indicating that the East Asian participants had no discernable cultural bias related to trustworthiness. The mean D score for Caucasian participants, on the other hand, was -0.23 (SD=0.25), which is statistically significantly below zero ($p<.0001$), indicating that the Caucasian participants had an implicit pro-Caucasian bias.

From the demographics survey, two participants stood out as potential outliers relative to their East Asian cohort. The first participant registered as East Asian (Taiwanese), but in the demographics survey responded as White/Caucasian. Additional information in the demographics survey indicated that he was, in fact, part East Asian and part Hispanic. Given that this individual's D score (0.006) was well within a single standard deviation of the East Asian D score mean, he was included in the East Asian cohort for the remaining analyses. The other individual had a large number of Euro-American close friends and family members. His D score (-0.84) was an outlier relative to both the East Asian and Caucasian cohorts, with the lowest D score across both data sets, indicating the greatest pro-Caucasian bias. Given that both his D score and his demographic responses indicate that this participant is an outlier, all analyses described herein were conducted both with this participant included as an East Asian and without this participant's data altogether. The results presented herein exclude this participant from the East Asian cohort.

3.2 Trust Discrimination Based upon Neural Activations

The d' (discriminability) estimates from the MVPA analysis in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform gyrus (FFA), posterior superior temporal sulcus and ventrolateral prefrontal cortex were statistically significantly

greater than zero, indicating that the neural patterns associated with viewing trustworthy and untrustworthy-appearing stimuli are dissociable in these regions (Figure 3). With the exception of the VLPFC and the cuneus, the mean d' estimates for East Asian and Caucasian participants were not statistically significantly different. In the VLPFC, the mean d' estimate for East Asians was larger than the d' estimate for Caucasians ($t(100)=1.85$, $p=.03$). In the cuneus, the mean d' estimate for the East Asians was larger than the d' estimate for the Caucasians ($t(109)=-1.74$, $p=.04$). The cuneus also stands out in an analysis of the correlation of the d 's in the seven regions tested (Table 4), where only the DLPFC and pSTS are significantly correlated with the cuneus.

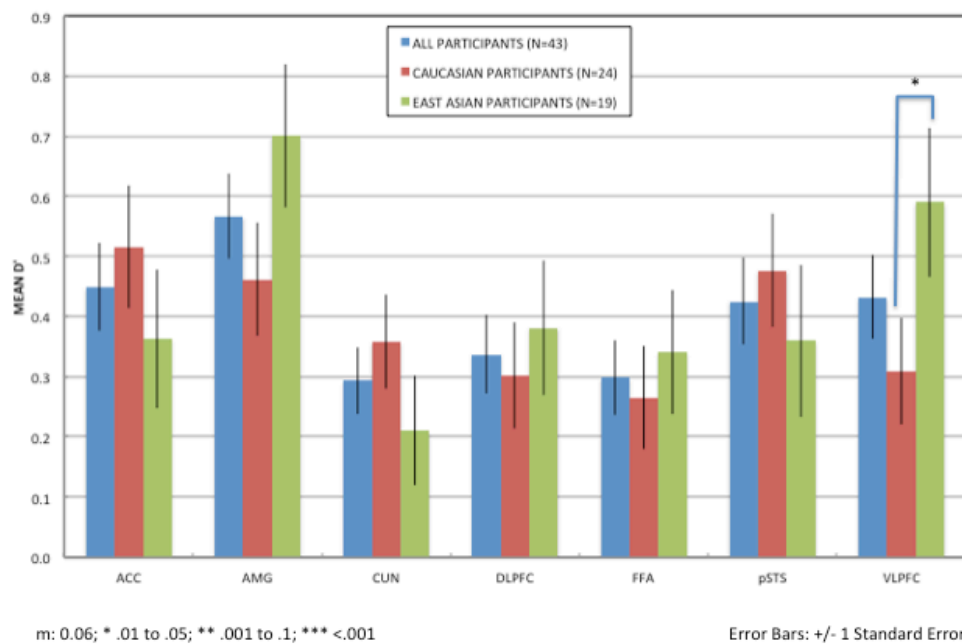


Figure 3. Discriminability of Trustworthy and Untrustworthy–Appearing Stimuli based upon neural patterns of activation in the anterior cingulate cortex (ACC), amygdala (AMG), cuneus (CUN), dorsolateral prefrontal cortex (DLPFC), fusiform gyri / fusiform face area (FFA), posterior superior temporal sulcus (pSTS) and ventrolateral prefrontal cortex (VLPFC).

From the previous analysis, we see that the spatial and temporal information contained across all four TRs (2 seconds per TR, a total of 8 seconds) of stimulus presentation supports discriminating between trustworthy and untrustworthy stimuli. But, from this analysis it is not clear whether information key to discriminating between the two conditions was primarily early in the stimulus presentation or later. To explore this question, the classifier was run on the average of just the first half of the stimulus presentation (first two TRs). In this analysis I found

that the d' for the cuneus was no longer statistically significantly greater than zero (Figure 4, All participants: $t(117)=0.89$, $p=.37$, 2-tailed; Caucasian participants: $t(67)=1.04$, $p=.30$, 2-tailed; East Asian participants: $t(49)=0.13$, $p=.89$, 2-tailed). Across all participants, the d' in the fusiform gyrus was greater than zero ($t(117)=2.09$, $p=.02$, 1-tailed), but at the significance level of .05, the d' was not greater than zero for either the East Asian or Caucasian cohorts (East Asian: $t(49)=1.52$, $p=0.07$, 1-tailed; Caucasian: $t(67)=1.43$, $p=.08$, 1-tailed).

Table 4. Correlation (Pearson's rho) of d 's in the regions of interest.

	ACC	AMG	CUN	DLPFC	FFA	pSTS	VLPFC
ACC		0.45 ($p<.0001$)	0.16 ($p=.09$)	0.36 ($p<.0001$)	0.36 ($p<.0001$)	0.47 ($p<.0001$)	0.49 ($p<.0001$)
AMG			0.15 ($p=.10$)	0.36 ($p<.0001$)	0.33 ($p=.0003$)	0.51 ($p<.0001$)	0.50 ($p<.0001$)
CUN				0.27 ($p=.002$)	-0.01 ($p=.95$)	0.30 ($p=.001$)	0.11 ($p=.21$)
DLPFC					0.32 ($p=.003$)	0.37 ($p<.0001$)	0.45 ($p<.0001$)
Fusiform Gyrus						0.16 ($p=.07$)	0.33 ($p=.0002$)
pSTS							0.37 ($p<.0001$)

When the classifier was run on the average of just the last half of the stimulus presentation (last two TRs, Figure 5), I found that the d 's in the cuneus were once again greater than zero (All participants: $t(117)=4.29$, $p<.0001$, 1-tailed; Caucasian participants: $t(67)=2.29$, $p=.01$, 1-tailed; East Asian participants: $t(49)=4.03$, $p<.0001$, 1-tailed). Although the discriminability in all other regions was similar (at the $\alpha=.05$ level), the d 's in the amygdala, cuneus and dorsolateral prefrontal cortices for the East Asian participants were numerically greater in the last two TRs than the same regions for the Caucasian participants, and the d' for the Caucasians was numerically greater in the posterior STS than that for the East Asians. None of these differences by race rose to a level of statistical significance ($\alpha = .05$).

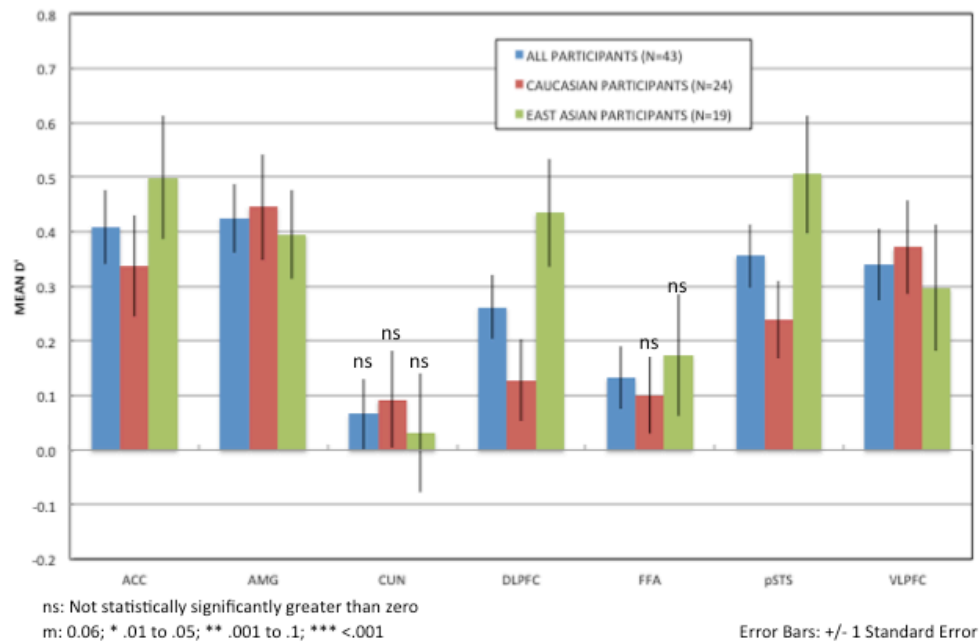


Figure 4. Discriminability of Trustworthy and Untrustworthy–Appearing Stimuli based upon neural patterns of activation during the first two TRs of stimulus presentation in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform gyri / fusiform face area, posterior superior temporal sulcus and ventrolateral prefrontal cortex.

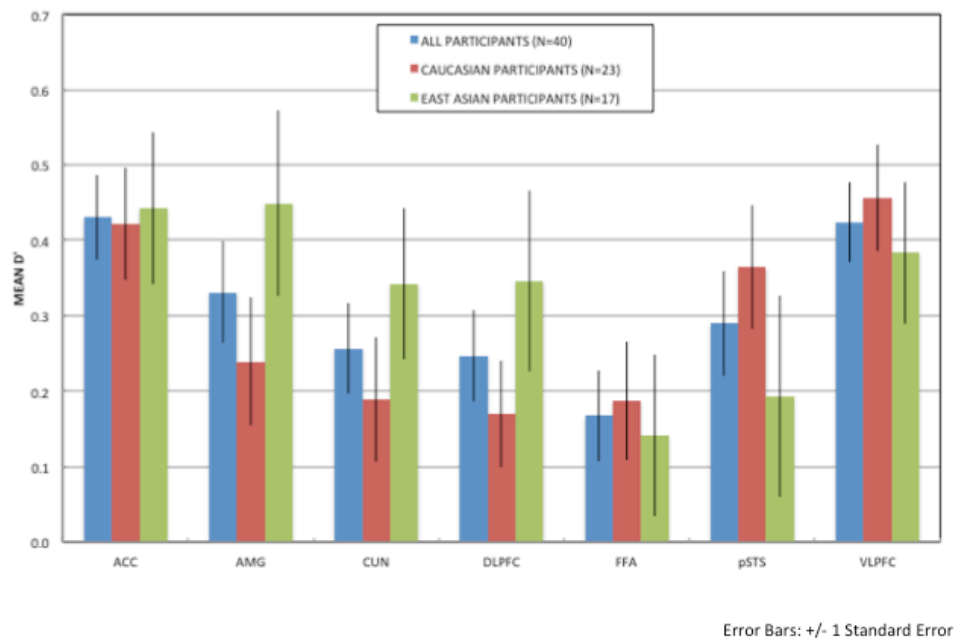


Figure 5. Discriminability of Trustworthy and Untrustworthy–Appearing Stimuli based upon neural patterns of activation during the last two TRs of stimulus presentation in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform gyri / fusiform face area, posterior superior temporal sulcus and ventrolateral prefrontal cortex.

3.3 Comparison of MVPA and Explicit Trust Predictions

Overall, the trustworthiness ratings provided by the participants of this study were higher than those used to initially select and categorize the videos ($F(1,61)=10.55$, $p<.001$), and the Caucasian ratings were higher than the East Asian ratings ($F(1, 61)=7.58$, $p=.008$) (Figure 6), particularly with respect to videos that had previously been categorized as “untrustworthy” in appearance. Across all participants, the correlation between the explicit judgments by the participants in this study and the original raters was moderate, but not significant ($ICC(15)=0.49$, $p=.10$). The disparity between the scanner participant ratings and the original ratings is not particularly surprising, given that the scanner participants had more information upon which to base their judgments, both due to exposure to the full-length videos and their auditory content.

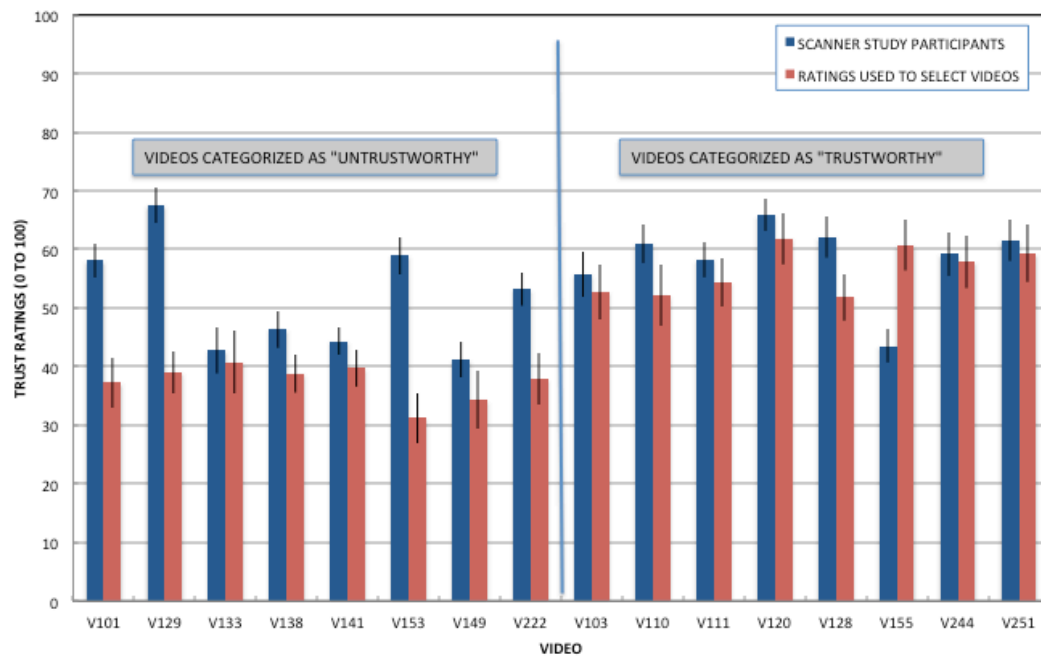


Figure 6. Comparison of explicit judgments by participants in this study and the ratings used to select stimulus videos for use in the study. (The x-axis labels indicate a video identification; error bars represent +/- one standard error.)

Using Signal Detection Theory, the mean discriminability (d') of the explicit trustworthiness judgments 0.38 ($SD=0.72$), which is statistically significantly greater than zero ($t(42)=3.39$, $p<.001$, 1-tailed), indicating successful discrimination between trustworthy and

untrustworthy-appearing faces. The mean d' for East Asian participants was not statistically significantly different from the mean d' for Caucasian participants ($t(32)=-.18$, $p=.86$).

The mean d' 's in the anterior cingulate cortex, cuneus dorsolateral and ventrolateral prefrontal cortices, the fusiform gyri (FFA), and posterior superior temporal sulcus were not statistically significantly different from the mean d' from the explicit trust judgments (Figure 7). The mean d' in the amygdala for East Asian participants was higher than the d' from their explicit judgments ($t(49)=-1.94$, $p=.03$, 1-tailed). The d' in the amygdala for all participants was higher than that of the explicit judgments ($t(116)=-1.95$, $p=.03$, 1-tailed).

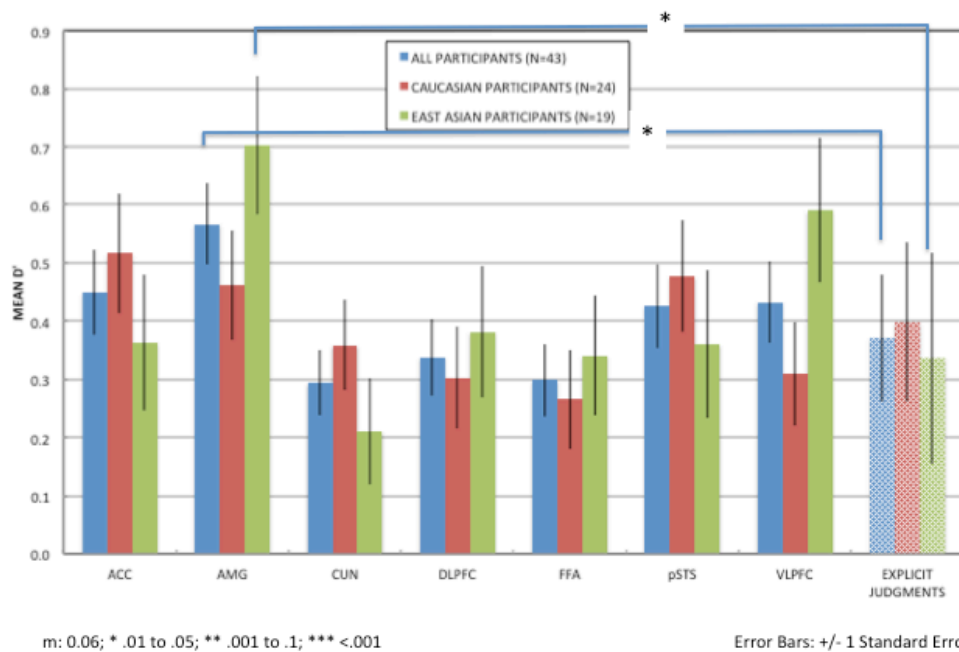


Figure 7. Comparison of Discriminability of Explicit Judgments and Discriminability in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform gyri / fusiform face area, posterior superior temporal sulcus and ventrolateral prefrontal cortex. A “*” indicates that the d' score for that region is statistically significantly greater or less than that of the d' for the explicit judgments.

A comparison of magnitude of the d' scores from the neural responses with the d' scores from the explicit judgments implies that *overall* the explicit judgments discriminating trustworthy from untrustworthy-appearing stimuli are consistent with discrimination based upon neural activation patterns. However, the correlation of the explicit and neural d' scores across participants provides a better indication of whether participants with high d' values based upon neural activation also have high d' values in their explicit judgments, and *vice versa*. In all

regions expect for the DLPFC, the explicit judgment discriminability was not well correlated with the discriminability based upon neural activations (Table 5). In the DLPFC, the neural measure of discriminability was significantly correlated with that from the explicit judgments (All Participants: $r=0.34$, $p=.0002$; Caucasian Participants: $r=0.30$, $p=.01$; East Asian Participants: $r=0.31$, $p=0.03$).

Table 5. Correlation of discriminability (d') based upon neural correlations with discriminability of post-scan explicit trust judgments

Region	Pearson's Correlation	P-value	Region	Pearson's Correlation	P-value
ACC	-0.04	.69	FFA	0.005	.96
AMG	0.10	.28	pSTS	-0.05	.59
CUN	0.04	.65	VLPFC	0.09	.31
DLPFC	0.34	.0002			

3.4 Dependence of Trust Discrimination on Magnitudes of Activation

Based upon the results presented thus far, it is apparent that neural activations can be used to discriminate between stimuli that appear trustworthy and stimuli that appear untrustworthy. To further examine this discrimination, the imaging data were normalized, the classifier rerun, and the discrimination without variability in activation magnitudes was compared to the earlier results (Figure 8). The d' 's based upon both magnitudes and pattern information in most regions were similar to the d' 's based upon only pattern information ($\alpha = .05$). The exception to this finding was in the cuneus, where the d' with both magnitude and pattern information was larger than the d' with only pattern information ($t(116)=-2.26$, $p=.01$, 1-tailed). This suggests that the strength of the BOLD signal in the cuneus is important to differentiating between trustworthy and untrustworthy-appearing stimuli.

3.5 Race Discrimination Based upon Neural Activations

A set of MVPA analyses was performed for which the data from all three runs was combined. With this combined data set, it would have been possible to perform multivariate pattern analysis using only East Asian or only Euro-Caucasian stimuli. In this way, we could compare discriminability under same-race conditions with discriminability under other-race conditions. Unfortunately, hemodynamic shift across the thirty minutes of scanning led to overfitting, and the results were not meaningful. However, I performed multivariate pattern

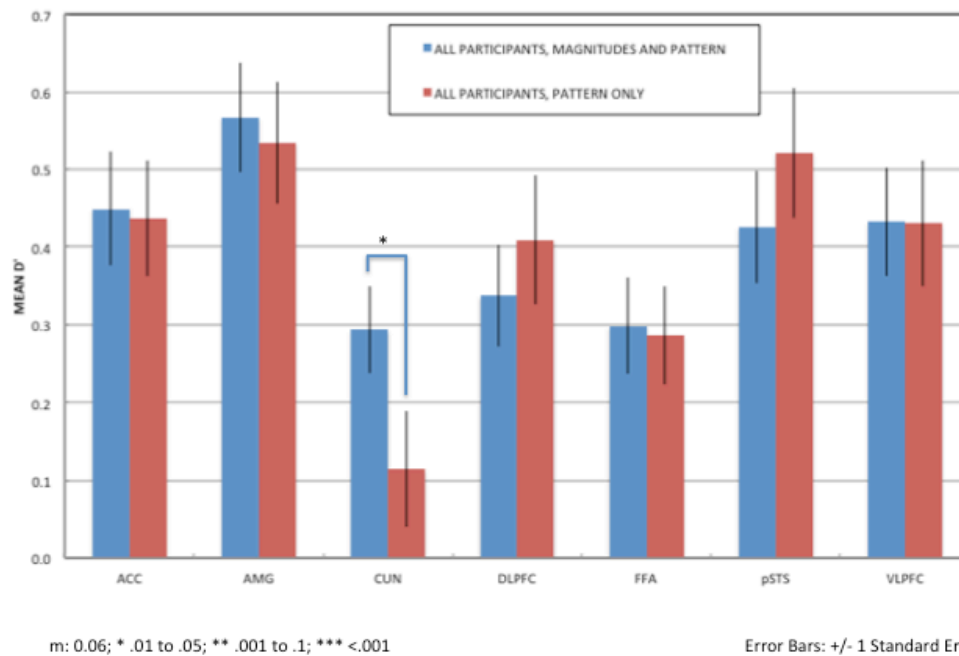


Figure 8. Mean discriminability between trustworthy and untrustworthy-appearing stimuli with magnitudes of activation and with pattern information only, in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform gyri / fusiform face area, posterior superior temporal sulcus and ventrolateral prefrontal cortex.

analysis on the independent runs to discern how well the neural patterns of activation in the East Asian and Caucasian cohorts discriminated between East Asian and Caucasian stimuli (Figure 9). Notably, d' was greater than zero ($\alpha = .05$) for all regions, except the dorsolateral ($t(66)=0.73$, $p=.47$) and ventrolateral prefrontal cortices ($t(66)=0.90$, $p=.37$) for the Caucasian subjects. In contrast, the only region for which d' was statistically significantly greater than zero for East Asian subjects was the amygdala ($t(49)=2.47$, $p=.009$, 1-tailed). These findings suggest that the neural activation patterns for the Caucasian participants viewing East Asian and Caucasian stimuli are more dissociable than for East Asian participants.

Again, to better understand the temporal component of the neural responses, the pattern classifier was run on the average of data from just the first two brain volumes (first two TRs). The d 's for the Caucasians for the amygdala ($t(67)=0.52$, $p=.30$, 1-tailed), cuneus ($t(67)=0.19$, $p=.42$, 1-tailed) and DLPFC ($t(67)=1.07$, $p=.14$, 1-tailed) were no longer statistically significantly greater than zero (Figure 10). The d' for the East Asians for the amygdala was also no longer greater than zero ($t(49)=1.14$, $p=.13$, 1-tailed). The d' for the fusiform gyrus, however,

was greater than zero for both the Caucasians ($t(67)=1.92$, $p=.03$, 1-tailed) and, marginally, the East Asians ($t(49)=1.59$, $p=.06$).

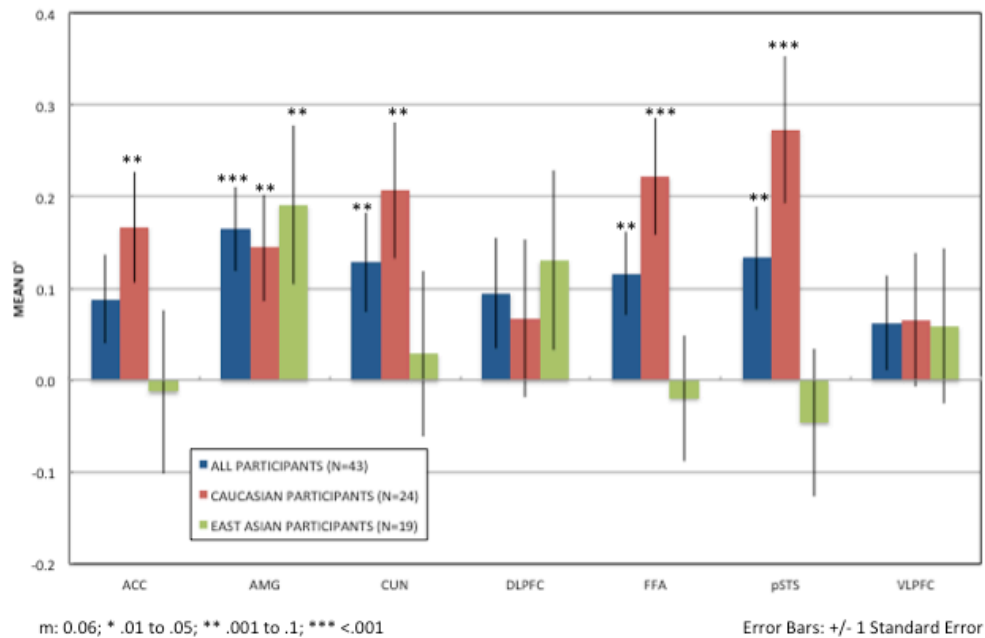


Figure 9. Discriminability of East Asian and Caucasian Stimuli based upon neural patterns of activation in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform gyri / fusiform face area, posterior superior temporal sulcus and ventrolateral prefrontal cortex.

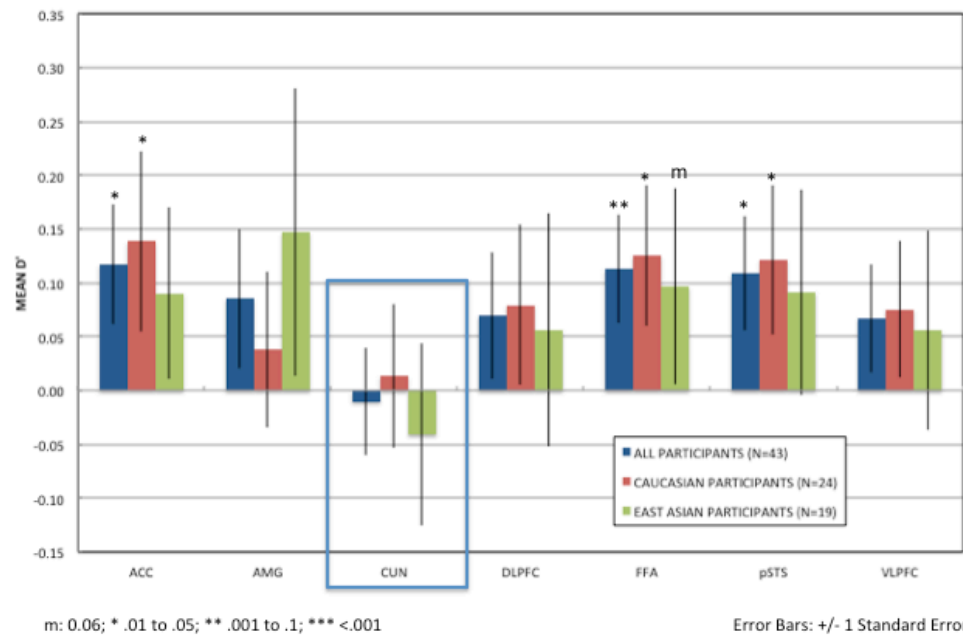


Figure 10. Discriminability of East Asian and Caucasian Stimuli based upon neural patterns of activation during the first two TRs of stimulus presentation.

In contrast, when the pattern classifier was run on the average of the data from just the last two brain volumes (last two TRs), only the ACC had a d' greater than zero for Caucasian participants (Figure 11). Additionally, for the Caucasians, d' was statistically significantly smaller for the last two brain volumes than the first two in fusiform gyrus and pSTS ($\alpha = .05$). Among East Asian participants, none of the brain regions were statistically significantly different in the first and last two brain volumes. The d' in the amygdala of East Asian participants was statistically significantly greater than zero ($t(49)=4.93$, $p<.0001$, 1-tailed) and greater than that for the Caucasian participants ($t(113)=3.50$, $p=.0003$, 1-tailed). It is worth noting that the d' for Caucasian participants in the fusiform gyrus was less than zero during the last two brain volumes ($t(67)=-2.06$, $p=.02$, 1-tailed). This finding suggests that the classifier may be overfitting the data, making the evidence for discrimination suspect.

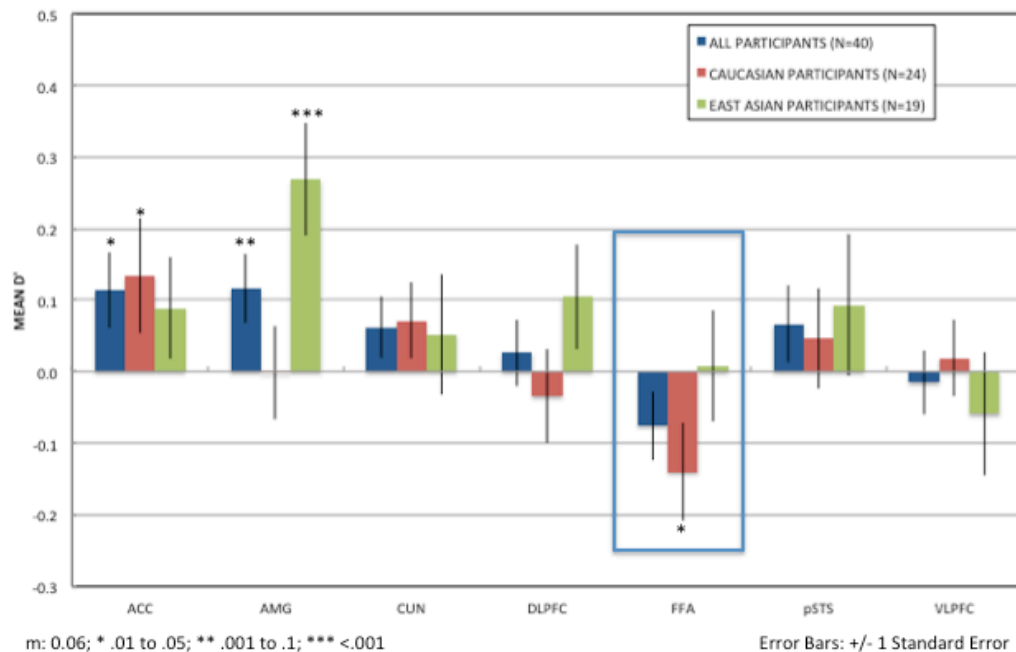


Figure 11. Discriminability of East Asian and Caucasian Stimuli based upon neural patterns of activation during the last two TRs of stimulus presentation.

Overall, the evidence for discriminability in the later two TRs was much weaker than in the first two TRs, suggesting that discrimination on race occurs early in the stimulus presentation. The evidence for discriminability for East Asian participants was much weaker than for Caucasian participants. Furthermore, even for the Caucasians, discrimination was

largely dependent upon magnitudes of activation, as shown by the reduced d' 's using all four TRs in a pattern only condition in Figure 12.

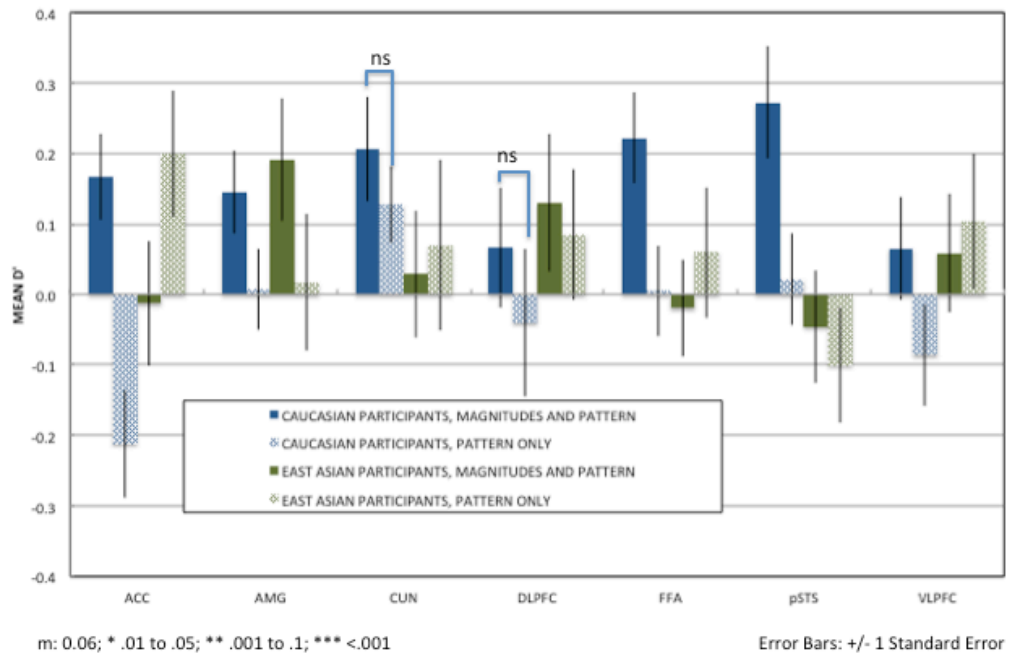


Figure 12. Mean discriminability between East Asian and Caucasian stimuli with magnitudes of activation and with pattern information only across all four TRs in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform gyri / fusiform face area, posterior superior temporal sulcus and ventrolateral prefrontal cortex.

CHAPTER 4

DISCUSSION

4.1 Dissociating on Trust

The central finding of this study was evidence of neural pattern dissociation based upon stimuli varying in appearance of trustworthiness. Evidence of dissociation was found in the anterior cingulate cortex, the amygdala, the dorsolateral prefrontal cortex, the fusiform gyrus, the posterior superior temporal sulcus, the ventrolateral prefrontal cortex and, to a lesser degree, the cuneus.

Existing literature provides strong evidence that the amygdala has a key role in trustworthiness evaluations. Early researchers observed a linear response to faces varying along a trust dimension, with untrustworthy faces resulting in increased BOLD signals associated with the amygdala (Todorov & Engell, 2008; Winston et al., 2002). More recent studies indicate that there is a quadratic response in the amygdala, with increased BOLD signals for both highly untrustworthy and highly trustworthy faces (Rule, Krendl, Ivcevic, & Ambady, 2013; Said, Baron, & Todorov, 2009; Said et al., 2011; Todorov et al., 2008). Given the likelihood of a parabolic response, the finding of discriminability in the amygdala using a linear classifier is somewhat surprising and warrants further research. One potential explanation is that, although the response of the amygdala is quadratic, it is not parabolic, so that activations to untrustworthy face are, in fact, greater than activations to trustworthy faces. Thus, an asymmetric quadratic equation, rather than a parabola, might better describe the relationship between the appearance of trustworthiness and discrimination. However, discrimination was successful within the amygdala even without variability in magnitudes of activation, indicating that the *pattern* of activation is sufficient to discriminate trustworthy and untrustworthy-appearing stimuli. To further examine this finding, additional spatial resolution within the amygdala may be required.

Dissociation on trustworthiness in the anterior cingulate cortex was not a foregone conclusion for this study. Previous cross-race studies have observed a stronger ACC BOLD

signal in *explicit* trustworthiness judgments of other-race faces (Stanley, Phelps, & Banaji, 2008), and additional recruitment of the ACC and right pSTS has been observed in trust games in the context of making decisions to trust partners deemed untrustworthy (Grèzes, Frith, & Passingham, 2004a; Winston et al., 2002). One interpretation of the ACC in these studies is that the ACC is responding to the conflict between the automatic response in the amygdala and the conscious intentions to respond fairly, in spite of this automatic response (Richeson et al., 2003). If this interpretation is correct, we would not expect the ACC to have significant activations in an implicit trust study and, might not, therefore, expect it to support discrimination between trustworthy and untrustworthy participants under implicit evaluation conditions. However, not only did this study differ from other cross-cultural trust studies in that it was entirely implicit, but it also differed in the use of dynamic stimuli which were presented over an 8 second duration. Even without the requirement for an explicit judgment of trustworthiness on the part of the participant, it is plausible that this length of time is sufficient to trigger activations in the ACC in response to a conflict between the automatic assessment of the amygdala and a top-down evaluation by the dorsolateral prefrontal cortex. This hypothesis is consistent with a study of implicit measures of racial attitudes by Cunningham et al. (2004) in which the elevated BOLD signal in the amygdala observed when other-race faces were presented for 30ms was significantly reduced when the faces were presented for 525 ms. The evidence of discriminability in the ACC under conditions allowing for the additional time for deliberate, conscious processing may indicate top-down control of the automatic amygdala response, even without the requirement of an explicit judgment.

Dissociation on trustworthiness in the cuneus was also not a foregone conclusion for this study. Although it is part of the basic visual processing system, the cuneus has been associated with attentional processing as needed to distinguish and classify emotions (Habel et al., 2007), and Kilts et al. (2003) found activation in the cuneus in response to dynamic stimuli with emotional expression, particularly of happiness. The finding of discriminatory power in the cuneus is consistent, therefore, with the finding by Engell et al. (2007) that the cuneus is linearly modulated with judgments of trustworthiness. However, it is, perhaps, surprising that the cuneus did not have a significant d' score in the first two TRs of the stimulus presentation, suggesting

that in spite of being part of the basic visual processing system, the role of the cuneus in discriminating trustworthy from untrustworthy stimuli occurs after the response of not only the amygdala, but also the anterior cingulate, dorsolateral prefrontal and ventrolateral prefrontal cortices. The difference in the response of the cuneus from the other regions may suggest that its role in processing these stimuli is more closely related to attention than, in fact, trait evaluations.

4.2 Comparison of East Asian and Caucasian Discriminability of Trustworthy Stimuli

Another key finding of this study was evidence that discriminability based upon the neural activation patterns across the duration of exposure is similar in the East Asian and Caucasian participants. Although prior cross-cultural behavioral studies were not specific to trustworthiness evaluations, they found strong agreement across cultures on traits of related traits, such as dominance, strength, warmth, honesty, babyfacedness and attractiveness (Albright et al., 1997; McArthur & Berry, 1987; Walker et al., 2011; Zebrowitz et al., 1993; Zebrowitz et al., 2012). Thus, the similarity in results for East Asian and Caucasian participants is not unexpected. It does, however, contradict the study by Na and Kitayama (2011) indicating that East Asian Americans do not automatically make trait evaluations.

4.3 Discriminability with and without Magnitudes of Activation

Based upon prior literature, I anticipated that magnitudes of activation would be required for dissociation of neural activation patterns during presentation of trustworthy and untrustworthy-appearing stimuli, at least in regions having previously been identified as having a linear response to faces varying in trustworthiness. In fact, the only region for which discrimination was negatively impacted by normalizing the magnitudes of activation was the cuneus. The finding that the cuneus requires magnitudes of activation is consistent with the findings in the study by Engell et al. (2007), in which linear modulation was observed in the cuneus during trustworthiness judgments. The more interesting outcome is, perhaps, the *lack* of sensitivity to the absence of magnitudes of activation in regions such as the fusiform gyrus, dorsolateral prefrontal cortex and ventrolateral prefrontal cortex, which have been previously thought to respond linearly to faces varying in trustworthiness. Given the inconsistency of the results of studies attempting to characterize the linearity or lack of linearity in activation of

various regions during trust judgments, it is likely that higher resolution and smaller regions of interest may be required to better understand how magnitudes of activation contribute to discrimination.

4.4 Comparison of MVPA and Explicit Judgment Discriminability

The explicit trust judgments by participants in this study were moderately correlated with the judgments used to categorize and select stimuli for use in the scanner. In spite of the pro-Caucasian bias found among Caucasian participants in the implicit association test, East Asian videos were rated higher than Caucasian videos and East Asian and Caucasian participants rated similarly. Although the explicit trust discriminability (d') was greater than zero, it was relatively low ($M=0.38$), with 47% correct judgments overall. Since this study was conducted in conjunction with a deception detection study, 50% of the individuals in the stimulus data set were lying. Thus, not only did participants in this study have additional content (both auditory and dynamic visual content) than is typically used in trustworthiness trait evaluation studies, but some of that content included deception, which may have influenced the trust judgments.

Although the mean discriminability of the explicit judgments was relatively low, it was similar to the mean discriminability based upon neural activation patterns with two notable exceptions. The mean d' across all participants in the amygdala was statistically significantly greater than the mean d' from the explicit judgments. This finding was consistent with studies showing that the BOLD signal in the amygdala is not correlated with explicit measures of racial attitudes (Stanley et al., 2012). The mean d' in the cuneus, on the other hand, was low relative to the d' from explicit trust judgments for East Asian participants. Given that the cuneus has been implicated exclusively in other-race trust evaluation studies, the weaker evidence of discriminability in the cuneus for East Asians may be related to the lack of implicit bias observed in East Asians using the implicit association test.

Although mean discriminability from the neural patterns was similar to mean discriminability from the explicit trust judgments, the d' scores were only correlated in the DLPFC. The role of the DLPFC in both cross-race and trait evaluation studies is still under investigation. However, it appears to have a role in regulation and control of autonomic responses that conflict with conscious intent (MacDonald, Wilson, Cohen, Stenger, and

Cameron, 2000; Kubota et al., 2012; Stanely et al., 2008; Knutson, Mah, Manly and Grafman, 2007). Alternatively, since the DLPFC is a highly integrative region of the brain, thought to contribute clarity to complex social situations and perspective-taking (van den Bos, van Dijk, Westenberg, Robvouts & Crone, 2011; Miller & Cummings, 2007). The correlation between the neural and explicit judgment discriminability measures may, therefore, reflect that the DLPFC was engaged in similar types of social cognition functions during the passive scanner task as in the explicit judgment task.

4.5 Race Dissociation

Although the number of trials was insufficient to run the classifier using exclusively same-race or other-race stimuli, I was able to run an analysis to determine whether East Asian and Caucasian stimuli could be discriminated based upon neural activations. Although the finding was weaker than for discriminating on trust, dissociation based upon neural activations across all four timepoints (8 seconds) was found with Caucasian participants in the anterior cingulate cortex, amygdala, cuneus, fusiform gyrus, posterior superior temporal sulcus and ventrolateral prefrontal cortex. These results are consistent with the study by Natu et al. (2011), which found an important temporal difference between the response of the fusiform face area and other ventral temporal regions to same and other-race faces. Natu et al.'s (2011) study found that same-race faces elicit a response in the fusiform face area and other ventral temporal regions that is initially strong, but attenuates quickly, while other-race faces elicit a response in the fusiform face area and ventral temporal regions that is initially weak, but increases over time. In this study, the dissociation based upon neural activations of Caucasian participants was dependent upon magnitudes of activation, and occurred primarily during the first 4 seconds of exposure, with no significant evidence of discriminability in the last 4 seconds of exposure. However, because the stimulus set included both same and other-race faces, it is not possible to fully replicate the Natu et al. (2011) findings, other than to merely identify that both spatial and temporal data contribute to race discrimination.

The finding that Caucasians' neural activations dissociated on race and East Asian neural activations did not was consistent with the results of the implicit association test, in which Caucasians were found to have an implicit pro-Caucasian bias, and East Asians were not found

to have any measurable bias. It should be noted that in the last four seconds of stimulus presentation, discriminability was found in the East Asian population in the amygdala. The late-time appearance of the amygdala is somewhat surprising, given that the response of the amygdala is thought to precede activations in the anterior cingulate and prefrontal cortices. However, most research into the amygdala response to other-race stimuli uses static stimuli without context. Future research, particularly using exclusively same or other-race stimuli is warranted to better understand the time course of the amygdala response to same and other-race dynamic, naturalistic stimuli for which there is context.

It is notable that the overall discrimination on race was statistically significantly weaker than the discrimination on trust in every region. Furthermore, the discrimination on race appears to occur primarily early during the stimulus presentation, and the discrimination on trustworthiness occurs throughout the 8 seconds. The lack of race discrimination late in the stimulus presentation may indicate some habituation to the stimuli, while the evidence of trust discrimination throughout the stimulus presentation is consistent with a level of processing involving more individuation and in-depth processing. Future research using blocks of same and other-race stimuli could be useful to build upon Natsu et al.'s (2011) work by exploring how the temporal component of trust evaluations may differ under same and other-race conditions. In addition, future research comparing the discriminability by participants with and without the context of prior exposure to full videos with auditory content will verify whether context strengthened the response to the video clips in the scanner, and provide insight into how a lack of context impacts implicit trustworthiness dissociation.

4.6 Limitations and Recommendations for Future Studies

With a passive task, as was used in this study, it is impossible to say definitively that trust evaluations were, in fact, occurring during the presentation of stimuli. The response in the cuneus may, in fact, be related to attention, rather than a trust evaluation. Winston et al. (2002) showed that implicit trustworthy evaluations occur, at least in the amygdala and fusiform gyrus. However, evidence of other regions being involved in implicit trustworthy evaluations is weaker. Given the close relationship between the appearance of trustworthiness and facial structure and emotion recognition (Todorov, 2008), it is not possible to say whether the discriminations found

herein reflect implicit judgments of trustworthiness or observation of some other aspect of the visual stimuli that corresponds with the appearance of trustworthiness.

This study was conducted in conjunction with a deception study. For that reason, half of the stimuli that appeared trustworthy lied in the full-length videos about their opinion of the movie they described and, likewise, half of the stimuli that appeared untrustworthy told the truth about their opinion of the movie they described. Participants were not told that the stimuli might be lying about their opinions. Although the brain regions associated with deception detection are still being researched, regions such as the amygdala, anterior cingulate cortex, ventrolateral cortex, dorsolateral cortex, and posterior superior temporal sulcus are likely to be implicated (Grèzes et al., 2004a; Grèzes, Frith, & Passingham, 2004b; Grèzes & Passingham, 2006; Harada et al., 2009). Given that several regions are likely to be common to both trust and deception detection a future study is recommended that removes the deception factor.

The experimental design for this study did not offer the ability to compare dissociations with same-race stimuli to dissociations with other-race stimuli. Although the results herein, along with behavioral results, suggest that trust evaluations are largely universal, a future study should be designed specifically to compare the neural activation patterns associated with same-race stimuli to those associated with other-race stimuli. A second limitation to this study was that all East Asian participants were citizens and long-time or second-generation residents of the United States. Future studies would be particularly informative if the East Asian participants live in East Asia and have comparatively little exposure to Euro-Americans. A larger sample size with a study conducted in the US would be beneficial to examine how relatively new East Asian residents of the US differ from second-generation or longer residents.

CHAPTER 5

CONCLUSION

This study advanced previous research into appearance-based trustworthiness trait evaluations by introducing a cross-cultural context using videos, rather than static images, and by applying machine learning to understand how neural patterns of activation are dissociated in response to trustworthy and untrustworthy-appearing stimuli. I found strong evidence that the patterns of activation elicited while passively viewing trustworthy and untrustworthy-appearing faces are dissociable in the anterior cingulate cortex, amygdala, cuneus, dorsolateral prefrontal cortex, fusiform face area, posterior superior temporal sulcus and ventrolateral prefrontal cortex. The discrimination occurs throughout the stimulus presentation and, with the exception of the cuneus, dissociation in these regions is not dependent upon magnitudes of activation. Discriminability based upon the neural patterns of activation among East Asian participants was similar to that among Caucasian participants. I also found evidence that patterns of activation elicited while passively viewing same and other race faces are dissociable. The discriminability is weaker than for trust discriminations, and was observed only among Caucasian participants.

Prior to this study, most studies of the neural activations in response to trustworthy and untrustworthy-appearing stimuli have used inferential statistics, particularly, general linear modeling (GLM). GLM has proliferated as the method of choice in fMRI data analysis because it is conceptually simple, readily available in standard packages, flexible, amenable to standard statistical tests and not overly computationally intensive (Poline & Brett, 2012). However, GLM methods are predicated on the assumptions of normality, which is difficult to check, and independence, which is highly improbable. In an effort to expand beyond the GLM approach, some researchers are adopting methods such as independent or principle component analysis, machine learning, clustering techniques and multi-voxel pattern analysis. These approaches are particularly appealing in the context of social neuroscience, where the number and complexity of associated cognitive processes, which serve to exacerbate the inherent risk of failure associated

with hypothesis or model-driven approaches, have thus far impeded progress. MVPA, in particular, provides an opportunity to look at both the temporal and spatial relationships and, in so doing, capture within and between-group variations that may be overlooked using General Linear Modeling. Even with the advantages of MVPA, the results of this study challenge the research community to look more closely at the impacts of context and dynamic stimuli on trait evaluations and to apply machine learning and other exploratory methods to further elucidate both the network of regions that contribute to trait evaluations and the specific contributions that each region makes.

APPENDIX A
CROSS-CULTURAL QUESTIONNAIRE

1. Gender:

☐ Male ☐ Female

2. Age:

_____ yrs

3. Which of the following are you attending?

☐ Undergraduate college

☐ Graduate school

4. Which year of college or graduate school are you in?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8+

5. Major:

6. Country where you attended high school:

☐ U.S. ☐ Other (please specify): _____

7. Country where you were born:

☐ U.S. ☐ Other (please specify): _____

8. Country of citizenship:

☐ U.S. ☐ Other (please specify): _____

9. How long have you lived in the United States?

☐ All my life ☐ _____ year(s)

10. Number of parents born in the U.S.?

☐ 0 ☐ 1 ☐ 2

11. Number of grandparents born in the U.S.?

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4

12. If you are an American citizen, permanent resident, or have a green card, what is your racial/ethnic background? If you identify with more than one racial or ethnic group, please circle/write all that apply.

- ☐ White/Caucasian
- ☐ African American
- ☐ Asian American
- ☐ Hispanic/Latino
- ☐ Native American
- ☐ Others (please specify): _____

13. If you are an international student, what is your ethnic background? If you identify with more than one ethnic group, please write all that apply.

14. What is/are your native language/s? (the language/s you speak at home)

15. What is the highest educational attainment of your father?

- ☐ Some high school
- ☐ Completed high school
- ☐ Some college
- ☐ Completed college (bachelor's)
- ☐ Some post-graduate
- ☐ Post-graduate degree (MD, Ph.D., LLB, MS, etc.)

16. What is the highest educational attainment of your mother?

- ☐ Some high school
- ☐ Completed high school
- ☐ Some college
- ☐ Completed college (bachelor's)
- ☐ Some post-graduate
- ☐ Post-graduate degree (MD, Ph.D., LLB, MS, etc.)

17. What is the income of your immediate family (to the best of your knowledge)?

- ☐ Less than \$15,000
- ☐ \$15,001-25,000
- ☐ \$25,001-35,000
- ☐ \$35,001-50,000
- ☐ \$50,001-75,000
- ☐ \$75,001-100,000
- ☐ \$100,001-150,000
- ☐ more than \$150,000

18. What is your religious background?

- ☐ Catholic
- ☐ Hindu

- ☐ Jewish
☐ Muslim
☐ Protestant (Baptist, Methodist, Episcopalian)
☐ Other (please specify)

FOR ASIAN PARTICIPANTS

19a – Rate your overall English Language Ability	Know Little		Moderate Knowledge		Fluent	
	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
19b – How often do you watch English language films and TV?	Never	Rarely	Sometimes	Often	Mostly	Always
	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
19c – How often do you watch Asian language films and TV?	Never	Rarely	Sometimes	Often	Mostly	Always
	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
19d – How often do you socialize with European Americans?	Never	Rarely	Sometimes	Often	Mostly	Always
	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
19e – Are any of your close friends European Americans?	<input type="checkbox"/> Yes	<input type="checkbox"/> No	If yes, how many? _____			
19f – Are any of your family members European Americans?	<input type="checkbox"/> Yes	<input type="checkbox"/> No	If yes, how many? _____			

FOR EUROPEAN AMERICAN PARTICIPANTS

19a – Do you speak any East Asian language?	<input type="checkbox"/> Yes	<input type="checkbox"/> No	If yes, which one(s) _____			
If yes, how fluent are you?	Know Little		Moderate Knowledge		Fluent	
	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
19b – Have you visited any East Asian countries??	<input type="checkbox"/> Yes	<input type="checkbox"/> No				
If yes, which countries and for how long?	_____					
19c – How often do you socialize with East Asians?	Never	Rarely	Sometimes	Often	Mostly	Always
	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
19d – Are any of your close friends East Asians?	<input type="checkbox"/> Yes	<input type="checkbox"/> No	If yes, how many? _____			
19e – Are any of your family members East Asians?	<input type="checkbox"/> Yes	<input type="checkbox"/> No	If yes, how many? _____			

REFERENCES

- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, 393(6684), 470. doi:10.1038/30982
- Albright, L., Malloy, T. E., Dong, Q., Kenny, D. A., Fang, X., Winquist, L., & Yu, D. (1997). Cross-cultural consensus in personality judgments. *Journal of Personality and Social Psychology*, 72(3), 558-569. doi:10.1037/0022-3514.72.3.558
- Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., ... Moran, R. (2014). SPM12 Manual.
- Bakmazian, A. (2014). The man behind the beard: perception of men's trustworthiness as a function of facial hair. *Psychology*, Vol.05, No.03, 7. doi:10.4236/psych.2014.53029
- Blanz, V., & Vetter, T. (1999, July). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (pp. 187-194). ACM Press/Addison-Wesley Publishing Co.
- Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization*, 43(1), 57-73. doi:10.1016/S0167-2681(00)00108-6
- Chiao, J. Y., Iidaka, T., Gordon, H. L., Nogawa, J., Bar, M., Aminoff, E., . . . Ambady, N. (2008). Cultural specificity in amygdala response to fear faces. *Journal of Cognitive Neuroscience*, 20(12), 2167-2174. doi:10.1162/jocn.2008.20151
- Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, 51(2), 143-159. doi:10.1016/S0167-2681(02)00092-6
- Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology*, 85(4), 639-649. doi:10.1037/0022-3514.85.4.639
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, 15(12), 806-813. doi:10.2307/40064052

- Derntl, B., Habel, U., Robinson, S., Windischberger, C., Kryspin-Exner, I., Gur, R. C., & Moser, E. (2009). Amygdala activation during recognition of emotions in a foreign ethnic group is associated with duration of stay. *Social Neuroscience*, 4(4), 294-307.
- Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, 100(6), 999-1014. doi:10.1037/a0023026, 10.1037/a0023026.supp (Supplemental)
- Dotsch, R. O. N., Wigboldus, D. H. J., & Knippenberg, A. D. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology*, 43(1), 116-125. doi:10.1002/ejsp.1928
- Dzhelyova, M. P., Ellison, A., & Atkinson, A. P. (2011). Event-related repetitive TMS reveals distinct, critical roles for right OFA and bilateral posterior pSTS in judging the sex and trustworthiness of faces. *Journal of Cognitive Neuroscience*, 23(10), 2782-2796.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, 19(9), 1508-1519. doi:10.1162/jocn.2007.19.9.1508
- Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception*, 39(7), 931-941. doi:10.1068/p6633
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77-83. doi:10.1016/j.tics.2006.11.005
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1), 32-41. doi:10.1016/j.neuropsychologia.2006.04.015
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17-41. doi:10.1037/a0015575, 10.1037/a0015575.supp (Supplemental)
- Grèzes, J., Frith, C., & Passingham, R. E. (2004a). Brain mechanisms for inferring deceit in the actions of others. *J Neurosci*, 24(24), 5500-5505. doi:10.1523/JNEUROSCI.0219-04.2004

- Grèzes, J., Frith, C. D., & Passingham, R. E. (2004b). Inferring false beliefs from the actions of oneself and others: an fMRI study. *NeuroImage*, 21(2), 744-750. doi:10.1016/S1053-8119(03)00665-7
- Grèzes, J., S., B., & Passingham, R. E. (2006). Amygdala activation when one is the target of deceit: did he lie to you or to someone else? *NeuroImage*, 30(2), 601-608.
- Habel, U., Windischberger, C., Derntl, B., Robinson, S., Kryspin-Exner, I., Gur, R. C., & Moser, E. (2007). Amygdala activation and facial expressions: Explicit emotion discrimination versus implicit emotion processing. *Neuropsychologia*, 45(10), 2369-2377. doi: 10.1016/j.neuropsychologia.2007.01.023
- Harada, T., Itakura, S., Xu, F., Lee, K., Nakashita, S., Saito, D. N., & Sadato, N. (2009). Neural correlates of the judgment of lying: A functional magnetic resonance imaging study. *Neuroscience Research*, 63(1), 24-34. doi:10.1016/j.neures.2008.09.010
- Hellström, Å., & Tekle, J. (1994). Person perception through facial photographs: effects of glasses, hair, and beard on judgments of occupation and personal qualities. *European Journal of Social Psychology*, 24(6), 693-705.
- Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage*, 18(1), 156-168. doi: 10.1006/nimg.2002.1323
- Krumhuber, E., Manstead, A., & Kappas, A. (2007). Temporal Aspects of Facial Displays in Person and Expression Perception: The Effects of Smile Dynamics, Head-tilt, and Gender. *Journal of Nonverbal Behavior*, 31(1), 39-56. doi:10.1007/s10919-006-0019-x
- Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7(4), 730-735. doi:10.1037/1528-3542.7.4.730
- Leder, H., Forster, M., & Gerger, G. (2011). The Glasses Stereotype Revisited. *Swiss Journal of Psychology*, 70(4), 211-222. doi:10.1024/1421-0185/a000059
- Liang, C. (2015, August 2015). *Trustworthiness evaluations of East Asian and Caucasian faces*. [Database and data analysis spreadsheets]. Unpublished research, University of Texas at Dallas.
- Maldjian, J. A., Laurienti, P. J., & Burdette, J. H. (2004). Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *NeuroImage*, 21(1), 450-455.
- Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, 19(3), 1233-1239.

- Mattavelli, G., Andrews, T. J., Asghar, A. U. R., Towler, J. R., & Young, A. W. (2012). Response of face-selective brain regions to trustworthiness and gender of faces. *Neuropsychologia*, 50(9), 2205-2211. doi:10.1016/j.neuropsychologia.2012.05.024
- McArthur, L. Z., & Berry, D. S. (1987). Cross-cultural agreement in perceptions of babyfaced adults. *Journal of Cross-Cultural Psychology*, 18(2), 165-192.
- Miller, B. L., & Cummings, J. L. (Eds.). (2007). *The human frontal lobes: Functions and disorders*. Guilford press.
- Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, 27(4), 237-254. doi:10.1023/A:1027332800296
- Muscarella, F., & Cunningham, M. R. (1996). The evolutionary significance and social perception of male pattern baldness and facial hair. *Ethology and Sociobiology*, 17(2), 99-117. doi: 10.1016/0162-3095(95)00130-1
- Na, J., & Kitayama, S. (2011). Spontaneous trait inference is culture-specific: Behavioral and neural evidence. *Psychological Science*, 22(8), 1025-1032. doi:10.1177/0956797611414727
- Natu, V., Raboy, D., & O'Toole, A. J. (2011). Neural correlates of own-and other-race face perception: Spatial and temporal response differences. *NeuroImage*, 54(3), 2547-2555.
- Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, 6(1), 101-115. doi:10.1037/1089-2699.6.1.101
- Olivola, C., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, 34(2), 83-110. doi:10.1007/s10919-009-0082-1
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087-11092. doi:10.1073/pnas.0805664105
- Pinkham, A. E., Hopfinger, J. B., Pelphrey, K. A., Piven, J., & Penn, D. L. (2008). Neural bases for impaired social cognition in schizophrenia and autism spectrum disorders. *Schizophrenia Research*, 99(1-3), 164-175. doi:10.1016/j.schres.2007.10.024
- Poline, J. B., & Brett, M. (2012). The general linear model and fMRI: does love last forever? *Neuroimage*, 62(2), 871-880.

- Quadflieg, S., Flannigan, N., Waiter, G. D., Rossion, B., Wig, G. S., Turk, D. J., & Macrae, C. N. (2011). Stereotype-based modulation of person perception. *NeuroImage*, 57(2), 549-557. doi:10.1016/j.neuroimage.2011.05.004
- Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, 106(6), 897-911. doi:10.1037/a0036498
- Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Research*, 1376(0), 66-75. doi: 10.1016/j.brainres.2010.12.041
- Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., & Shelton, J. N. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, 6(12), 1323-1328. doi:10.1038/nn1156
- Rule, N. O., Ambady, N., Adams, R. B., Jr., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. *Journal of Personality and Social Psychology*, 98(1), 1-15. doi:10.1037/a0017673, 10.1037/a0017673.supp (Supplemental)
- Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology*, 104(3), 409-426. doi:10.1037/a0031050
- Said, C. P., Haxby, J. V., & Todorov, A. (2011). Brain systems for assessing the affective value of faces. *Philos Trans R Soc Lond B Biol Scie*, 366(1571), 1660-1670. doi: 10.1098/rstb.2010.0351
- Santos, I. M., & Young, A. W. (2010). Inferring social attributes from different face regions: Evidence for holistic processing. *The Quarterly Journal of Experimental Psychology*, 64(4), 751-766. doi:10.1080/17470218.2010.519779
- Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research*, 20(1), 81-91. doi:10.1016/j.cogbrainres.2004.01.008
- Stanley, D., Phelps, E., & Banaji, M. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science*, 17(2), 164-170. doi:10.1111/j.1467-8721.2008.00568.x
- Stanley, D. A., Sokol-Hessner, P., Fareri, D. S., Perino, M. T., Delgado, M. R., Banaji, M. R., & Phelps, E. A. (2012). Race and reputation: Perceived racial group trustworthiness

- influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1589), 744-763. doi:10.1098/rstb.20110300
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, 21(3), 349-354. doi:10.1177/0956797610362647
- Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, 1124, 208-224. doi:10.1196/annals.1440.012
- Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, 25(3), 395-410. doi:10.1080/02643290802044996
- Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cognitive & Affective Neuroscience*, 3(4), 303-312. doi:10.1093/scan/nsn033
- Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, 45(1), 163-173. doi:10.1016/j.neuropsychologia.2006.04.018
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308, 1623-1626. doi:10.1126/science.1110589
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, 66(1), 519-545. doi:10.1146/annurev-psych-113011-143831
- Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotions in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Research*, 1284(0), 100-115. doi:10.1016/j.brainres.2009.05.075
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273-289.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, 43(2), 161-204. doi:10.1080/14640749108400966

- van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A., & Crone, E. A. (2011). Changing brains, changing perspectives the neurocognitive development of reciprocity. *Psychological Science*, 22(1), 60-70.
- Walker, M., Jiang, F., Vetter, T., & Sczesny, S. (2011). Universals and cultural differences in forming personality trait judgments from faces. *Social Psychological and Personality Science*, 2(6), 609-617. doi:10.1177/1948550611402519
- Walker, M., & Vetter, T. (2009). Portraits made to measure: Manipulating social judgments about individuals with a statistical face model. *Journal of Vision*, 9(11), 12.
- Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice. *Psychological Science (Wiley-Blackwell)*, 16(1), 56-63. doi:10.1111/j.0956-7976.2005.00780.x
- Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, 5(3), 277. doi:doi:10.1038/nn816
- Wogalter, M. S., & Hosie, J. A. (1991). Effects of cranial and facial hair on perceptions of age and person. *Journal of Social Psychology*, 131(4), 589-591.
- Zebrowitz, L. A., & Montepare, J. M. (2005). Appearance DOES matter. *Science*, 308(5728), 1565-1566. doi:10.2307/3841598
- Zebrowitz, L. A., & Montepare, J. M. (2015). Faces and first impressions. In M. Mukulincer, P. R. Shaver, E. Borginda, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition* (pp. 251-276). Washington, D.C., US: American Psychological Association.
- Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individual impressions of other racial groups. *Journal of Personality and Social Psychology*, 65(1), 85-101. doi:10.1037/0022-3514.65.1.85
- Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First impressions from faces among u.s. and culturally isolated tsimane' people in the bolivian rainforest. *Journal of Cross-Cultural Psychology*, 43(1), 119-134. doi:10.1177/0022022111411386

BIOGRAPHICAL SKETCH

Ms. Young has worked research for the Department of Defense for the last 24 years. She is a nationally recognized expert in blast injuries. In 2006, Ms. Young was the principal investigator on a test program for the Combating Terrorism Technology Support Office's (CTTSO) Technical Support Working Group (TSWG) that provided the first ever histological evidence of primary blast traumatic brain injuries in a very sub-lethal thoracic injury environment. She designed and served as the system integrator for the Quantico Breacher Injury Study, which was the first prospective study of blast on humans in the United States, and was a key developer of the ongoing Low Level Research Blast pre-clinical study for identifying mild traumatic brain injury thresholds under single and repeated exposure blasts (DoD-funded). Ms. Young served as the designer and technical director for the Office of Naval Research's Human Injury and Treatment program, a future naval capability program to develop a modeling and simulation tool capable of predicting the injuries and the associated incapacitation of injured personnel, medical response requirements and effectiveness, and overall mission impact associated with a blast-attack on a ship.

For the last three years, Ms. Young has been on a sabbatical, completing a doctorate in cognition and neuroscience, while providing limited consulting services as a subject matter expert in blast-induced injuries for DoD's Technical Support Working Group, a scientific advisor for Applied Research Associates, Inc. and a Senior Advisor to Vista Life Sciences, Inc. Although her dissertation topic was in the area of social neuroscience, her research while at The University of Texas at Dallas Center for Brain Health included, in equal parts, both social neuroscience and research into virtual-reality based approaches to characterize and remediate frontal lobe impairment caused by traumatic brain injuries.

CURRICULUM VITAE

Leanne R. Young

Education

Bachelor of Science: Mechanical Engineering	1992
University of Texas	Austin, TX
Masters of Arts: Applied Statistics	1995
University of New Mexico	Albuquerque, NM
Doctorate: Cognition and Neuroscience	(Defending) September 1, 2016
University of Texas at Dallas	Dallas, TX

Employment History

Senior Advisory Board Member	April 2015 to Present
Vista Life Sciences, Inc.	Parker, CO
Research Assistant	September 2013 to Present
University of Texas at Dallas Center for Brain Health	Dallas, TX
Cognition, Neuroscience, Blast Injury Consultant	September 2015 to Present
Clients: Applied Research Associates, Inc. Vista Life Sciences, Inc.	
Science and Technology Advisor	August 1992 to September 2015
Applied Research Associates, Inc.	Albuquerque, New Mexico
<i>Positions Held:</i>	Science and Technology Advisor (Jun 2006 to Sep 2015); Division Manager (Jul 2003 – Jun 2006); Senior Engineer (Jun 1998 – Jun 2003); Staff Engineer (Aug 1992 – Jun 1998).
<i>Responsibilities:</i>	<u>Division Manager (2003-2006)</u> – Personnel, Finance and Operations Management, Recruiting, Quality Control. Division grew from 6 to 14 people (Exceeded Growth Objectives). <u>Project Manager (1994-2015)</u> – Test Execution, Data Analysis, Modeling; <u>Marketing and Business Development (1994-2015)</u> – Customer Interactions, Strategic Planning,

Proposal Writing, Cross-Corporation Coordination, Multi-Disciplinary Collaborations.

Profile

Ms. Young has worked research for the Department of Defense for the last 24 years. She is a nationally recognized expert in blast injuries. In 2006, Ms. Young was the principal investigator on a test program for the Combating Terrorism Technology Support Office's (CTTSO) Technical Support Working Group (TSWG) that provided the first ever histological evidence of primary blast traumatic brain injuries in a very sub-lethal thoracic injury environment. She designed and served as the system integrator for the Quantico Breacher Injury Study, which was the first prospective study of blast on humans in the United States, and was a key developer of the ongoing Low Level Research Blast pre-clinical study for identifying mild traumatic brain injury thresholds under single and repeated exposure blasts (DoD-funded). Ms. Young served as the designer and technical director for the Office of Naval Research's Human Injury and Treatment program, a future naval capability program to develop a modeling and simulation tool capable of predicting the injuries and the associated incapacitation of injured personnel, medical response requirements and effectiveness, and overall mission impact associated with a blast-attack on a ship.

Currently, Ms. Young is on a sabbatical, pursuing a doctorate in cognition and neuroscience, while providing limited consulting services as a subject matter expert in blast-induced injuries for DoD's Technical Support Working Group and serving as a senior advisor to Vista Life Sciences, Inc. Her dissertation is in the area of social neuroscience, but her research while at the UTD Center for Brain Health has included, in equal parts, both social neuroscience and research into virtual-reality based approaches to characterize and remediate frontal lobe impairment caused by traumatic brain injuries. Ms. Young has completed all coursework toward the doctorate, and is scheduled to defend September 1, 2016.

Key Scientific Contributions

- 2003 - Ms. Young served as the principal investigator for a test program through the Combating Terrorism Technology Support Office's Technical Support Working Group designed to evaluate the effectiveness of tactical ballistic gear in a blast environment. During this study, improvised explosive device blast tests were performed on vests and helmets currently worn by first responders using anthropomorphic test devices, porcine specimens and cadavers. This effort provided two seminal advances in the area of blast bio-effects: (1) It was the first study to provide scientific (histological) evidence of primary blast

traumatic brain injuries in a very sub-lethal thoracic injury environment; and (2) It was the first study to experimentally identify the risk of behind armor blunt trauma caused by a blast impact to Level IV protective gear.

- 2006 through 2008 - Ms. Young served the Defense Advanced Research Projects Agency and the Office of Naval Research as the lead for the Quantico Breacher Injury Study, a program designed to evaluate US Marine breachers to determine and characterize the effects of repeated low-level blasts on the brain. This was the first prospective study ever performed on human subjects to quantify blast traumatic brain injuries. The results of this study have led to no fewer than six follow-on breacher research initiatives exploring the risk of cumulative neurological effects resulting from repeated blast exposures.
- 2009 - Ms. Young was contracted by the Office of Naval Research to lead the “Human Injury and Treatment” future naval capability program, which developed a modeling and simulation tool capable of predicting the injuries and the associated incapacitation of injured personnel, medical response requirements and effectiveness, and overall mission impact associated with a blast-attack on a ship. In addition to developing the modeling and simulation tool, this effort provided notable contributions to the science of blast injuries and combat casualty care, including the development of a military-centric injury severity scoring model and acquiring test data demonstrating the effect of high sea states on the ability of medical personnel to perform life-saving procedures.

Honors and Recognitions

Friends of Brain Health New Investigator Award 2014

Business Development Award - FY2009, FY2010, FY2011, FY2012, FY2013, FY2014 (Top business developer at ARA in 2014, and among the top 10 business developers since FY2010)

Robert E. Walker Team Player Award - FY2008

Harry E Auld Young Engineering Scientist Award - FY2001

Technical Achievement Award - FY1999

Selected Publications

Selected Peer-Reviewed Publications:

Vincent, A.S., Bailey, C.M., Cowan, C., Cox-Fuenzalida, E., Dyche, J., Gorgens, K.A., Krawczyk, D.C., Young, L.R. (2016). Normative data for evaluating mild traumatic brain injury with a handheld neurocognitive assessment tool. Appl Neuropsychol Adult. Accepted for publication.

- Carr, W., Stone, J. R., Walilko, T., Young, L. A., Snook, T. L., Paggi, M. E., Tsao, J. W., Jankosky, C. J., Parish, R. V., & Ahlers, S. T. (2016). Repeated low level blast exposure: A descriptive human subjects study. *Military Medicine*, 181, 28-39.
- Young, L., Rule G.T., Bocchieri, R.T., Walilko, T.J., Ling, G. "When Physics Meets Biology: Low and High-Velocity Penetration, Blunt Impact and Blast Injuries to the Brain" *Frontiers in Neurology* 6 (2015): 89.
- Needham, C.E., Ritzel, D., Rule, G.T., Wiri, S., Young, L. "Blast testing issues and TBI: experimental models that lead to wrong conclusions" *Frontiers in Neurology*. 6 (2015).
- Young, L.A., Rule, G.T., Bocchieri, R.T., Walilko, T.J., Burns, J.M. "Biophysical Mechanisms of Traumatic Brain Injuries" *Seminars in Neurology* 2015; 35(1):5-11
- Carr, W., Polejaeva E, Grome A, Crandall B, Lavalley, C, Eonta S, Young L "Relation of Repeated Low-level Blast Exposure with Symptomatology Similar to Concussion" *J Head Trauma Rehabil* 2015; 30(1):47-55.
- Lawnick, M, Champion HR, Gennarelli T, Galarneau MR, D'Souza E, Vickers RR, Wing V, Eastridge BJ, Young L, Dye J, Spott MA, Jenkins DH, Holcomb JR, Blackburne LH, Ficke JR, Kalin EJ, Flaherty S, "Combat Injury Coding: A Review and Reconfiguration," *J of Trauma*. 2013; 75(4):573-581.
- Jacobs, B. Young L, Champion HR, Lawnick M, Galarneau M, Wing V, Krebs WR, "Applying Modeling and Simulation to Predict the Effects of a Blast Attack on a Shipboard Environment," *Human Factors and Ergonomics Society's 56th Annual Meeting*, Boston, 22-26 Nov 2012.
- Walilko T, North C, Young L, Lux W, Warden D, Jaffee M, Moore D. Head Injury as a Predictor of PTSD among Survivors of the Oklahoma City Bombing. *J Trauma*. 2009;67:1311-1319.
- Champion, HR. Holcomb, JB. Young, LA Injuries from explosions: Physics, biophysics, pathology, and required research focus. *J of Trauma*. 66(5):1468-1477, May 2009.
- Walilko, T.J., LA Young, "Evaluation of the Effectiveness of Ballistic Protective Gear in a Blast Environment," *Applied Research Associates, Inc. Technical Review publication*, May 2006.
- Young, L.A., K Hilton, S Kirkpatrick, G Ogg, Thermobaric Hellfire Lethality Analysis, Final report prepared for Defense Threat Reduction Agency, April 2003.

Book Chapters:

- Wade, C.E., Ritenour, A.E., Eastridge, B.J., Young, L.A., Blackburne, L.H., Holcomb, J.B. (2008). Explosion Injuries Treated at Combat Support Hospitals in the Global War on Terrorism. In N.M. Elsayed and J.L. Atkins (Ed.), *Effects of Explosion and Blast from Military Operations and Acts of Terrorism* (pp. 41-70). Amsterdam: Elsevier.

Selected Presentations/Posters:

- Young, L.A. (April 2015). *Using the Virtual World to Understand the Real World of Brain Injury*. Presentation at the annual Feed the Mind luncheon, Center for Brain Health, Dallas, TX.
- Young, L.A., Rogers, J., Ling, G., Walilko, T.J. (2013, December). *Blast Gauge Threshold Settings for Detecting Risk of Blast-induced Neurotrauma*. Poster presented at the annual meeting of the Special Operations Medical Association, Tampa, FL.
- Maestas, F.A., LA Young, "Weapon Effectiveness Models: Are They Appropriate for Use in Force Protection Analyses?" SAFE Conferences Proceedings, Rome, Italy, June 2005.
- Young, L.A., KL Becvar, RL Moriarty, "Vehicle Bomb Mitigation Guide for the Personal Digital Assistant," Department of Defense Explosives Safety Board conference, August 2004.
- Young, L.A., DJ Stevens, SB Meyer, K Becvar, R Hayda, CR Bass, A Kersul, D Sunshine, Glass Shard Flyout and Penetration Model, 11th ISIEM Conference Proceedings, Mannheim, Germany, May 2003.