PRENATAL TESTOSTERONE, EMPATHY, EMOTION RECOGNITION, AND FACIAL MIMICRY IN WOMEN

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of
Master of Arts
in
Experimental and Applied Psychology
University of Regina

By
Emilie Gay Kossick
Regina, Saskatchewan
December, 2013
Copyright 2013: E. Kossick
Emilie Gay Kossick, candidate for the degree of Master of Arts in Experimental & Applied Psychology, has presented a thesis titled, *Prenatal Testosterone, Empathy, Emotion Recognition, and Facial Mimicry in Women*, in an oral examination held on December 16, 2013. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner: Dr. Kerri Staples, Faculty of Kinesiology & Health Studies

Co-Supervisor: Dr. Laurie Sykes Tottenham, Department of Psychology

Co-Supervisor: Dr. Dennis Alfano, Department of Psychology

Committee Member: Dr. Katherine Arbuthnott, Campion College

Committee Member: Dr. Christopher Oriet, Department of Psychology

Chair of Defense: Dr. Ronald Camp, Kenneth Levene Graduate School
Abstract

This study examined whether individual differences in facial mimicry, emotion recognition, and empathy are related to prenatal testosterone (PT) exposure. Previous studies have found relationships among facial mimicry, emotion recognition, and self-report empathy scores; further, imaging studies suggest that mimicry, emotion recognition, and empathy have shared neural bases. Previous evidence also suggests that the development of some of these shared brain regions is influenced by prenatal sex hormones (Goldstein et al., 2001), and a recent study demonstrated that exogenous testosterone administration decreases facial mimicry in women (Hermans, Putman, & van Honk, 2006). The present study examined the relationships among PT exposure (estimated using the 2D:4D ratio), facial mimicry, emotion recognition, and self-reported empathy in 53 women between the ages of 18 and 25 years. Facial mimicry was induced and emotion recognition was tested using a newly developed Facial Expression Viewing and Recognition Task (FEVRT). During the first block of the FEVRT participants passively viewed dynamic facial expressions of emotion presented on a computer screen; in the second block they viewed and identified the emotion that was displayed. During both blocks the participants’ spontaneous facial reactions were recorded covertly using a hidden video camera. The video recordings were assessed to determine the extent of participants’ facial mimicry of the FEVRT stimuli, using Ekman, Friesen, and Hager’s Facial Action Coding System (FACS; 2002). The left 2D:4D ratio significantly predicted facial mimicry intensity, emotional empathy, and emotion recognition accuracy scores. For all of the relationships, higher estimated PT levels were associated with lower scores on the emotion measures. No relationships were found between mimicry and empathy or
between mimicry and emotion recognition accuracy. Mediating relationships were not observed amongst the 2D:4D ratio and mimicry, empathy, or emotion recognition scores. These findings suggest that mimicry, emotion recognition, and empathy are distinct processes, which are each independently influenced by PT. The findings support the hypothesis that prenatal sex hormones have an effect on adult emotion processing. By understanding the role that the prenatal sex hormone environment has on the development of emotion processing, we can gain a better understanding of how individual differences in emotion processing arise.
Acknowledgements

I would like to extend my heartfelt thanks to my supervisors, Dr. L. Sykes Tottenham and Dr. D. Alfano, without whose advice and assistance this thesis would not be possible. I would also like to thank my supervisory committee, Dr. C. Oriet and Dr. K. Arbuthnott, who helped to bring this thesis to fruition. My deepest thanks are also extended to my lab mates, especially Denis Gavigan who helped with scoring. Lastly, I would like to thank the University of Regina Faculty of Graduate Studies and Research for their funding support through 3 teaching assistantships and one graduate research assistantship.
Post Defense Acknowledgements

I would like to thank my external examiner, Dr. K. Staples, for her time and insight, and my chair, Dr. R. Camp, for making my defense a positive experience.
Dedication

I would like to thank my friends who provided me with willing ears and shoulders, in addition to much needed perspective. And I would like to dedicate this thesis to my mother who was there through my good days and my bad, and offered reassurance when there was doubt. She, as always, stuck with me as I stuck with this project. Mom you have my unending love and respect.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Post Defense Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>Dedication</td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xi</td>
</tr>
<tr>
<td>1. OVERVIEW</td>
<td>1</td>
</tr>
<tr>
<td>2. LITERATURE REVIEW</td>
<td>2</td>
</tr>
<tr>
<td>2.1. Facial Mimicry, Empathy, and Emotion Recognition</td>
<td>2</td>
</tr>
<tr>
<td>2.2. Neural Correlates of Empathy, Facial Mimicry, and</td>
<td>8</td>
</tr>
<tr>
<td>Emotion Recognition</td>
<td></td>
</tr>
<tr>
<td>2.3. Sex Differences and Sex Hormone influences on Neural Regions</td>
<td>11</td>
</tr>
<tr>
<td>underlying Facial Mimicry, Empathy, and Emotion Recognition</td>
<td></td>
</tr>
<tr>
<td>2.4. Sex and Sex Hormone-Related Differences in Empathy and</td>
<td>13</td>
</tr>
<tr>
<td>Mimicry</td>
<td></td>
</tr>
<tr>
<td>2.5. Prenatal Sex Hormone Measures</td>
<td>16</td>
</tr>
<tr>
<td>3. PURPOSE AND HYPOTHESES</td>
<td>19</td>
</tr>
<tr>
<td>4. METHOD</td>
<td>21</td>
</tr>
<tr>
<td>4.1. Participants</td>
<td>21</td>
</tr>
<tr>
<td>4.2. Measures</td>
<td>21</td>
</tr>
</tbody>
</table>
5.5.2. Hypothesis 2: 2D:4D Ratio and Empathy ...........................................38
5.5.3. Hypothesis 3: 2D:4D Ratio and Emotion Recognition .................39
5.5.4. Hypothesis 4: Facial Mimicry and Empathy .................................39
5.5.5. Hypothesis 5: Facial Mimicry and Emotion Recognition .............40
5.5.6. Hypothesis 6: Mediating Relationships ..................................41

6. DISCUSSION .........................................................................................47

6.1. Prenatal Testosterone and Emotion Processing ..........................47

   6.1.1. Prenatal Testosterone and Mimicry during Active versus Passive
          Emotion Viewing ........................................................................51

   6.1.2. Prenatal Testosterone and Behavioural versus Self-Report
          Emotion Measures ......................................................................51

6.2. Relationships among Mimicry, Empathy, and Emotion Recognition ..........................................................53

6.3. Mediating Relationships among Prenatal Testosterone and Emotion Measures ..............................................................57

6.4. Limitations .......................................................................................57

6.5. Future Directions ...........................................................................60

6.6. Conclusion .......................................................................................61

7. REFERENCES .....................................................................................63

8. APPENDIX ..........................................................................................73
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Procedural flow chart</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Outline of mediation analysis</td>
<td>43</td>
</tr>
<tr>
<td>Figure 3a</td>
<td>Mediation analysis for 2D:4D, AU mimicry, and empathy</td>
<td>44</td>
</tr>
<tr>
<td>Figure 3b</td>
<td>Mediation analysis for 2D:4D, mimicry intensity, and empathy</td>
<td>44</td>
</tr>
<tr>
<td>Figure 4a</td>
<td>Mediation analysis for 2D:4D, AU mimicry, and emotion recognition</td>
<td>45</td>
</tr>
<tr>
<td>Figure 4b</td>
<td>Mediation analysis for 2D:4D, mimicry intensity, and emotion recognition</td>
<td>45</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Mediation analysis for 2D:4D, emotion recognition, and BEES scores</td>
<td>46</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>2D</td>
<td>Second digit</td>
<td></td>
</tr>
<tr>
<td>2D:4D</td>
<td>Second to fourth digit ratio</td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>Fourth digit</td>
<td></td>
</tr>
<tr>
<td>AIS</td>
<td>Androgen insensitivity syndrome</td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>Action Unit</td>
<td></td>
</tr>
<tr>
<td>BEES</td>
<td>Balanced Emotional Empathy Scale</td>
<td></td>
</tr>
<tr>
<td>CAH</td>
<td>Congenital adrenal hyperplasia</td>
<td></td>
</tr>
<tr>
<td>CES-D</td>
<td>Centre for Epidemiologic Studies Depression Scale</td>
<td></td>
</tr>
<tr>
<td>EMG</td>
<td>Electromyography</td>
<td></td>
</tr>
<tr>
<td>EQ</td>
<td>Empathy Quotient</td>
<td></td>
</tr>
<tr>
<td>FACS</td>
<td>Facial Action Coding system</td>
<td></td>
</tr>
<tr>
<td>FEVRT</td>
<td>Facial expression viewing and recognition task</td>
<td></td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional Magnetic Resonance Imaging</td>
<td></td>
</tr>
<tr>
<td>IFG</td>
<td>Inferior Frontal Gyrus</td>
<td></td>
</tr>
<tr>
<td>IFO</td>
<td>Inferior Frontal Operculum</td>
<td></td>
</tr>
<tr>
<td>JACFEE-</td>
<td>Japanese and Caucasian Facial Expressions of Emotion</td>
<td></td>
</tr>
<tr>
<td>JACNeuF-</td>
<td>Japanese and Caucasian Neutral Faces</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td>PAM</td>
<td>Perception Action Mechanism</td>
<td></td>
</tr>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>Prenatal testosterone</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>TAS-20</td>
<td>Toronto Alexithymia Scale</td>
<td></td>
</tr>
<tr>
<td>VIF</td>
<td>Variance Inflation Factor</td>
<td></td>
</tr>
<tr>
<td>WHQ-R</td>
<td>Waterloo Handedness Questionnaire-Revised</td>
<td></td>
</tr>
</tbody>
</table>
Empathy is the ability to understand another person’s mental state. This ability is critically important to social functioning because it allows a person to establish a sympathetic relationship with another person, based on mutual understanding (de Waal, 2008). Empathy occurs on both cognitive and emotional levels, meaning that an observer understands both the thoughts and emotions of another person (Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004). Emotional empathy encompasses the initial response to another person’s mental state – the acts of observing, mimicking, and experiencing another person’s emotions (Bons et al., 2012; Nummenmaa, Hirvonen, Parkkola, & Hietanen, 2008). Cognitive empathy involves recognizing and understanding another person’s mental state (Bons et al., 2012; Nummenmaa et al., 2008). In combination, emotional and cognitive aspects of empathy allow humans to share needs and experiences with others.

It is important to know how empathy develops and how individual differences in empathy arise in order to understand empathy as both emotional and cognitive abilities. Recent evidence suggests that sex hormones present in the prenatal environment affect the development of neural regions underlying empathy (Goldstein et al., 2001; Phan, Wager, Liberzon, & Taylor, 2002), and some studies have found a relationship between prenatal testosterone (PT) exposure and behavioural measures of cognitive empathy in children (Chapman et al., 2006; Knickmeyer, Baron-Cohen, Raggatt, Taylor, & Hackett, 2006). Collectively, these studies suggest that developmental differences in the brain,
caused by differences in the prenatal sex hormone environment, could be one source of individual differences in cognitive empathy, at least in children. The purposes of the present study were to determine: 1. whether the previously observed relationship between PT and cognitive empathy in children could be observed in adults, and 2. whether the relationship extends to emotional empathy. Specifically, using both behavioural and self-report measures, the present study examined the relationships among an index of PT exposure, facial mimicry, emotion recognition, and empathy in women.

2. LITERATURE REVIEW

2.1. Facial Mimicry, Empathy, and Emotion Recognition

Many lines of research suggest that facial mimicry facilitates empathy (Blairy, Herrera, & Hess, 1999; de Waal, 2008). Facial mimicry may facilitate empathy by allowing a person to feel what another person is feeling, or by simply allowing a neural representation to be formed of the emotion (without the experience of feeling it). Further, mimicry appears to facilitate emotion recognition (e.g., Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001). Research examining these relationships is reviewed below.

The association between facial mimicry and empathy was first hypothesized by Lipps (1903, as cited in Blairy et al., 1999), who argued that facial mimicry was a key element of understanding how an observer could relate to and understand the experience of another. Lipps proposed a three step model: 1. the observer sees a facial expression, 2. that expression is mimicked leading the observer to create the affect within himself or herself, and 3. understanding of the subject’s state of mind (empathy) occurs because the observer is now feeling the same affect.
Blairy et al. (1999) tested Lipps’ theory using photographs selected from Matsumoto and Ekman’s (1988) datasets. Blairy et al. had participants evaluate facial expressions of emotion to identify the emotion displayed, and describe the feelings they experienced while observing the picture. Mimicry was measured using Electromyography (EMG) while participants viewed the facial expressions. The results of their study indicated that, although mimicking occurred and appeared to facilitate emotion recognition, it did not lead the participant to feel the emotion, as Lipps had suggested. Instead, this evidence might better support the theory put forth by de Waal (2008), called the Perception Action Mechanism (PAM), which suggests that the observer does not feel the emotion, but rather has a neural representation of similar states that are automatically activated. This neural representation leads to understanding, but not experiencing, the emotion (de Waal, 2008).

Evidence for the importance of facial and gestural mimicry in everyday interactions was demonstrated by Chartrand and Bargh (1999). In their studies, confederates worked with participants. During the first testing session a confederate would display various behaviours, either rubbing his or her face or shaking his or her leg, and either smiling or not smiling; in the second session a second confederate would display actions opposite to that which the previous confederate displayed (e.g., if the first confederate had smiled and shook his or her leg throughout the testing session, the second confederate would not smile and would rub his or her face). Both sessions were video-recorded. Analysis of the video recordings showed that participants mimicked the behaviours the confederates had displayed in both sessions. Chartrand and Bargh also administered a self-report measure of cognitive empathy that involved perspective taking.
Scores on this measure were associated with the frequency of mimicked behavior, such that participants who scored high on the perspective taking measure mimicked the mannerisms of the confederates more frequently than participants who scored low on the measure. In the second study, Chartrand and Bargh, studied the adaptive function of mimicry experimentally. In the experimental condition a confederate mimicked the participants’ behaviours throughout the session, while in the control condition the confederate did not mimic the participants. After the session, participants were asked to rate how much they liked the confederate and how smoothly they thought the interaction had gone. Participants who had experienced mimicking from the confederate rated the confederate and the experience more positively than those in the control condition. The results of these studies suggest that mimicry is associated with cognitive empathy, interpersonal affinity, and the perceived quality of social interactions.

Sato and Yoshikawa (2007) measured facial mimicry in a situation in which participants were asked to passively view emotions. They hypothesized that facial mimicry would occur even when participants were not asked to identify a specific emotion. Participants were video-recorded while passively watching dynamic facial expressions of anger and happiness. These dynamic expressions were created using computer morphing technology. Each dynamic expression was created using 26 intermediate photographs, progressing from a neutral expression to an emotional expression, displayed in sequence as a video clip. Participants were asked to view the stimuli, and were unknowingly recorded while they viewed them. The recordings of participants’ facial mimicry responses were then evaluated using the Facial Action Coding System (FACS). The FACS is a system in which facial expressions are
objectively coded into 30 action units (AU) and 14 miscellaneous actions, and analyzed for intensity (Ekman, Friesen, & Hager, 2002). Although participants did not have to identify the emotion, results showed that participants mimicked both angry and happy emotional facial expressions, suggesting that mimicry happens even during passive viewing of emotional stimuli.

Sonnby-Borgstrom, Jonsson, and Svensson (2003) also explored the relationship between empathy and facial mimicry. In their study participants’ facial muscle movements were recorded using EMG while they were shown several faces displaying happy and angry expressions. Following this, participants were given a self-report measure of empathy (the Questionnaire Measure of Emotional Empathy; Mehrabian & Epstein, 1972). Results showed that all participants displayed some mimicking behaviour, but those with high self-reported empathy scores had higher levels of facial mimicry.

Emotion recognition tasks are often used in studies of empathy and facial mimicry because, it is argued that, in order to understand another person’s affective state one has to have the basic ability to identify that state (Clark, Winkielman, & McIntosh, 2008). In order to study the role of facial mimicry in emotion recognition, Niedenthal and colleagues (2001) had participants view video clips of faces transitioning from happy to sad expressions, and vice versa. Participants were asked to indicate when the first expression (either happy or sad) disappeared. One half of the participants were asked to view the clips while holding a pen between their teeth and lips, whereas the other participants were allowed to behave naturally while watching the clips. It was hypothesized that by blocking the participants’ ability to mimic expressions (by having
them hold the pen) their ability to detect the emotional change in expression would be adversely affected. The results showed that those participants holding the pen were significantly slower at recognizing when the first expression disappeared. These results suggest that facial mimicry facilitates the quick identification of changes in another’s emotional expression.

Niedenthal and colleagues’ study (2001) was extended by Oberman, Winkielman, and Ramachandran (2007) to determine whether blocking mimicry impairs recognition of all emotions or specific emotions. In the first part of the experiment, Oberman et al. instructed participants to produce facial expressions in response to both verbal and visual cues. Participants were divided into four groups to test different methods of blocking mimicry: 1. holding a pen with one’s lips (Lip), 2. chewing gum (Gum), 3. holding a pen between one’s teeth (Bite), and 4. no manipulation (Control). From this part of the experiment they determined that the “Bite” condition caused the most blocking of mimicry, as measured using EMG. In the second part of the experiment they had participants view pictures of facial emotional expressions from Ekman and Friesen’s 1976 database, which were edited to display emotions at varying intensities. These pictures were edited using computer morphing software to create seven distinct intensity levels of each emotional expression (happy, sad, fearful, and disgusted). For each picture, participants were asked to indicate which of four emotions was expressed by the person shown, while performing one of the three mimicry blocking conditions or the control condition described above. In the case of the “Bite” condition, which in part 1 caused the most blocking of mimicry, participants’ recognition of happy and disgusted expressions was significantly reduced compared to the other conditions, while
recognition of sad and fearful expressions remained the same across conditions. Oberman and colleagues suggested that this result was not surprising because happy and disgusted expressions require more facial movement than expressions of sadness and fear, suggesting that the blocking of facial mimicry allowed for less facial feedback, subsequently leading to poorer recognition of these emotions in particular.

Stel and Knippenberg (2008) further examined the relationship between facial mimicry and emotion recognition. Specifically, they sought to determine whether facial mimicry would affect the speed of emotion recognition in addition to affecting recognition accuracy. Stel and Knippenberg proposed that individuals who are more empathetic may be so because they respond quickly to the expressed emotions of others. Following from this they hypothesized that participants restrained from mimicry would be slower to recognize emotional expressions, and that women’s recognition of the expressions would be more adversely affected by the restraint than men’s, because women demonstrate more empathy than men. Participants viewed photographs of positive and negative facial expressions chosen from Ekman and Friesen’s (1976) and Matsumoto and Ekman’s (1988) databases. Participants were asked to indicate whether the emotion they saw was negative or positive. Half of the participants were asked to avoid facial movements and had a large plaster bandage placed on their foreheads to provide physical feedback about involuntary facial movements, whereas the other half were allowed to view the photographs without any facial movement restrictions. The results showed that only women’s response times were adversely affected by the instruction to avoid facial movements. Overall, the results suggest that mimicry was more important for women’s ability to classify emotions quickly compared to men.
Together, the studies discussed above suggest that there is a relationship between empathy and facial mimicry – participants who have higher self-reported empathy scores also show more facial mimicry. Further the studies suggest that facial mimicry improves the accuracy and speed of emotion recognition, and also demonstrate that mimicry is displayed when a participant is passively viewing emotional expressions.

2.2. Neural Correlates of Empathy, Facial Mimicry, and Emotion Recognition

A number of studies have suggested that sex hormones may affect empathy. In order to understand why sex hormones might affect empathy, one has to understand which brain regions are involved in the processes underlying empathy and how these regions are affected by sex hormones. As described earlier, these processes include facial mimicry and emotion recognition.

Some researchers suggest that a specific class of neurons may underlie the empathetic response. A group of Italian neurophysiologists discovered an interesting class of brain cells in macaque monkeys (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). These brain cells, located in the inferior frontal cortex, fired when a monkey was mimicking another monkey’s behaviour. Surprisingly, the cells also fired when the monkey was simply observing the behaviour of another monkey. The term ‘mirror neuron’ was coined to reflect the function of these neurons. The discovery of these neurons in monkeys led to studies in humans; researchers posited that mirror neurons found in monkeys might be found in humans as well (Iacoboni, 2007) and that these neurons may underlie facial and gestural mimicry (Chartrand & Bargh, 1999).

In order to test for the presence of mirror neurons in the human brain, Carr, Iacoboni, Dubeau, Mazziotta, and Lenzi (2003) used fMRI to examine brain activity
during observation and imitation tasks. Participants were presented with 36 trials of emotionally expressive faces, with each face displaying one of six basic emotions. Participants were asked to imitate and internally generate the emotion they were viewing, or to simply observe the emotion. Carr et al. found similar patterns of brain activation during active imitation and internal generation of the emotion, and during simple observations of the emotional expressions (without imitation or internal generation), suggesting the presence of mirror neurons in humans.

A study by Leslie, Johnson-Frey, and Grafton (2004) used fMRI to examine neural activation while participants watched videos of expressive faces and hands, and while they performed a control task which required them to generate verbs when a noun was presented. For the experimental tasks, participants were asked to: 1. passively watch a movie of a model (View task); 2. watch a movie of a model and imitate the movement made by the model (Imitate task); and 3. either always smile or always frown regardless of what the model in the movie was doing (Move Task). In line with the findings of Carr et al. (2003), Leslie and colleagues found similar patterns of neural activation for imitation and observation of face and hand movements, though for observation the activation was in the right hemisphere whereas imitation produced bilateral activation. Both Carr et al. and Leslie et al. found that brain activation during observation and imitation occurred in the dorsal premotor cortex, ventral premotor cortex, and the inferior frontal lobe in both hemispheres. Leslie et al. also found that simple observation of others yielded a greater right hemisphere activation and greater activation of the sensorimotor cortex.
Lee, Josephs, Dolan, and Critchley (2006) examined patterns of neural activation associated with emotional and non-emotional facial mimicry. In their study, fMRI was used to examine neural activation in 18 participants while they actively imitated emotional and non-emotional (control) facial expressions displayed in video clips. The three emotions they were asked to mimic were anger, sadness, and happiness, and the control conditions were chewing and licking. Brain activity was increased in the right inferior frontal gyrus (IFG) and right insula for all three emotion conditions compared to the control conditions.

Jabbi and Keysers (2008) used fMRI to study the activity of the inferior frontal gyrus (IFG), and the anterior insula and adjacent frontal operculum (collectively referred to as the inferior frontal operculum, IFO) while participants passively viewed a video of an actor making either a disgusted or neutral face. These areas are hypothesized to be important in empathetic behaviours because they are more active in individuals who self-report having a higher level of empathetic concern for the distress of others (Jabbi & Keysers, 2008). Jabbi and Keysers hypothesized that the IFO might be controlled by the IFG, which appears to underlie facial mimicry. They found that activity in the IFG and IFO increased when viewing disgusted expressions. Because the IFG has been shown to be activated when facial expressions are produced, Jabbi and Keysers hypothesized that humans view facial expressions and mimic them, and that as a result of mimicry we activate that emotion internally so that it is experienced. Therefore, the IFG is activated when mimicking the facial expression, which in turn activates the IFO, leading to the “sharing” of that emotion.
Adolphs, Damasio, Tranel, Cooper, and Damasio (2000) conducted a large scale analysis of 108 patients who had lesions in differing regions of the brain to examine the neural correlates of emotion recognition. Each patient was asked to complete three different tasks: identify emotions (happiness, anger, fear, disgust, sadness, and surprise) from pictures of emotionally expressive faces, assign emotional labels from a selected list to pictures of emotionally expressive faces, and sort photographs of emotionally expressive faces into categories based on the emotion shown. Patients who performed poorly on all three tasks had damage in the right somatosensory cortices (specifically the S-I, S-II) and the insula, suggesting that these areas are important for recognition of emotions in human facial expressions (Adolphs et al., 2000). Further, fMRI and PET studies have shown that the anterior insula and amygdala are activated when participants are asked to recognize emotional facial expressions of others or evaluate emotional stimuli (Iacoboni, 2007; Phan et al., 2002). The medial prefrontal cortex is also commonly activated during tasks associated with emotion processing, including emotion recognition (Phan et al., 2002).

Collectively, these studies suggest that structures in the fronto-limbic system, including the premotor cortex, IFG, IFO, insula, and amygdala, play important roles in emotion mimicry, empathy, and emotion recognition.

2.3. Sex Differences and Sex Hormone influences on Neural Regions underlying Facial mimicry, Empathy, and Emotion Recognition

Of relevance to the current study, many of the brain regions that underlie mimicry, empathy, and emotion recognition are sexually dimorphic in volume in adults, and many have a high density of sex hormone receptors during early development,
especially the premotor cortex, IFG, IFO, insula, and amygdala (Goldstein et al., 2001). The presence of sex hormone receptors allows sex hormones to influence the structural and functional development of these brain regions. These are classified as organizational sex hormone effects because they are permanent changes in the brain that occur during critical periods of development. In contrast, sex hormones also have transient effects on the brain and behaviour throughout life that occur as hormone concentrations fluctuate, which are known as activational effects (for review see: Schulz, Molenda-Figueira, & Sisk, 2009). The mechanism that allows sex hormones to have organizational effects on brain development is somewhat unclear; however, both estrogens and androgens appear to influence brain development in humans through numerous genomic and nongenomic effects (for review see: Falkenstein, Tillman, Christ, Feuring, & Wehling, 2000). Genomic effects are displayed in the synthesis, release, and metabolism of many neuropeptides essential to central nervous system regulation (Genazzani, Monteleone, & Gambacciani, 2002; Romeo & McEwen, 2004). Sex hormones also have non-genomic effects, such as regulating the permeability of ion channels in the cell membrane (Romeo & McEwen, 2004). With the multiple effects of sex hormones at the cellular level, it is apparent that prenatal sex hormones can have long term effects on brain structure (Goldstein et al., 2001).

In addition to structural sex differences, the neural regions that underlie empathy also demonstrate functional sex differences. In an fMRI study, Schulte-Ruther, Markowitsch, Shah, Fink, and Piefke (2008) presented participants with facial expressions of angry and fearful emotions and asked participants to complete two tasks: 1. indicate the emotional state of the presented expression (OTHER), and 2. indicate their
own emotional response to the expression (SELF). They were also given questionnaires, which included a self-report measure of empathy called the Balanced Emotional Empathy Scale (BEES; Mehrabian, 2000). The results showed increased brain activity in the IFG during viewing of facial expressions compared to a baseline measure. Scores on the BEES (in both sexes) were positively correlated with IFG activity. The BEES also showed a sex difference with females scoring higher than males. Activation in the IFG was significantly higher for females, particularly in the right IFG in both the SELF and OTHER condition compared to males. In contrast males showed higher left IFG activation in the SELF condition compared to the OTHER condition, although not as high as that seen in the female participants. Such response differences suggest that there are sex differences in the way emotion is evaluated and experienced.

2.4. Sex and Sex Hormone-Related Differences in Empathy and Mimicry

Most self-report measures of empathy demonstrate a consistent sex difference, with females scoring higher than males (Eisenberg & Lennon, 1983); such measures include Baron-Cohen’s Empathy Quotient (Lawrence et al., 2004), which measures both cognitive and emotional empathy, and the aforementioned BEES (Mehrabian, 2000), which measures emotional empathy. However, the sex difference in empathy is not limited to self-report measures. Knickmeyer et al. (2006) found that girls used more affective terms on a behavioural measure of cognitive empathy, which required them to describe social relationships in abstract cartoons. A sex difference in emotional empathy was also suggested by Dimberg and Lundquist (1990), who found that facial mimicry is stronger in females compared to males.
One possible reason for the sex difference in empathy could be varying levels of prenatal sex hormone exposure in males and females (Knickmeyer & Baron-Cohen, 2006). Research suggests that hormonally influenced differences observed in brain structure are related to later differences in behaviours and abilities. Knickmeyer, Baron-Cohen, Raggatt, and Taylor (2005) examined four-year-old children and found that lower levels of PT (as measured by amniocentesis) were associated with better quality social relationships and a broader range of social interests, as assessed by parental report. Chapman et al. (2006) examined this relationship in children ages 6-9 years by examining PT (previously determined through amniocentesis) and administering two measures of empathy – a self-report measure (the EQ) and a behavioural measure (“Reading the Mind in the Eyes”). They found that boys scored lower on the measures compared to girls, and that PT levels were negatively associated with empathy levels in boys. Boys who had lower PT scored higher on both the self-report and behavioural empathy measures.

The relationship between PT and emotional abilities has also been studied in clinical populations. Manning, Baron-Cohen, Wheelwright, and Sanders (2001) studied 72 children with autism spectrum disorders. Autism spectrum disorders are characterized by poor emotional and social functioning and show a strong sex difference, with males diagnosed four times more than females. Using an estimator of PT exposure (the 2D:4D ratio, which is the ratio between the lengths of the 2nd and 4th digits of each hand), Manning et al. (2001) found that children with autism spectrum disorders had significantly lower 2D:4D ratios than same-sex controls, suggesting a relatively higher presence of PT. Similar results were found by de Bruin, Verheij, Wiegman, and
Ferdinand (2006) in their study of a large spectrum of developmental disorders, including autism and ADHD. While these findings suggest that relatively higher PT exposure is associated with disorders characterized by poor social functioning, these findings may not generalize to nonclinical populations.

Collectively, these studies suggest that PT is associated with empathy and social-emotional processing. This relationship between PT and emotional abilities is not surprising because, as outlined above, brain regions underlying empathy, mimicry, and emotion recognition have high densities of sex hormone receptors during early development (Goldstein et al., 2001).

As previously mentioned, facial mimicry is related to empathy (Carr et al., 2003; Chartrand & Bargh, 1999; Iacoboni, 2007), and there is a relationship between prenatal sex hormones and empathy in children (Knickmeyer et al., 2006). However, few studies have investigated the relationship between facial mimicry and sex hormone concentrations, and none have examined prenatal sex hormone effects in particular. Hermans, Putman, and van Honk (2006) tested whether circulating testosterone had an effect on facial mimicry. In their study, exogenous testosterone was administered sublingually to a group of female students; the experiment was conducted over two days, and participants served as their own controls. During one testing session participants received testosterone, and during the other they were given a placebo. The participants were asked to observe dynamic facial expressions of anger and happiness, which were made using computer morphing of still photographs, creating a video clip. Movements of participants’ facial muscles were recorded using EMG while they were examining the dynamic expressions. Testing of the participants was restricted to the pre-ovulatory phase.
of the menstrual cycle, when estrogen is high and progesterone is low; further, to limit the influence of daily testosterone fluctuations, all testing was conducted in the afternoon. Results showed that when participants were administered testosterone they showed less facial mimicry than when they were given a placebo, supporting the hypothesis that circulating testosterone decreases the intensity of facial mimicry. If facial mimicry is a component of emotional empathy, these results may further suggest that testosterone decreases emotional empathy, and perhaps even cognitive empathy. Hermans and colleagues’ study shows that circulating testosterone has activating effects on facial mimicry, but it is still unknown whether there is an organizational effect that precedes this relationship.

2.5. Prenatal Sex Hormone Measures

There are a number of ways in which prenatal sex hormone concentrations can be measured or estimated (for review see: Cohen-Bendahan, van de Beek, & Berenbaum, 2005). The most direct measures of prenatal sex hormone concentrations occur during pregnancy. These are tests involving blood samples taken from the umbilical cord or the mother, and amniocentesis, in which amniotic fluid is sampled and tested. Because the present study examined an adult population in which such data were unavailable, these measures could not be used.

Another common way to study the effects of prenatal sex hormone exposure is by examining clinical populations of individuals who have experienced atypical endogenous sex hormone activity, such as those who have congenital adrenal hyperplasia (CAH) or androgen insensitivity syndrome (AIS). CAH is caused by a 21-hydroxylase deficiency, which causes high levels of adrenal androgen production (Collaer & Hines, 1995). AIS is
characterized by a lack of androgen receptors (Hines, Ahmed, & Hughes, 2003). Other clinical populations include those who have been exposed to exogenous hormones in utero, often due to drugs given to the mother during pregnancy (Reinisch, 1981). Although studies involving samples from these populations produce strong support for prenatal sex hormone influences on behaviour, it is unclear if the findings generalize to a normal population. These clinical populations are also rare, making samples difficult to obtain for research studies.

Indirect measures are frequently used to study prenatal sex hormone exposure in normal populations. Such measures include: otoacoustic emissions, fingerprint analyses, and the 2D:4D ratio. Otoacoustic emissions are sounds produced in the inner ear, likely through displacement of sensory hair cells (Cohen-Bendahan et al., 2005). These emissions show distinct sex differences, with women experiencing them more than men (McFadden, 1998). While most evidence suggests that sex differences in otoacoustic emissions are caused by genetic differences, there are some studies that suggest that prenatal sex hormone levels also have an effect (McFadden, Loehlin, & Pasanen, 1996). However, otoacoustic emissions are not often used as estimators of prenatal sex hormone exposure because of the specialized equipment required to record them, and the inconsistent findings concerning the relationship between the emissions and hormones (Cohen-Bendahan et al., 2005). Fingerprint analyses are also thought to be an indirect measure of prenatal sex hormone exposure, since ridge count and ridge count asymmetry between hands show consistent sex differences. However, studies have shown that these differences are likely a result of genetic differences between men and women, not a result of the prenatal sex hormones present during pregnancy (Cohen-Bendahan et al., 2005).
Since there is not much data to support their relationship to prenatal sex hormones, these measures are rarely used.

The most commonly used indirect measure of prenatal sex hormone exposure is the 2D:4D ratio – the ratio between the lengths of the 2nd and 4th digits of each hand. Lutchmaya, Baron-Cohen, Raggatt, Knickmeyer, and Manning (2004) used amniocentesis and demonstrated that a relatively low 2D:4D ratio is associated with a relatively high uterine PT level and a relatively low uterine prenatal estradiol (PE) level, whereas high 2D:4D ratios are associated with low PT and high PE in utero. A low 2D:4D ratio has also been observed in females who suffer from CAH, providing further evidence for its relationship with PT levels (e.g., Puts, McDaniel, Jordan, & Breedlove, 2008). The 2D:4D ratio is a reliable measure that has been widely used to examine the influence of PT on a large array of abilities and traits that demonstrate sex differences and other behaviours, including depression in men (Bailey & Hurd, 2005), sexual orientation and spatial abilities in women (van Anders & Hampson, 2005), and self perceived attractiveness in women (Wade, Shanley, & Imm, 2004). Further, the sexual dimorphism in the 2D:4D ratio is consistently observed within differing age and ethnic groups, although there are differences in this ratio between ethnic and age groups (Manning, 2002; Manning & Quinton, 2007). Because the 2D:4D ratio is a widely used and easy to obtain index of prenatal sex hormone exposure that can be used in a normal adult population, it was chosen for use in the present study.
3. PURPOSE AND HYPOTHESES

The studies reviewed above demonstrate that numerous brain regions underlying empathy, facial mimicry, and emotion recognition contain sex hormone receptors, and that structural and functional sex differences are characteristics of these regions. Additionally, many studies have found performance differences between males and females on empathy, facial mimicry, and emotion recognition tasks, and many studies show that empathy, facial mimicry, and emotion recognition are related to one another. Importantly, a few studies have suggested that prenatal testosterone exposure influences cognitive empathy in children (Chapman et al., 2006; Knickmeyer et al., 2005; Knickmeyer et al., 2006), and others have shown that exogenous testosterone administration decreases facial mimicry and emotion recognition in women (Hermans et al., 2006; van Honk & Schutter, 2007), however, the relationships among prenatal sex hormone exposure, empathy, facial mimicry, and emotion recognition in adults have not yet been examined. As such, the present study examined relationships among a PT index (the 2D:4D ratio) and both emotional and cognitive empathy by examining facial mimicry, emotion recognition, and self-reported empathy in women.

Only women were examined in the present study because women have been previously shown to have higher levels of self-reported empathy, mimicry, and emotion recognition performance compared to men, increasing the likelihood that relationships between PT and the empathy measures would be observable. In addition, the present study was, in part, motivated by Hermans and colleagues’ 2006 study, which examined the influence of exogenous testosterone administration on facial mimicry in an exclusively female sample.
Based on the evidence reviewed above it was hypothesized that:

1) Women with lower 2D:4D ratios (higher PT) will display less facial mimicry while observing dynamic facial expressions than women with higher 2D:4D ratios (lower PT).

2) Women with lower 2D:4D ratios (higher PT) will score lower on self-report measures of empathy than women with higher 2D:4D ratios (lower PT).

3) Women with lower 2D:4D ratios (higher PT) will have lower emotion recognition accuracy scores than women with higher 2D:4D ratios (lower PT).

4) Facial mimicry scores will predict scores on self-report measures of empathy, such that as facial mimicry scores increase empathy scores will increase.

5) Facial mimicry scores will predict emotion recognition accuracy scores, such that as facial mimicry scores increase emotion recognition accuracy will increase.

6) If the above relationships exist, then mimicry and emotion recognition are expected to mediate the relationship between the 2D:4D ratio and self-reported empathy, and mimicry is expected to mediate the relationship between the 2D:4D ratio and emotion recognition.
4. METHOD

4.1. Participants

Ethics approval was received from the University of Regina Research Ethics Board prior to the start of data collection. Participants were recruited from the University of Regina Department of Psychology Research Participant Pool and received one credit towards a first or second year psychology course for their participation. The sample consisted of 53 right-handed, Caucasian females between the ages of 18-25 years ($M = 20.47$ years). All participants had no previous major injuries or breakages of the index or ring fingers (as injury would affect the 2D:4D ratio). In addition, participants had normal or corrected-to-normal vision (to ensure the stimuli could be properly viewed), and were fluent in English. All participants were tested during the menstrual phase to limit between-subject variability due to activational hormone effects that occur at different stages of the menstrual cycle.

4.2. Measures

4.2.1. 2D:4D Ratio. The 2D:4D ratio was used as an estimate of PT exposure. Photos were taken of participants’ hands during the test procedure. After data collection was complete Corel Paint Shop Pro Photo X2 was used to view the photos and obtain measurements. Measuring pixels on a computer has demonstrated the highest interrater agreement and lowest level of equipment error of various 2D:4D measurement approaches (Kemper & Schwerdtfeger, 2009). The second (2D) and fourth (4D) digits were measured on the ventral surface from the tip of the finger to the basal crease. The 2D:4D ratio was computed by dividing the length of the 2D by the length of the 4D on
each hand. All measurements were obtained by two separate raters, and the ratios were compared to assess interrater reliability.

4.2.2. Facial Expression Viewing and Recognition Task. The Facial Expression Viewing and Recognition Task (FEVRT) is a novel task, developed at the Behavioural Neuroscience Research Laboratory at the University of Regina, which was used to display dynamic facial emotion stimuli to participants. Participants were shown video clips of faces morphing from a neutral expression to one of six basic facial expressions: happiness, anger, fear, disgust, surprise, and sadness. Video clips were created using a program called Fantamorph. The program was used to edit pairs of neutral and expressive pictures, taken from the previously validated Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Japanese and Caucasian Neutral Faces (JACNeuF) databases (Matsumoto & Ekman, 1988), to create a video clip which morphs from neutral pictures to expressive pictures, allowing a changing facial expression to be displayed. These databases were used because all images have been previously FACS coded and standardized, and are thus known to be comparable in expression intensity (Matsumoto & Ekman, 1988). However, because the FACS coding protocol for videos differs from the protocol used for still photographs (Ekman et al., 2002), the newly created video clips were FACS coded by the researcher, who is a certified coder.

The video clips were displayed using E-prime software. For each of the six emotions, three representative video clips were created giving a total of 18 separate video clips. Each block of trials consisted of two consecutive randomized presentations of the 18 clips, resulting in 36 trials per block. The video clips were two seconds in duration and, like in Hermans et al.’s (2006) study, the final frame of full expression remained
displayed for four seconds. Between trials there was a blank screen displayed for one second before the next clip was displayed. For the first block of trials participants passively viewed the facial expressions (allowing for spontaneous facial mimicry), and in the second block of trials they were asked to identify the emotion that was displayed, via key press on the keyboard number pad. Two rows of three keys each on the number pad were labeled from left to right as ‘anger,’ ‘disgust,’ and ‘fear’ (Row 1 – keyboard numbers 4, 5, and 6, respectively) and ‘happy,’ ‘sad,’ and ‘surprise’ (Row 2 – keyboard numbers 1, 2, and 3, respectively). To ensure participants were attending to the stimuli in the first block a control task was used, which required them to attend to the clips and indicate verbally whether the same clip had been presented consecutively.

Participants were video-recorded during both blocks of trials using a hidden camera present in the computer monitor, so that mimicry could later be assessed. Upon completion of the study the recordings of participants’ facial expressions were coded for specific muscle movements and intensity using the FACS. In this system, facial expressions are coded into 30 action units (AU) and 14 miscellaneous actions based on known movement patterns (Ekman et al., 2002). Examples of the AUs include brow lowerer (AU4) and upper lid raiser (AU5; Ekman et al., 2002). Some of the AUs are further coded for level of intensity. FACS coding was done at the end of all data collection by the researcher; during the coding of participants’ facial expressions, the researcher was blind to the expressions being displayed in the stimuli. In their experiments, Sayette, Cohn, Wertz, Perrott, and Parrott (2001) found good to excellent reliability for the FACS in coding spontaneous facial expressions.
4.2.3. Self-Report Empathy Measures. Two self-report empathy measures, the Empathy Quotient (EQ; Baron-Cohen & Wheelwright, 2004) and the Balanced Emotional Empathy Scale (BEES; Mehrabian, 1997), were chosen because they are commonly used in the literature and both demonstrate the expected sex difference, suggesting that scores on these measures may be related to sex hormone differences (Lawrence et al., 2004; Mehrabian, 2000). Additionally, these measures were chosen because they assess both cognitive empathy (the EQ) and emotional empathy (both the EQ and the BEES). Lastly, an additional reason for choosing the BEES was because an earlier version of this measure (the Questionnaire Measure of Emotional Empathy; Mehrabian & Epstein, 1972) has been found to correlate with facial mimicry (Sonnby-Borgstrom et al., 2003).

4.2.3.1. The Empathy Quotient (EQ). The EQ, also known as the Cambridge Behavioural Scale, is a 40 item self-report measure of empathy that requires participants to respond to questions on a 4-point scale ranging from “Strongly Agree” to “Strongly Disagree” (Baron-Cohen & Wheelwright, 2004). This scale measures cognitive empathy, emotion reactivity, and social skills. The EQ has demonstrated reliability across samples, and concurrent validity with another empathy measure (the Interpersonal Reactivity Scale; Davis, 1980). The EQ also has demonstrated construct validity (Lawrence et al., 2004). The EQ summary score was computed using the published scoring criteria (Baron-Cohen & Wheelwright, 2004). The lowest possible score on this measure was zero and the highest possible score was 80.

4.2.3.2. Balanced Emotional Empathy Scale (BEES). The BEES is a self-report measure of emotional empathy. The BEES was developed to examine individual
differences in the ability to feel the emotional experiences of others (Mehrabian, 1997). This 30-item measure of empathy uses a 9-point scale ranging from “Very Strongly Agree” to “Very Strongly Disagree,” and includes statements such as “unhappy movie endings haunt me for hours afterward” (Mehrabian, 2000). The internal consistency of this scale is .87, and it has been determined to be a one factor test of empathy (Mehrabian, 1997). The score for the BEES was also computed using the published scoring criteria (Mehrabian, 2000). To do this, a total score was calculated for the positively worded questions and a total score was calculated for the negatively worded questions. The total score for the negative questions was then subtracted from the total score for the positive questions to produce a final score. The lowest possible score on this measure was -120 and the highest possible score was 120.

4.2.4. Laterality and Demographics Questionnaire. A modified version of the Waterloo Handedness Questionnaire- Revised (WHQ-R; Elias, Bryden, & Bulman-Fleming, 1998) was used to assess individual differences in handedness, as past research has shown that handedness is related to brain organization. The WHQ-R assesses handedness using a 5-point scale, ranging from “Left Always” to “Right Always,” to assess the hand typically used to perform 15 different activities. This questionnaire was modified to include questions about age, race, and menstrual phase, because these factors have previously been shown to be related to sex hormone concentrations (Hausmann, 2000; Manning, 2002). Questions added to this measure also assessed pre-existing injury to the digits, because of the use of the 2D:4D ratio. The handedness score was computed by summing the scores on the 15 items of the WHQ-R, which has a minimum score of 15 (extreme left handedness) and a maximum of 75 (extreme right handedness).
4.2.5. Centre for Epidemiologic Studies Depression Scale (CES-D). The CES-D is a commonly used measure of mood and depression consisting of 20 questions pertaining to how an individual has felt and behaved over the past week (Radloff, 1977). Various studies have shown that mood has an effect on facial mimicry, with the emotional state of the participants biasing them to mimic certain emotions more than others (Niedenthal et al., 2001). Further, individuals suffering from depression have failed to show normal facial mimicry in past research (Wexler, Levenson, Warrenburg, & Price, 1994). The CES-D summary score was computed using the published scoring criteria (Radloff, 1977). The lowest possible score on this measure was zero and the highest possible score was 60.

4.2.6. Toronto Alexithymia Scale (TAS-20). The TAS-20 (Bagby, Parker, & Taylor, 1994) was used to assess alexithymia, which is a condition characterized by the inability to express emotion, discern the emotional states of others, and display empathy (Levant, Hall, Williams, & Hasan, 2009). The TAS-20 is a 20-item self-report measure containing questions that are rated on a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). Factor analysis has confirmed a three-factor structure which corresponds to the three subscales in this measure: difficulty identifying feelings (DIF), difficulty describing feelings (DDF), and externally oriented thinking (EOT; Parker, Keefer, Taylor, & Bagby, 2008). However, since this measure was included as a screening for alexithymia in general, only the total score was computed. The TAS-20 total score was computed using published scoring criteria (Bagby et al., 1994). The lowest possible total score on the scale was 20 and the highest possible total score was 100; scores over 51 are indicative of the presence of alexithymia.
4.2.7. Pre-screening questionnaire. The prescreening questionnaire consisted of 11 questions regarding: age, English fluency (spoken and written), vision, handedness, prior finger injury (that could affect the 2D:4D ratio), oral contraceptive use, and menstrual cycle regularity and phase.

4.3. Procedure

As illustrated in Figure 1, participants were pre-screened before they were scheduled for their sessions to ensure that they would be menstruating at the time of testing. Prescreening was done through email; the study was advertised on the Department of Psychology Research Participant Pool website, and prospective participants contacted the researcher via email to express their interest in participating. The researcher then emailed the prescreening questionnaire to the prospective participants, along with information on why the prescreening questions were being asked, a statement confirming consent, and instructions indicating that they were to complete the questionnaire only if they were interested in further participation. Once the questionnaire was completed, the testing session time was arranged based on menstrual cycle phase.

At the testing session, following informed written consent, participants completed the FEVRT and then the questionnaires (EQ, BEES, CES-D, TAS-20, and the modified WHQ-R). Once the FEVRT and the questionnaires were completed a digital photograph was taken of participants’ right and left hands so that the 2D:4D ratio could be later assessed for each hand. Because the major testing concerns were fatigue and knowledge of the hypotheses, the FEVRT was always completed first, followed by the administration of the questionnaires in a random order.
During the FEVRT, participants’ reactions to the stimuli were video-recorded using a hidden camera in the computer monitor. Given the need for covert video-recording of participants, disclosure of this did not occur at the time of initial written informed consent to ensure that recorded reactions were spontaneous and not a product of demand characteristics. Once testing was complete a debriefing was performed and participants were informed that they were video-recorded, and the rationale for the hidden camera and video footage was fully disclosed. Once this was done participants’ written informed consent to use the video footage was obtained. One participant did not consent to the use of her video footage, and therefore the footage was immediately destroyed and her data were excluded from the analyses.
<table>
<thead>
<tr>
<th>Event</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Screening</td>
<td>•This was to ensure testing was done during menstrual phase for each participant</td>
</tr>
<tr>
<td>Participants arrived</td>
<td>•Written informed consent was obtained</td>
</tr>
<tr>
<td>FEVRT completed</td>
<td></td>
</tr>
<tr>
<td>Questionnaires completed</td>
<td>•These were given in random order</td>
</tr>
<tr>
<td>Digital photos of hands were taken</td>
<td>•Both hands were photographed in order to calculate the left and right 2D:4D ratios</td>
</tr>
<tr>
<td>Participants were debriefed on video recording</td>
<td></td>
</tr>
<tr>
<td>Second written informed consent was obtained</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Procedural flow chart
4.4. Scoring

4.4.1. Scoring Protocol. The researcher performed all testing and analyses; because of this, care was taken to avoid researcher bias by completing the scoring of each measure blindly and separately. The video stimuli that comprised the FEVRT were FACS scored by the researcher prior to the start of any data collection in order to obtain the AU and intensity coding against which participants’ AU and intensity reactions would be compared; this was done to allow for the later computation of the mimicry scores (described in detail below). After data collection was complete, all video footage obtained from participants was scored prior to the other data to ensure that the researcher had no knowledge of the participants’ 2D:4D ratios or self-reported empathy scores. Because the FEVRT stimuli were displayed in a random order, the researcher was blind to which emotional expression the participants were viewing at the time of the recordings. To ensure that the scoring reflected mimicked expressions, participants’ facial expressions were not scored until after the time at which the stimuli started to display the emotion, which was determined to be .21 seconds into each trial. Detailed scoring information for the FEVRT is provided below.

Next the questionnaires were scored. Following this, the digital photographs that were taken of participants’ hands were downloaded and the 2D:4D ratios were determined. Detailed scoring information for the 2D:4D ratio is provided below. Once the above scoring was completed, the order of the trials displayed during the FEVRT was determined for each participant, and their mimicry scores were computed. Accuracy for the second block of the FEVRT was recorded by the E-prime software, allowing for the computation of the emotion recognition scores (described in detail below).
4.4.2. 2D:4D ratio. Using Corel Paint Shop Pro Photo X2, the researcher and a second rater examined the photos of participants’ hands and measured the length of each participant’s index and ring fingers on each hand; both raters were blind to the participants’ scores on the other measures while doing this. The measurements were used to compute the left and right hand 2D:4D ratios, by dividing the length of the index finger (2D) by the length of the ring finger (4D) for each hand. The ratios obtained by each rater were then analyzed using paired samples t-tests to check for significant differences; no significant differences were found between the ratios determined by each rater (all $p > .05$). Following procedure used in past research (e.g., van Anders & Hampson, 2005), the ratios determined by each rater for each hand of each participant were averaged to compute a single 2D:4D ratio score for each hand. The left and right hand ratios were not averaged with each other; it was important to analyze each hand’s ratio separately, since previous studies have shown relationships with either both hands’ ratios, or the right or left hand ratios independently (Honekopp et al., 2007).

4.4.3. FEVRT. Both the stimuli and the video recordings garnered from the FEVRT were scored using the FACS. The FACS is used to categorize facial expressions into nominal action units, with some of these action units also having intensity ratings associated with them. For each of the stimuli presented in the FEVRT the researcher determined the AU score, and the intensity score as applicable. The researcher performed the same coding on the participants’ facial reactions to the stimuli. The scores of the participants were then compared to the stimulus scores in order to compute the mimicry scores.
The first set of mimicry scores are AU scores that compare the action units shown in the stimuli to the action units displayed by the participants. For each matching action unit the participant was given one point, and the score for each trial was converted to a percentage that was calculated based on the number of matched AUs and the total AUs in the trial. For example, if the stimulus presented was scored as having AUs 1 (inner brow raiser), 6 (cheek raiser and lid compressor), 9 (nose wrinkler), and 28 (lip suck), and the participant displayed AUs 1, 9, and 28, then she would have an AU mimicry score of 75% on that trial (i.e., 3 matched AUs / 4 FEVRT AUs x 100 = 75%). This was done for each of the 36 trials in block 1 (passive viewing) and block 2 (active viewing – recognition).

For those action units that have an intensity score, a separate intensity mimicry score was also computed for each trial. The FACS uses the letter labels A through E to rate intensity. An intensity score of A indicates the lowest intensity level and an intensity score of E indicates the highest intensity level; for our purposes, an intensity level of A was assigned a score of 1, B a score of 2, and so forth, with the maximum level of E assigned a score of 5. If a participant correctly displayed an AU that had an associated intensity, the intensity score was compared to the stimulus’s intensity score for the trial. For example, if the scoring for a stimulus was 1A (presence of inner brow raiser), 6A (presence of cheek raiser and lid compressor), 9C (marked evidence of nose wrinkler), and 28D (severe-lip suck), giving it an intensity rating of 9 (i.e., 1 + 1 + 3 + 4 = 9), and the participant displayed 1B (slightly-inner brow raiser), 6A (presence of cheek raiser and lid compressor), 9B (slightly-nose wrinkler), and 28C (moderate-lip suck), giving her a intensity rating of 8 (i.e., 2 + 1 + 2 + 3 = 8), then her mimicry intensity score would be
88.9% on that trial (i.e., participant’s intensity score of 8 / FEVRT intensity score of 9 x 100 = 88.9%). Because of the way in which the mimicry intensity scores were calculated, it was possible that participants’ intensity scores could be higher than the stimulus intensity scores, resulting in a mimicry intensity score greater than 100%; however, this did not occur in the present sample.

Overall average AU mimicry scores and overall average mimicry intensity scores were computed, as well as an average AU mimicry score and an average mimicry intensity score for each block of the FEVRT and for each emotion type. Total emotion recognition accuracy scores were also computed for the second block of the FEVRT, both overall and by emotion type.

5. RESULTS

5.1. Inter-rater Reliability of 2D:4D Ratios

As described above, the 2D:4D ratios were determined by two independent raters. Paired sample t-tests were used to assess whether there were significant differences between the ratios determined by each rater. There were no significant differences between the 2D:4D ratios determined by each rater for the left hand, \( t(47) = 1.75, p > .05 \) (rater 1: \( M = .94, SD = .04 \); rater 2: \( M = .93, SD = .04 \)), and the right hand, \( t(47) = -.13, p > .05 \) (rater 1: \( M = .94, SD = .04 \); rater 2: \( M = .94, SD = .05 \)). In addition, the 2D:4D ratios determined by each rater were found to be significantly correlated for both the left \((r = .74, p < .001)\) and right \((r = .69, p < .001)\) hands. Since there were no significant differences between the 2D:4D ratios determined by each rater, and since they demonstrated large intercorrelations, they were averaged to create a single 2D:4D ratio.
for the left hand, and a single 2D:4D ratio for the right hand. Both the left 2D:4D ratio and the right 2D:4D ratio were used in the following analyses.

5.2. Exclusion of Individuals with possible Alexithymia and Depression, and Non-Mimickers

Alexithymia is a condition characterized by the inability to express or discern emotion and the absence of empathetic display (Levant et al., 2009). The TAS-20 was used in the present study to identify and exclude individuals with a high degree of alexithymia; a score greater than 51 on the TAS-20 indicates that an individual has or possibly has alexithymia (Bagby et al., 1994). In the present study, 16 of the 53 participants had TAS-20 scores over 51 and were removed from further analyses.

As noted in the measures section, depression has been shown to impact facial mimicry in previous studies (Wexler et al., 1994). In the present study the CES-D was used to identify and exclude individuals who appeared to be experiencing a depressed mood. Using a cutoff score of 27, based on previous research that used an all-female Canadian sample (Costello & Devins, 1989), 13 participants were identified who appeared to be experiencing a depressed mood, which resulted in the exclusion of an additional 7 participants (as 6 of the 13 also had TAS-20 scores above 51).

In addition, participants who displayed no mimicry were removed from analyses examining mimicry. This was done for two reasons: 1. the primary objective of the present study was to determine how PT relates to individual differences in the extent of mimicry, and 2. inclusion of participants who demonstrated no mimicry at all resulted in severely skewed distributions on the mimicry scores that could not be normalized through data transformations. This resulted in the removal of one participant from overall
FEVRT mimicry analyses, 3 participants from analyses examining mimicry in Block 1, and 7 participants from analyses examining mimicry in Block 2. These exclusions resulted in a final sample of 30 for the analyses that did not involve mimicry variables, and final samples of 27, 23, and 29 for analyses involving Block 1, Block 2, and overall FEVRT scores, respectively.

5.3. Data Transformations

To assess normality, descriptive statistics were obtained for all variables (i.e., all FEVRT, empathy, emotion recognition, TAS-20, and CES-D scores, and the 2D:4D ratios). The presence of outliers and non-normal distributions were discovered in the FEVRT AU mimicry and intensity scores and the emotion recognition scores. Because the chosen analyses required normality, procedures outlined by Tabachnick and Fidell (2007) were used to correct the distributions. Specifically, log10 transformations were performed on FEVRT AU mimicry and intensity scores (total, by block, and by emotion), which were all positively skewed, and reversals and square root transformations were performed on the emotion recognition scores (total and by emotion), which were all negatively skewed. The transformed values were used in all subsequent analyses.

5.4. Preliminary Analyses

First, a Pearson correlation analysis was used to determine if the scores on the two self-report measures of empathy were significantly correlated with each other. As expected, the BEES and EQ scores were significantly positively correlated ($r = .52, p < .002$). Pearson correlation analyses were also used to determine if the FEVRT AU mimicry and intensity scores (total and by block) were correlated with each other. All
FEVRT AU mimicry and intensity scores were significantly positively correlated with each other (all \( p < .005 \)).

Next, Pearson correlations were computed to determine if age was related to the FEVRT AU mimicry, intensity, and emotion recognition scores (total, by block, and by emotion type) or the self-reported empathy scores. Age was not significantly correlated with any of these scores (all \( p > .07 \)).

For the final preliminary analyses, independent samples \( t \)-tests were used to determine if FEVRT AU mimicry, intensity, emotion recognition (total, by block, and by emotion type), and the self-reported empathy scores differed between oral contraceptive users and naturally cycling women. Oral contraceptive users and non-users did not significantly differ on any of these variables (all \( p > .11 \)). As such, oral contraceptive use was not used as a factor in the main analyses.

5.5. Main Analyses

Multiple regression analyses were to be used to address the main hypotheses of the study. Tests of multicollinearity were initially run. Because multicollinearity was an issue for all multiple regression analyses, simple regressions were run instead, examining the influence of each predictor variable on the outcome variables separately. Specifically, it was found that the left and right 2D:4D ratios (predictor variables for addressing Hypotheses 1, 2, 3, and 6) had large condition indices (> 47), indicating that multicollinearity would be an issue if both predictors were used within the same analyses. In addition, the FEVRT AU mimicry and intensity scores (predictor variables in Hypotheses 4, 5, and 6) had large variance inflation factor (VIF) levels (> 15), similarly indicating that multicollinearity would be an issue if both predictors were used in the
same analyses. In addition, because all hypotheses were directional, one-tailed significance tests were used throughout.

5.5.1. Hypothesis 1: 2D:4D Ratio and Facial Mimicry. Women with lower 2D:4D ratios (higher PT) were expected to display less facial mimicry while observing dynamic facial expressions than women with higher 2D:4D ratios (lower PT). This hypothesis was addressed using left and right 2D:4D ratios as predictor variables and the overall average FEVRT AU mimicry and mimicry intensity scores as criterion variables. The left 2D:4D ratio showed a nonsignificant trend towards predicting AU mimicry scores, $\beta = 2.90, t(27) = 1.44, p = .08$, and trended towards explaining a significant portion of the variance in AU mimicry scores, $R^2 = .07, F(1, 27) = 2.06, p = .08$. The left 2D:4D ratio significantly predicted the mimicry intensity score, $\beta = 4.38, t(27) = 1.94, p = .03$, and explained a significant portion of the variance in the intensity scores, $R^2 = .12, F(1, 27) = 3.78, p = .03$. For both the AU mimicry and mimicry intensity scores, as the left 2D:4D ratio increased the level of mimicry increased, suggesting that lower PT is associated with higher mimicry. The right 2D:4D ratio did not significantly predict AU mimicry, $\beta = 0.83, t(27) = .37, p = .36$, or mimicry intensity, $\beta = 1.58, t(27) = .63, p = .27$.

Next, analyses were performed on the two blocks of the FEVRT separately (block 1- passive viewing, block 2- emotion recognition) in order to further examine the significant relationship found between the left 2D:4D ratio and the overall average FEVRT mimicry intensity score, and to further examine the trend found between the left 2D:4D ratio and the AU mimicry scores. No significant relationship was observed between AU mimicry in block 1 and left 2D:4D ratio, $\beta = 2.15, t(25) = 1.28, p = .11$. 
However, the left 2D:4D ratio was found to significantly predict mimicry intensity in block 1, $\beta = 3.42, t(25) = 1.86, p = .04$, with a significant portion of the variance in intensity explained, $R^2 = .12, F(1, 25) = 3.47, p = .04$. In block 2, the left 2D:4D ratio significantly predicted mimicry intensity, $\beta = 5.04, t(21) = 1.74, p = .05$, with a significant portion of the variance explained, $R^2 = .13, F(1, 21) = 3.03, p = .05$, and marginally predicted AU mimicry, $\beta = 3.78, t(21) = 1.61, p = .06$. Similar to what was observed for the total scores, these analyses showed that as the left 2D:4D ratio increased mimicry also increased, suggesting that lower PT exposure is related to higher mimicry.

Pearson correlations were computed for on each of the six emotion types to further examine the relationships between left 2D:4D and AU mimicry and mimicry intensity. No significant correlations were found between left 2D:4D and AU mimicry or intensity for the individual emotion types (all $p$s > .05).

5.5.2. Hypothesis 2: 2D:4D Ratio and Empathy. Women with lower 2D:4D ratios (higher PT) were expected to score lower on self-report measures of empathy than women with higher 2D:4D ratios (lower PT). To address this hypothesis, left and right 2D:4D ratios were used as predictor variables, and BEES and EQ scores as the criterion variables. The left 2D:4D ratio significantly predicted the BEES scores, $\beta = 197.12, t(28) = 1.76, p = .04$, with a significant portion of variance explained, $R^2 = .10, F(1, 28) = 3.11, p = .04$. This showed that as the left 2D:4D ratios increased the BEES scores also increased. The right 2D:4D ratio did not significantly predict the BEES scores, $\beta = 19.68, t(28) = .17, p = .44$, and EQ scores were not significantly predicted by either the left 2D:4D ratio or the right 2D:4D ratio, $\beta = -13.24, t(27) = -.37, p = .36$, and $\beta = 30.52, t(27) = .82, p = .21$, respectively.
5.5.3. Hypothesis 3: 2D:4D Ratio and Emotion Recognition. Women with lower 2D:4D ratios (higher PT) were expected to have lower emotion recognition accuracy scores than women with higher 2D:4D ratios (lower PT). To address this hypothesis, left and right 2D:4D ratios were used as predictor variables and emotion recognition scores were used as the criterion variable. The left 2D:4D ratio significantly predicted emotion recognition accuracy, $\beta = -4.35$, $t(28) = -2.17$, $p = .02$, explaining a significant portion of the variance, $R^2 = .14$, $F(1, 28) = 4.70$, $p = .02$. As left 2D:4D ratios increased the reversed emotion recognition scores decreased; examination of the non-transformed emotion recognition scores shows that as 2D:4D ratios increased emotion recognition scores also increased. The right 2D:4D ratio did not significantly predict emotion recognition accuracy, $\beta = 3.18$, $t(28) = 1.51$, $p = .07$.

The relationship between left 2D:4D and emotion recognition accuracy was examined for the six emotion types using Pearson correlations. Left 2D:4D and sadness were significantly positively correlated, $r = .31$, $p = .05$, indicating that as the left 2D:4D ratio increased the accuracy of recognizing sadness increased. All other correlations were non-significant (all $p$s > .06).

5.5.4. Hypothesis 4: Facial Mimicry and Empathy. Facial mimicry scores were expected to predict scores on self-report measures of empathy, such that as facial mimicry scores increased empathy scores were also expected to increase. To test this hypothesis, overall average FEVRT AU mimicry and mimicry intensity scores were used as the predictor variables, and BEES and EQ scores as the criterion variables. BEES scores were not predicted by either AU mimicry or mimicry intensity scores, $\beta = 6.42$, $t(27) = .60$, $p = .28$, and $\beta = 7.52$, $t(27) = .81$, $p = .21$, respectively. Similarly, EQ scores
were not predicted by either AU mimicry or mimicry intensity scores, $\beta = -1.08, t(26) = -.33, p = .37$, and $\beta = -1.33, t(26) = -.47, p = .32$, respectively.

5.5.5. Hypothesis 5: Facial Mimicry and Emotion Recognition. Facial mimicry scores were expected to predict emotion recognition accuracy scores, such that as facial mimicry scores increased emotion recognition accuracy scores were expected to increase. To address this hypothesis overall average FEVRT AU mimicry and mimicry intensity scores were used as predictor variables and emotion recognition accuracy as the criterion variable. Emotion recognition accuracy was not significantly predicted by either FEVRT AU mimicry or mimicry intensity scores, $\beta = -.21, t(27) = -1.07, p = .15$, and $\beta = -.18, t(27) = -1.05, p = .15$, respectively.

Mimicry scores from block 2 of the FEVRT were used to further examine the potential relationship between mimicry and emotion recognition, because explicit emotion recognition was required only in this block. Emotion recognition accuracy was not significantly predicted by either the block 2 FEVRT AU mimicry or mimicry intensity scores, $\beta = -.03, t(21) = -.14, p = .89$, and $\beta = -.05, t(21) = -.23, p = .41$, respectively.

Pearson correlations were computed for each of the six emotion types to determine if accuracy on each was related to overall AU mimicry and mimicry intensity scores or to the AU mimicry and mimicry intensity scores for block two, specifically, in which emotion recognition was required. No significant correlations were found between the individual emotion types and any of these mimicry scores (all $ps > .07$).
5.5.6. Hypothesis 6: Mediating Relationships. The sixth hypothesis predicts that, if the five previously hypothesized relationships were found, then:

1. mimicry was expected to mediate the relationship between the 2D:4D ratio and self-reported empathy;
2. mimicry was expected to mediate the relationship between the 2D:4D ratio and emotion recognition;
3. emotion recognition was expected to mediate the relationship between 2D:4D ratio and self-reported empathy.

Using the method proposed by Baron and Kenny (1986), a stepwise procedure was followed for each of the hypothesized relationships to determine if mediation occurred (see Figure 2 for an overview of this procedure). In order for there to be a mediational relationship, the predictor variable (X) needs to be significantly related to the criterion variable (Y; Step 1), X must be significantly related to the mediator variable (M; Step 2), and M must be significantly related to Y, while controlling for X (Step 3). If these first three steps are satisfied, then the final step to test for mediation involves determining if the relationship between X and Y still exists when M is controlled for – if this relationship is no longer present when M is controlled for, then full mediation can be assumed. Alternatively, if only the first three steps are satisfied, then partial mediation is likely. Mediation is not present if the first three steps are not satisfied. This procedure is used to determine if a more complex relationship exists between the variables being tested. In this study, the purpose of the first and second mediation analysis was to determine if facial mimicry was mediating the relationships between 2D:4D ratio and empathy, and between 2D:4D ratio and emotion recognition, respectively. The purpose of
the third mediation analysis was to determine if emotion recognition accuracy was
mediating the relationship between 2D:4D ratio and empathy.

Only the left 2D:4D ratio was used in each of the mediation analyses because only
the left 2D:4D ratio was significantly related to any of the variables in the previous
analyses (i.e., Hypotheses 1 - 3). Further, overall average AU mimicry and mimicry
intensity scores were used in the mediation analyses because these scores were either
significantly related to left 2D:4D or showed a trend towards significance (Hypothesis 1).
Similarly, the BEES scores were used for the measure of empathy in the first model
because they were significantly related to the left 2D:4D ratio, whereas the EQ was not
significantly related to either 2D:4D ratio (Hypothesis 2).

Step one was satisfied for each of the analyses, because the left 2D:4D ratio
significantly predicted the BEES and emotion recognition accuracy scores (see Figures
3a, 3b, 4a, 4b, and 5). Step 2 was satisfied for most of the models; left 2D:4D ratio
significantly predicted mimicry intensity (Figures 3b and 4b) and emotion recognition
accuracy scores (Figure 5), while marginally failing to significantly predict AU mimicry
scores (Figures 3a and 4a). The third step was not satisfied for any of the models, as
mimicry scores did not significantly predict BEES (Figures 3a and 3b) or emotion
recognition accuracy scores (Figures 4a and 4b), and emotion recognition scores did not
significantly predict BEES scores (Figure 5). Because all three steps were not satisfied in
any of the models, it can be concluded that there was no mediation occurring.
Figure 2: Outline of mediation analysis
Figure 3a: Mediation analysis for 2D:4D, AU mimicry, and empathy

Figure 3b: Mediation analysis for 2D:4D, mimicry intensity, and empathy
Figure 4a: Mediation analysis for 2D:4D, AU mimicry, and emotion recognition

Figure 4b: Mediation analysis for 2D:4D, mimicry intensity, and emotion recognition
Emotion recognition accuracy

Left 2D:4D ratio \( p = .04 \)

BEES \( p = .16 \)

\( p = .03 \)

Figure 5: Mediation analysis for 2D:4D, emotion recognition, and BEES scores
6. DISCUSSION

The primary goal of the present study was to examine the relationships among PT and various components of empathy, facial mimicry, and emotion recognition. Previous research has shown that testosterone administration affects facial mimicry in women (Hermans et al., 2006), and that PT concentrations are related to cognitive empathy and social functioning in children (Chapman et al., 2006; Knickmeyer et al., 2005, 2006). The present study extended this area of research by examining the relationships among PT exposure (estimated using the 2D:4D ratio) and various putative components of cognitive and emotional empathy in women, using both objective behavioural measures and self-report measures of emotion.

6.1. Prenatal Testosterone and Emotion Processing

The first three hypotheses suggested that PT would demonstrate negative relationships with facial mimicry, self-reported empathy, and emotion recognition scores. Specifically, it was expected that higher PT exposure would be associated with less mimicry, lower empathy scores, and less accurate emotion recognition. As expected, high estimated PT levels were significantly associated with mimicry that was at lower intensities, and demonstrated a trend towards being associated with mimicry of fewer facial movements. The same negative relationship was observed between estimated PT and emotion recognition accuracy – high levels of PT were associated with significantly lower emotion recognition accuracy scores. This relationship was also found between estimated PT and scores on the self-report measure of emotional empathy (the BEES), such that higher PT was associated with lower empathy. The relationship between PT and empathy was not found with the self-report measure that assessed both cognitive and
emotional empathy (the EQ), despite both self-report empathy measures being highly correlated with each other.

Previous studies examining the relationship between PT and empathy have found negative relationships in clinically normal children (Knickmeyer et al., 2005, 2006) and children with autism and Asperger syndrome (Manning et al., 2001), and have largely focused on cognitive empathy and social functioning using behavioural measures and parental reports. The present study was the first to examine the relationships between PT and both cognitive and emotional empathy measures in adults. Previous studies used behavioural measures of empathy that were quite different from those used in the present study, including assigning social relationships to abstract cartoons (Knickmeyer et al., 2006), and the “Reading the mind in the Eyes” test (Chapman et al., 2006). These measures assess cognitive empathy, as did the emotion recognition measure in the present study. While these measures are quite different from one another, they are similar in that each asks participants to identify emotional states in the stimuli. The present study also included facial mimicry measures, which were intended to index emotional empathy. Despite great differences in the measures used to assess empathy, and despite differences in the type of empathy assessed (cognitive versus emotional), the previous studies on PT and empathy and the present study all produced the same pattern of results: higher levels of PT were associated with poorer performance on behavioural measures of empathy. It is also noteworthy that this same pattern of results was observed across the studies, despite the use of differing measures of PT; Knickmeyer, Chapman, and colleagues’ studies used amniocentesis to index PT, whereas the present study used the 2D:4D ratio. Collectively, these results suggest that there is a robust relationship between PT and empathy that can
be observed using many different types of behavioural measures of empathy, and different measures of PT. It is also important to note that the present study found this relationship in adults, compared to the previous studies which observed this relationship in children. This suggests that the relationship between PT and empathy may be static and observable across the lifespan. This study’s findings of relationships between PT and emotion recognition, self-reported emotional empathy, and mimicry intensity in adults extend the literature by showing that the relationship between PT and empathy continues into adulthood, and can be observed using a variety of measures.

Further, these findings fit with the literature on neural regions affected by prenatal sex hormones. Areas of the brain underlying mimicry, empathy, and emotion recognition have high concentrations of sex hormone receptors during development and demonstrate volumetric sexual dimorphisms in adulthood (Goldstein et al., 2001); such regions include the premotor cortex, IFG, IFO, insula, and amygdala (Carr et al., 2003; Iacoboni, 2007; Leslie et al., 2004; Phan et al., 2002). Collectively, the results of the neural correlates studies, the present study, and the above-noted studies in children suggest that PT has organizational effects on neural regions underlying emotional processing, resulting in long lasting effects on empathy behaviours that are observable in both children and adults.

The present study’s results concerning organizational sex hormone effects on mimicry and emotion recognition extend those of Hermans et al. (2006) and van Honk and Schutter (2007) pertaining to activational sex hormone effects on these processes. Hermans and colleagues showed that a negative causal relationship exists between circulating testosterone concentrations and facial mimicry in women, by demonstrating
that exogenous testosterone administration decreased facial mimicry of happy and angry expressions compared to placebo in their all-female sample. In their study, van Honk and Schutter showed that administration of exogenous testosterone to women decreased their recognition of angry facial expressions compared to placebo, but recognition of other basic emotions was not affected. The present study’s results are consistent with organizational effects of PT on facial mimicry intensity and emotion recognition, in addition to the activational effects reported by Hermans et al. and van Honk and Schutter. Collectively, these findings support the idea that sex hormones have both organizational and activational effects on putative behavioural components of emotional (mimicry) and cognitive (emotion recognition) empathy in female adults.

It is likely the case that the organizational effects of PT exposure on emotion processing preprogram the brain for later activational effects of circulating testosterone in adulthood. Support for this idea comes from a recent study that examined both activational and organizational effects of testosterone in women performing a behavioural measure of cognitive empathy – Baron-Cohen’s “Reading the Mind in the Eyes” test (van Honk et al., 2011). Participants had their 2D:4D ratios measured and completed two testing sessions involving the empathy test, one in which they were given a dose of sublingual testosterone and one in which they were given a placebo. It was found that the introduction of exogenous testosterone significantly impaired participants’ performance on the empathy test, but only in those who were previously exposed to high prenatal testosterone (as indicated by 2D:4D). These results lend further support to the idea that, while circulating sex hormones do affect emotion processing, the prenatal sex hormone
environment preprograms individual abilities in emotion processing (van Honk et al., 2011).

6.1.1. Prenatal Testosterone and Mimicry during Active versus Passive Emotion Viewing. Consistent with the literature, the present study found that the relationship between estimated PT and emotion expression mimicry was largely observed for both passive and active viewing of the emotional expressions. During the active viewing condition, participants identified the expressed emotions. The mimicry intensity scores demonstrated significant relationships with PT in both the active and passive viewing conditions; however, the AU mimicry relationship only approached significance during the active viewing condition. The reason for the non-significant relationship between AU mimicry and PT in the passive viewing block might be explained by the differing patterns of brain activation accompanying passive and active viewing states. Studies show that passive viewing of emotional expressions activates brain regions which have large numbers of PT receptors during development (Carr et al., 2003; Goldstein et al., 2001; Leslie et al., 2004), however, active viewing in which emotion identification is required is associated with more activation in these areas compared to passive viewing (Carr et al., 2003). Thus, it is not surprising that both mimicry measures were associated with PT during the active viewing condition. Further, the fact that the significant relationship between PT and mimicry intensity was observed for both the passive and active viewing blocks of the present study suggests that mimicry intensity may be a more sensitive measure for this effect than is the number of mimicked facial movements.

6.1.2. Prenatal Testosterone and Behavioural versus Self-Report Emotion Measures. While the putative behavioural empathy measures (mimicry and emotion
recognition) in the present study demonstrated the expected relationships with estimated PT exposure, scores on the self-report empathy measures differed in their relationships with estimated PT exposure. A significant relationship between self-reported emotional empathy and PT was observed, but unexpectedly, no relationship was found between PT and scores on the self-report measure of both cognitive and emotional empathy. The data from the empathy questionnaires in the present study suggest that PT has a greater effect on emotional empathy (characterized by observing, mimicking, and experiencing another’s state) than on cognitive empathy (characterized by understanding another’s state). While the finding of a significant relationship between emotion recognition and PT in the present study would seem to contradict this, since emotion recognition is often considered to be part of cognitive empathy (e.g., Nummenmaa et al., 2008; Bons et al., 2012), studies of brain activation during emotional empathy versus cognitive empathy show that there is increased emotion processing, perceiving of faces, and understanding and simulating others’ actions during emotional empathy compared to cognitive empathy (Nummenmaa et al., 2008). Such findings would suggest that emotion recognition might be better classified as the last step of emotional empathy.

It is interesting that PT was related to the self-report measure of emotional empathy, but not the self-report measure of emotional and cognitive empathy in the present study’s adult sample, despite the fact that scores on these two measures were significantly correlated with each other. This finding may suggest that PT is more related to emotional empathy than to cognitive empathy in adults. Past studies’ findings in children (e.g., Chapman et al., 2006; Knickmeyer et al., 2006) have found relationships between PT and cognitive empathy, but have not assessed emotional empathy.
Collectively these studies may suggest that PT is initially related to cognitive empathy (and perhaps emotional empathy) in childhood, but is only related to emotional empathy in adulthood. If this is the case, it may be possible that cognitive empathy is affected to a greater extent by socialization, so the relationship between cognitive empathy and PT diminishes in adulthood, whereas emotional empathy may be more innate and less affected by socialization, demonstrating a relationship with PT that remains into adulthood.

Alternatively, the lack of significant relationships between PT and self-reported emotional and cognitive empathy scores in the present study could be explained in terms of the hormone sensitivity of neural regions underlying a self-report task. As discussed by Hermans et al. (2006), the main areas of the brain targeted by steroid hormones are located in the limbic system. This means that hormone action is more likely to modulate basic affective processes, like facial mimicry, while having only an indirect effect on the more cognitive self-evaluation processes that self-report measures require, which would involve higher cortical processing (Hermans et al., 2006). Related to this line of reasoning, of the two self-report measures used, it was the emotional empathy measure, which arguably taps into aspects of empathy that involve greater limbic activity, that showed a significant relationship with PT; the self-report measure of both emotional and cognitive empathy presumably taps aspects of empathy that involve limbic and cortical regions to a greater extent, and showed no relationship with PT.

6.2. Relationships among Mimicry, Empathy, and Emotion Recognition

The second group of hypotheses aimed to determine if there were relationships among different forms of emotion processing. The literature suggested that a significant
relationship would be observed between facial mimicry and self-report measures of empathy (Chartrand & Bargh, 1999; Sonnby-Borgstrom et al., 2003); however, the data in the present study did not show relationships between the facial mimicry AU and intensity scores and either the BEES or EQ self-report empathy scores. This difference in results may be due to the two studies mentioned above having measured mimicry in a very different way than how it was examined in the present study, in addition to a strictly cognitive empathy self-report measure being used by Chartrand and Bargh (1999). In Chartrand and Bargh’s study, both facial and bodily mimicry were assessed through a simple visual examination of whether specific movements (smile, face rub, etc.) were present or absent. In Sonnby-Borgstrom and colleagues’ study (2003), EMG was used to assess facial mimicry of anger and happiness by examining only hallmark movements associated with these expressions (corrugators supercilii muscle movements associated with anger, and zygomaticus major muscle movements associated with smiling). The present study assessed mimicry using the FACS by taking into account both the intensity of mimicked expressions and the percentage of all facial movements (i.e., AUs) that were mimicked, not simply the hallmark movements. These measurements were more precise than the gross ‘mimicry present-or-absent’ approach used by Chartrand and Bargh, and more comprehensive than the hallmark movements examined by Sonnby-Borgstrom et al. However, because the present study’s mimicry scores mathematically accounted for all facial movements occurring in the stimuli, the mimicry scores might have been less sensitive to the hallmark movements, and less sensitive than the gross present versus absent approach used in previous studies. These differing approaches may account for the discrepant findings concerning the mimicry and self-reported empathy relationships.
Alternatively, significant relationships between mimicry and empathy may have been observed in Sonnby-Borgstrom et al. (2003), but not in the present study, due to the greater level of sensitivity to muscle movements that EMG provides compared to the FACS. EMG is able to record minute muscle movements that are not observable to the human eye. While the EMG approach to assessing facial mimicry may be more sensitive to muscle movements than the FACS, the FACS assessment of mimicry is more ecologically valid since it is measuring what the human eye can actually perceive as movement. In addition, the FACS is a more ecologically valid approach to assessing mimicry than is EMG, because EMG involves the intrusive presence of electrodes on participants’ faces. The presence of the electrodes may cause an exaggeration of facial movements. Further, the use of electrodes likely makes participants aware that their facial movements are being recorded (despite the fact that this was not explicitly disclosed in Sonnby-Borgstrom and colleagues’ study), potentially causing demand characteristics, which could then cause a relationship to be observed with self-reported empathy scores, which are also subject to demand characteristics. In the present study, participants were not aware that they were being recorded during the FEVRT, thus avoiding the influence of demand characteristics on the mimicry scores.

The literature also suggested that a positive relationship would be found between emotion recognition and facial mimicry (Niedenthal et al., 2001; Oberman et al., 2007). However, when directly tested no significant relationships were found between facial mimicry (AUs and intensity) and emotion recognition accuracy. Previous research has shown that there is a decrease in emotion recognition speed when facial mimicry is blocked (Niedenthal et al., 2001; Stel & Knippenberg, 2008), and poorer recognition of
certain emotions when facial mimicry is blocked (Stel & Knippenberg, 2008). The present study differs from the previous literature in that participants’ mimicry was not restricted – participants mimicked freely and were not aware that mimicry was being assessed. Thus, it may be that blocking mimicry impairs emotion recognition accuracy and speed, but that individual differences in naturally-occurring mimicry are not related to individual differences in emotion recognition accuracy. Further, emotion recognition response time was not recorded in the present study. This was done to allow maximum mimicry to occur, by allowing participants to view the fully formed facial expression, instead of stopping the expression video as soon as the displayed emotion was recognized. This difference in methodology between the present study and previous research may also account for the differing findings. Only recording accuracy, while necessary to the design of this study, may have decreased the power to find individual differences in emotion recognition, since all participants performed quite well on this task (accuracy ranged from 86.1-100%); thus, the ceiling effect on the emotion recognition accuracy measure may have prevented the relationship between emotion recognition and mimicry from having been observed in the present study. Because greater individual variability is observed on reaction time measures compared to accuracy measures, reaction time may be a more sensitive metric of individual differences in emotion recognition that may in turn demonstrate relationships with mimicry scores. Alternatively, it may be that facial mimicry facilitates emotion recognition speed, as suggested by Niedenthal et al.’s (2001) and Stel and Knippenberg’s (2008) findings, but does not facilitate emotion recognition accuracy.
6.3. Mediating Relationships among Prenatal Testosterone and Emotion Measures

The last set of hypotheses suggested that mediating relationships might exist among PT and facial mimicry, emotion recognition, and self-reported empathy. While the expected relationships between PT and mimicry, emotional empathy, and emotion recognition accuracy were found, no significant relationships were observed between mimicry and emotional empathy, between mimicry and emotion recognition accuracy, or between emotion recognition accuracy and emotional empathy. Thus, no mediating relationships appear to exist amongst these variables. The data do not support the hypothesis of complex relationships whereby mimicry, influenced by PT, influences self-reported empathy and emotion recognition accuracy. Instead these findings suggest that mimicry and emotion recognition accuracy are distinct processes that are each independently influenced by PT.

6.4. Limitations

While significant results in line with the hypotheses were found, no study is without limitations. One limitation of the present study is the use of the 2D:4D ratio. Although numerous lines of evidence suggest that the 2D:4D ratio is associated with PT levels (e.g., Lutchmaya et al., 2004; Manning, 2002), it is still an indirect measure, as is the case with most measures of PT. The 2D:4D ratio was chosen because it was the best indirect measure available for the chosen sample. For future research, it is important to demonstrate converging findings using other PT measures and to further explore the relationships between PT and mimicry, emotion recognition, and empathy in special populations, such as women affected by CAH.
While both the right and left 2D:4D ratios were measured and analyzed in the present study, only the left 2D:4D ratio showed significant relationships with the measures. It has been argued that there are no persuasive theoretical reasons to suggest that right and left hands differ in their sensitivity to PT (Luxen & Buunk, 2005). In a meta-analysis on the use of 2D:4D ratios, it was clear that left and right 2D:4D ratios have been used in past research, with significant results being found with one hand or the other in some studies, and both hands in other studies (Honekopp, Bartholdt, Beier, & Liebert, 2007). There is no clear reason why different behavioural measures demonstrate a different relationship with either the left or right 2D:4D ratio (Honekopp et al., 2007). Future studies should further explore this issue.

Another limitation was that the facial emotion viewing and recognition task involved the facial expressions being viewed for a fairly lengthy period of time (4 seconds), in order to ensure that the participants had adequate time to mimic the facial expression shown. In real life, genuine expressions are often brief, and many emotion recognition tasks involve a brief exposure time, such as the Facialy Expressed Emotion Labeling task which displays the expressions for 300 milliseconds (Braun, Traue, Frisch, Deighton, & Kessler, 2005). The need to have the facial expression displayed for longer for the purposes of mimicry might have decreased individual differences in emotion recognition accuracy by creating a relatively easy emotion recognition task. In turn, the ease of this task might have contributed to the absence of a significant relationship between facial mimicry and emotion recognition in the present study, which has been found in earlier studies that measured facial mimicry using EMG. Further, only accuracy, but not response time, was assessed in the present study because, as noted previously, in
order to allow for maximum mimicry to occur participants had to be allowed to view the
fully formed facial expression instead of stopping the expression as soon as they
recognized the emotion displayed. Since response time is widely used to examine
individual differences in abilities, and would combat the ceiling effect on the emotion
recognition accuracy measure, being unable to gather this information due to study
design might have limited the full examination of individual differences in emotion
recognition. The FEVRT also had the limitation of the passive viewing block always
being presented first, instead of being counterbalanced with the active viewing block.
This was done so that the participants were less likely to know what was being examined
during the observation phase, so that participants would be more likely to react to the
stimuli in a spontaneous fashion. However, having the passive viewing block first may
have caused practice effects that could have impacted performance during the active
viewing block.

A final limitation was that the sample size was smaller than planned. The present
study had strict inclusion criteria to limit individual variability on factors related to sex
hormone levels, which severely limited the number of individuals in the participant pool
who were eligible to participate. Further, individuals who possibly had alexithymia and
those who appeared to be experiencing a depressed mood were excluded from the
analyses, as depressed mood and alexithymia have previously been shown to be related to
mimicry, empathy, and emotion processing (Levant et al., 2009; Wexler et al., 1994).
With these criteria in place, the sample likely produced cleaner results that were less
influenced by extraneous variables. However, these criteria greatly reduced the sample
size, leading to a reduction in power to detect effects in this study.
6.5. Future Directions

It is important to replicate the findings of the present study using differing measures of PT, mimicry, emotion recognition, and empathy so that convergent validity and generalization can be achieved across studies. It is also important in future studies to look at the role of prenatal and circulating sex hormones concurrently when examining facial mimicry, emotion recognition, and other empathy measures, to expand on this study and the ones performed by Hermans et al. (2006) and van Honk and Schutter (2007). van Honk et al. (2011) found that performance on a behavioural measure of cognitive empathy demonstrated an interactive relationship with PT and circulating testosterone concentrations, but these relationships have not yet been examined for mimicry and emotion recognition. Adapting the mimicry and emotion recognition task to incorporate a response time measure and a quicker display of the emotions may also allow for further exploration of the relationships among PT, facial mimicry, and emotion recognition.

Further, future studies might also explore the relationships between sex hormones and neural activation during empathy, emotion recognition, and mimicry, by performing functional neuroimaging on participants as they complete tasks assessing these abilities and behaviours. By examining prenatal as well as circulating sex hormone levels, differing patterns of neural activation between individuals can be studied to determine what effect sex hormones may have on the function of brain regions underlying these emotion processing tasks.

Lastly, it is important that future studies in this area examine both men and women. Women often score higher than men on both self-report and behavioural
measures of emotion processing, as discussed in the literature review section of this thesis. As well, there are large sex differences in prenatal sex hormone concentrations, and in circulating sex hormone concentrations throughout adulthood. Thus, it is important to examine the relationships among PT, facial mimicry, emotion recognition, and self-reported empathy scores within an adult male population, to determine if relationships are similar to or different from those which were observed in the present study’s female sample. Other future research in this area should also endeavor to study both men and women, either between-studies or within-studies depending upon the research designs and goals. Extending the research to include both sexes will enable a better understanding of how prenatal and circulating sex hormones influence adult emotion processing, and of the role these hormones play in influencing emotion processing differences between men and women.

6.6. Conclusion

The present study found relationships between PT exposure and mimicry, between PT exposure and emotion recognition, and between PT exposure and self-reported emotional empathy, lending support to the idea that prenatal sex hormones have effects on adult emotional behaviour and emotion processing. These results suggest that the sex hormones in the prenatal environment have organizational effects on brain development, allowing them to have long-lasting effects on behaviour, which continue to be observed in adulthood. By furthering the understanding of such effects, disorders like those in the autism spectrum may be better understood from a development perspective.

The lack of support for a more complex relationship among prenatal sex hormones, facial mimicry, self-reported empathy, and emotion recognition also expands our understanding
of how different types of empathy (emotional vs. cognitive) and individual differences in empathy develop.

The processing of emotion is a key component of our social functioning. By understanding the role that the prenatal sex hormone environment has on the development of emotion processing, we can gain a better understanding of how individual differences in emotion processing arise.
7. REFERENCES


DATE: June 3, 2011

TO: Emile Kossick
    Psychology Department

FROM: Dr. Bruce Plouffe
      Chair, Research Ethics Board

Re: The Relationship Between a Prenatal Sex Hormone Indicator and Emotion Processing (File # 98S1011)

Please be advised that the University of Regina Research Ethics Board has reviewed your proposal and found it to be:

☐ 1. APPROVED AS SUBMITTED. Only applicants with this designation have ethical approval to proceed with their research as described in their applications. For research lasting more than one year (Section 1F), ETHICAL APPROVAL MUST BE RENEWED BY SUBMITTING A BRIEF STATUS REPORT EVERY TWELVE MONTHS. Approval will be revoked unless a satisfactory status report is received. Any substantive changes in methodology or instrumentation must also be approved prior to their implementation.

☐ 2. ACCEPTABLE SUBJECT TO MINOR CHANGES AND PRECAUTIONS (SEE ATTACHED). Changes must be submitted to the REB and approved prior to beginning research. Please submit a supplementary memo addressing the concerns to the Chair of the REB. **Do not submit a new application. Once changes are deemed acceptable, ethical approval will be granted.

☐ 3. ACCEPTABLE SUBJECT TO CHANGES AND PRECAUTIONS (SEE ATTACHED). Changes must be submitted to the REB and approved prior to beginning research. Please submit a supplementary memo addressing the concerns to the Chair of the REB. **Do not submit a new application. Once changes are deemed acceptable, ethical approval will be granted.

☐ 4. UNACCEPTABLE AS SUBMITTED. The proposal requires substantial additions or redesign. Please contact the Chair of the REB for advice on how the project proposal might be revised.

Dr. Bruce Plouffe

cc: Dr. L. Sykes Tottenam
cc: Dr. D. Alfano

**supplementary memo should be forwarded to the Chair of the Research Ethics Board at the Office of Research Services (Research and Innovation Centre, Room 169) or by e-mail to research.ethics@uregina.ca

Phone: (306) 585-4775
Fax: (306) 585-4801